Received 5 February 2024, accepted 26 February 2024, date of publication 28 March 2024, date of current version 4 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3377699

RESEARCH ARTICLE

REN-A.I.: A Video Game for AI Security Education Leveraging Episodic Memory

MINE ARAI[1,2], KOKI TEJIMA[1], YUYA YAMADA[1], TAKAYUKI MIURA[1], KYOSUKE YAMASHITA[1], CHIHIRO KADO[1], REI SHIMIZU[1], MASATAKA TATSUMI[1], NAOTO YANAI[1], (Member, IEEE), AND GOICHIRO HANAOKA[2], (Member, IEEE)

[1]Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565-0871, Japan
[2]AIST, Koto, Tokyo 135-0064, Japan

Corresponding author: Mine Arai (m-arai@ist.osaka-u.ac.jp)

This work was supported in part by Japan Science and Technology Agency (JST) CREST under Grant JPMJCR21M5, in part by JST SAKIGAKE under Grant JPMJPR23P6, and in part by JSPS KAKENHI under Grant 22H03591 and Grant 23H00479.

ABSTRACT Education in cybersecurity is crucial in the current society, and it will be extended into the artificial intelligence (AI) area, called AI security, in the near future. Although many video games for education in cybersecurity have been designed, we have two problems for education in AI security: a helpful design of a video game for users to learn cybersecurity is still unclear, and there is no game for AI security, to the best of our knowledge. In this paper, we design a video game for education in AI security, REN-A.I., to address the above problems. In designing REN-A.I., we built some hypotheses: simulating damage caused by attacks on AI and the effectiveness of their countermeasures through a video game helps a user to improve awareness of AI security with the episodic memory of the user itself. We focus on game scenarios and game functionalities to learn AI security with episodic memory in accordance with the above hypothesis. We conducted a questionnaire survey with 48 users to evaluate REN-A.I.. As a result, we confirm that both game scenarios and game functionalities are effective for learning with episodic memory. Specifically, 74% of users consider game scenarios effective, and 81% of users consider game functionalities effective. Our survey results have revealed two suggestions for beneficial design aspects in video games for education in cybersecurity. In particular, users who read game scenarios in REN-A.I. can learn AI security by the game more effectively than the other users. Furthermore, the functionality for accuracy deterioration due to attacks in REN-A.I. is effective even for users who do not read the game scenario. REN-A.I. is publicly available (https://www-infosec.ist.osaka-u.ac.jp/software/ren-ai/REN-AI(EN).html).

INDEX TERMS AI security, episodic memory, questionnaire survey, security education, video game.

I. INTRODUCTION

Cybersecurity is important for not only experts but also non-experts because many electronic devices, such as smartphones, have been used widely in recent years. Cybersecurity threats become more complex, and their risk affects the connectivity of critical infrastructure systems, economics, and public safety [1]. Namely, cybersecurity is an important component of an organization's overall risk management. Cybersecurity is also extended into artificial intelligence (AI) recently [2], which is known as AI security because AI has been leveraged in various areas, e.g., medical image analysis [3] and natural language processing [4]. Improving knowledge and awareness is an essential step in education in cybersecurity [5].

Many educators teach security by showing some incidents and their countermeasures [6]. Nonetheless, learners may need high motivation for a good education because the current education methods are often passive, e.g., learners just listen to talk about the incidents [7]. Furthermore, to achieve a high level of learning effectiveness, learners need to maintain a high level of motivation [8]. Security is also a secondary

The associate editor coordinating the review of this manuscript and approving it for publication was Mahdi Zareei.

technology for non-experts [6]. Furthermore, the routine use of security mechanisms is sometimes undesirable, and they often lack motivation [9]. Enabling users to make appropriate decisions about security often needs at least a somewhat shifting mental model [10]. Consequently, many motivational methods for security education have been investigated, e.g., videos, games, and comics [5].

In this paper, we explore video games as a method for education in cybersecurity since a video game as a learning tool can cover a wide range of users from children to adults [5]. In particular, users can learn by trial and error in safe environments [8] through video games, and thus it is helpful for education in cybersecurity. Education in cybersecurity through a video game has been shown as effective [11], [12].

In spite of the fact that video games are more attracting attention, according to the recent survey [5], a helpful design for video games of education in cybersecurity is still unknown [7]. To develop more effective video games as educational tools, a helpful design for video games should be investigated deeply. We take the first step in revealing a helpful design for video games of education in cybersecurity.

To reveal the game design for education in cybersecurity, we take *episodic memory* [13], where a user can memorize personally experienced events and their particular emotions, such as anxiety and pleasure, into account. More specifically, in the context of cybersecurity, it is expected that a user will feel anxiety and pleasure when he/she is harmed by attacks and can defend against the attacks by introducing their countermeasures, respectively. We believe that the above standpoint is essential for improving the awareness of cybersecurity.

We then built the following hypothesis: *a user will create awareness of cybersecurity into episodic memory by experiencing damage caused by attacks and the advantages of their countermeasures throughout the game.* We focus on *game scenarios* and *game functionalities* to realize the above hypothesis. Namely, we test if a user can learn cybersecurity with episodic memory based on game scenarios and game functionalities.

Another motivation for our work is AI security. AI has been a serious issue for cybersecurity: for instance, adversarial examples [14] which force an AI model to misinfer, privacy inference [15] to extract training samples from the model, model extraction [16] to steal the model through model's API, and backdoors [17] that are embedded into the model as vulnerabilities. These attacks provide many attack surfaces with an adversary against AI models, and many real-world threats have been shown until now [18], [19], [20], [21], [22]. Since AI models have also been deployed into various services, these threats will become serious issues in the near future. Nevertheless, to the best of our knowledge, there is no educational tool for AI security. To mitigate threats in the real world, the education of users is crucial, and hence we are motivated to develop an education tool for AI security.

To this end, we propose REN-A.I.,[1] a video game that enables a user to learn AI security through the game scenarios and the game functionalities. REN-A.I. is a single-player game, and it is considered that implementing it in a real-world setting as an educational tool is potentially easy. As described above, although it is vital to protect AI from attacks as a part of cybersecurity [23], [24], [25], to the best of the authors' knowledge, no game about AI security has been proposed so far. Namely, REN-A.I. is the first video game and also the first educational tool for AI security. We designed REN-A.I. as a dating adventure for entertainment in order to maintain users' motivation. The design of REN-A.I. is also based on several learning principles [26], [27]. In REN-A.I., to obtain knowledge of AI security, a user can simulate damage caused by attacks on AI through the game scenarios and can resist them by their countermeasures through the game functionalities. (See Section III for more details.) REN-A.I. is publicly available (https://www-infosec.ist.osaka-u.ac.jp/software/ren-ai/REN-AI(EN).html).

We also conducted a questionnaire survey with 48 users and analyzed its results to test whether the above hypothesis is true. Specifically, we give research questions in the survey with respect to game scenarios and game functionalities and evaluate the survey results from 48 users in a statistical way. We then confirm, through statistical testing, that game scenarios and game functionalities are effective for learning cybersecurity. Specifically, we demonstrate that users who read game scenarios in REN-A.I. can learn AI security more effectively through the game than other users. We also demonstrate the functionality for accuracy deterioration due to attacks in REN-A.I. is effective even for users who do not read the game scenario. (See Section V for detail.) Finding this insight is our novelty and contribution as well as the design of REN-A.I.

To sum up, we make the following contributions:

- We propose a video game, REN-A.I., for learning cybersecurity by leveraging episodic memory.
- REN-A.I. is publicly available.
- We confirm that game scenarios and game functionalities are effective for learning cybersecurity through a questionnaire survey with 48 users. For instance, 74% of users consider game scenarios effective, while 81% of users consider game functionalitie effective.
- We demonstrate that users who read game scenarios in REN-A.I. can learn AI security through the game more effectively than the other users. We also demonstrate that the functionality for accuracy deterioration due to attacks in REN-A.I. is effective even for users who do not read the game scenario.

[1] "REN-A.I." is pronounced "ren-ai", which means romance in Japanese.

1) PAPER ORGANIZATION

The remaining parts of this paper are organized as follows. In Section II, we present backgrounds and related works on games for education in cybersecurity and episodic memory. In Section III, we describe the design of REN-A.I., including the game scenarios and the game functionalities, and present the learning principles we applied in designing the game. In Section IV, we present the user study, i.e., research questions and our questionnaire survey. In Section V, we discuss the game scenarios and the game functionalities based on the questionnaire survey and threats to the validity of our results. Finally, we conclude in Section VI.

II. RELATED WORK

In this section, we describe related works in terms of the design of games for education in cybersecurity and episodic memory.

A. DESIGN OF GAMES FOR CYBERSECURITY

Learning by video games has various benefits [8]: for instance, the motivation for learning is maintained; assignments can be given easily in accordance with the learner's understanding; and safe environments to enable users to learn by trial and error can be constructed. In the past decades, there are many games to learn cybersecurity (also known as serious game) [6], [28], [29], [30], [31], [32], [33], [34], [35], [36]. Users can learn how to attack and exploit vulnerabilities in a dynamic setting and how to react to the attacks by developing their countermeasures through games [6]. It means users' fast learning of cybersecurity [37].

There are two kinds of games for education in cybersecurity [5], i.e., table game [32], [40], [41], [42] and video game [28], [29], [31], [38], [43], [44], [46]. Whereas users can communicate with each other in the table game, they can concentrate on learning alone in the video game [8]. As for the video game types, web-based games are more widely used than computer games and mobile games. Users can access a web-based game instantly because users can play web-based games without installing, unlike computer games and mobile games [5].

In this paper, we focus on video games. As described in the previous section, a video game can be introduced as a learning tool for a wide generation of a user, from children to adults, and hence many games for learning security have been proposed [5]. However, it is still unclear about effective elements for security education [5], [7], except for expression in games [47]. Among the video games, Saeet [46] discussed game scenarios such that users can apply security concepts. However, it did not conduct any survey with user study. We shed light on this standpoint in this paper. Menelaos et al. [48] proposed a framework for the design and implementation of cybersecurity serious games, but they did not consider effective elements for security education. We aim to clarify such elements through the design of REN-A.I. in this paper.

We show features of existing games for education on security in Table 1.[2] There are three novelties in REN-A.I.: designed for learning AI security, including some romance stories for the dating adventure, and designed by the Gagné's nine events of instruction and the Merrill's principles of instruction. The design of a video game for education should be dependent on its learning topic [49]. Since AI security deals with attack methods based on AI architectures, it is different from existing security education methods. In general, training and inference of AI is a black box for both a model owner and an adversary, and hence, updating an AI with any specific method is difficult [50]. Due to the above unexplainable reason, education in AI security is quite different from education in other cybersecurity.

Meanwhile, games for education in cybersecurity need hooks, excluding contents of cybersecurity, to attract users' attention although existing games only include the contents of cybersecurity. We designed REN-A.I. as a dating adventure game. Dating adventure games are one of the major game genres in Japan. There is no educational game as a dating adventure game without design for the enhancement of human-communication skills [51], to the best of our knowledge. REN-A.I. was designed by two learning methods, i.e., the Gagné's nine events of instruction and the Merrill's principles of instruction. There is no game based on these learning methods, to the best of our knowledge. If we can demonstrate that the games based on these learning methods are effective for education in cybersecurity, revealing a helpful game design for education in cybersecurity is possible.

B. EPISODIC MEMORY

Memory is roughly classified into short-term memory and long-term memory, and long-term memory is classified into semantic memory and episodic memory. Semantic memory is the memory related to knowledge obtained from contents such as textbooks, while episodic memory is the memory of one's personal experiences [13]. episodic memory is strongly affected by emotions [52]. For instance, the existing experiments with rats showed that episodic memory is memorized, especially by rewards, fears, and anxiety [53]. We focus on episodic memory as a learning principle of cybersecurity. For education in cybersecurity focusing on episodic memory, to the best of our knowledge, no video game has been proposed (see Table 1), while a video was proposed [54]. Namely, our novelty is to design a video game as an educational tool with episodic memory for cybersecurity.

C. AI SECURITY

There are four major attacks in AI security, i.e., adversarial examples [14], privacy inference [15], [55], [56], model

[2] Capture-The-Flags (CTF) are often included in video games, but we did not mention them because CTFs are difficult for non-experts. We note that, to the author's knowledge, no game about AI security has been proposed.

TABLE 1. Related works of games for education in cybersecurity.

Title	Subjects	Primary audience	Game genre	Learning method
CyberCIEGE [28] [38]	Cybersecurity	General computer users	Resource management simulation	-
Hacking Simulator [29]	Network attacks	Students	Simulation	-
Info-Sec Consultant [29]	Logical security	Students	Role-playing	-
Space Fighter [29]	Web security	Students	Action-adventure	-
KMD Puzzle [29]	Key management	Students	Puzzle	-
Be-aware [29]	Social engineering	Students	Quiz	-
Snakes and Ladders [29]	Security policy	Students	Board game	-
Securitycom [30]	Information security	Information security teams	Simulation	Experiential learning theory, and the NCW tenet of SSA
Anti-Phishing Phil [31]	Phishing	End-users	Shooter game	Reflection, story-based agent, and conceptual–procedural
Control-Alt-Hack [32]	Cybersecurity	CS/STEM students	Card game	-
sD&D [33]	Security threats	-	Card game, Quiz	-
Snakes and ladders for digital natives [34]	information security	Primary school audience	Board game	BCE principles
DFI [35]	Digital Forensics	College and high school students	Quiz, Adventure, Simulation	A logic in [39]
Happy Onlife [36]	E-safety	Children	Sugoroku, Quiz	Positive psychology gamification, meaningful learning, and ecosystemic theories
CySEC [40]	Information security	Middle school students	Card game	Educational design research
CSRAG [41]	Software security awareness	System stakeholders	Board game	Inquiry-based learning, team-based learning, and learning through playing
D-D [42]	Infrastructure security	-	Board game	-
Passworld [43]	Password	Enterprise users	RPG	Experiential gaming model, Bloom's taxonomy, and conceptual-procedural
SherLOCKED [44]	information security	Undergraduate students	Adventure	A logic in [45]
REN-A.I.	AI security	Non-experts	Quiz, Simulation, Dating Adventure	the Gagné's nine events of instruction, and the Merrill's principles of instruction

extraction [16], and backdoors [17]. Adversarial examples are the most popular attack where an AI model is confused by generating perturbed inputs, and self-driving technology is affected significantly by adversarial examples [57]. Privacy inference is a well-known problem where the privacy of a data owner is revealed through the use of an AI model because the AI model memorizes most of its training data [58]. Next, model extraction is an attack where an adversary obtains a clone of the target model by sending queries and receiving their results. The model obtained by the adversary can be used in generating adversarial examples [59] or recovering training data [60]. Finally, backdoors (also known as poisoning/trojans) are a kind of attack where an adversary embeds vulnerabilities into an AI model through its training phase. Recent backdoors are invisible for human [61], [62], and backdoors can also be used in privacy inference [63], [64], [65]. The design of an educational tool, including the above attacks, is quite important since AI has been used widely. However, there is no video game as an educational tool regardless of expecting video games for education in both cybersecurity and AI [12] (see Table 1). Although there is a Capture-The-Flag (CTF) [66] as an educational tool to learn how to attack AI models, it is limited to experts who have enough knowledge for AI security. Namely, our goal is to design the first educational tool for AI security, especially for non-experts.

III. DESIGN OF REN-A.I.

In this section, we describe the design of REN-A.I., especially game scenarios and game functionalities. In this paper, we define a game scenario as the story of a game, including interaction with game characters, and game functionality as game mechanisms, including complete conditions. REN-A.I. is a simulation game, i.e., dating simulation and life simulation. The main concept of REN-A.I. is to leverage game scenarios and game functionalities to instantiate episodic memory. We implemented REN-A.I. with TyranoBuilder.[3]

A. REQUIREMENT

Our target is non-experts who want to learn AI security, such as undergraduate students. Our goal of education is that a user can be aware of AI security. In the Gagné's five types of learning outcomes [26], which classify the learning outcomes with teaching methods for each type, the above goal is identical to "attitude." The attitude is an internal state that can affect the learner's action. Learners should understand the reason and train with ideal movements for the attitude. Consequently, we set the goal of REN-A.I. as follows: a user can understand attacks on AI and their countermeasures in addition to the necessity of AI security.

[3] https://b.tyrano.jp/

Specifically, in REN-A.I., a user can learn the following four points to achieve the goal: (1) concept of AI; (2) major attacks on AI and their countermeasures; (3) damage caused by the attacks on AI; and (4) adversary's advantages by the attacks. These points are important because of the following reasons. First, if a user wants to understand AI security, he/she should also understand the concept of AI, including its architectures, because attacks on AI often depend on it [24], [25], as well as kinds of attacks on AI and their countermeasures. Moreover, to grasp the necessity of AI security, it is important for a user to understand the damages caused by attacks on AI and the adversary's advantages. We show the contents, where users can learn by REN-A.I., in Table 9 shown in Appendix VI.

We note that programming and formulas for machine learning theory are out of the scope of REN-A.I. because our target is non-experts as described above. For the sake of simplicity, a user can learn only deep learning as AI in REN-A.I.. We also note that REN-A.I. is a single-player game, i.e., a user can learn by playing REN-A.I. by him-/herself. It can also be utilized for a large group of users with environmental adjustments, e.g., allowing each user to use an individual terminal or facilitating the observation of one user by other users.

B. OVERVIEW OF REN-A.I.

We first describe the overview of REN-A.I.. The main concept of REN-A.I. is that a user can learn AI security by memorizing it into episodic memory. First, a user needs to improve AI in REN-A.I., and it will deteriorate due to attacks on AI itself in the game. Hence, the user also needs to introduce their countermeasures. The user can then understand AI and AI security deeply by experiencing them as episodic memory with emotions, pleasure, and anxiety. Figure 1 shows the flow of the game scenario in REN-A.I.

To realize the above concept, in REN-A.I., we focus on game scenarios and game functionalities. The user feels anxiety due to attacks on AI through the game scenarios because he/she does not know their countermeasures against these attacks. Likewise, the user feels pleasure from the game functionality when he/she can defend against the attacks by introducing countermeasures.

As the learning principles, we introduce the Gagné's nine events of instruction [67] and the Merrill's principles of instruction [27] for realizing our concept. More specifically, we apply the Merrill's principles of instruction for designing REN-A.I.; that is, a user first learns background knowledge about AI architectures and then learns AI security. To deeply understand attacks on AI and their countermeasures, the user also needs to understand AI architectures as described above. Next, the Gagné's nine events of instruction have four steps, i.e., preparation, presentation, practice, and retention, to support the user's learning. In REN-A.I., incidents about AI security are used to gain attention as the preparation step. The scenarios and the game functionalities provide knowledge of AI security as the presentation step, and hence the user can learn from them. The game functionalities help a user to try ideal movements as the practice step and then he/she can memorize them. The retention step of learning is given rise from episodic memory.

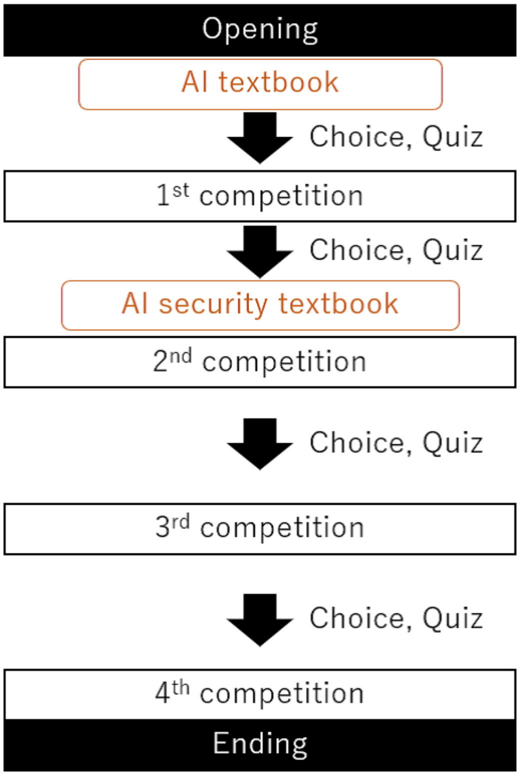


FIGURE 1. Flow of the game scenario in REN-A.I.. The black arrow means when a user can operate in the game. The orange parts mean when a user reads the textbook for the first time.

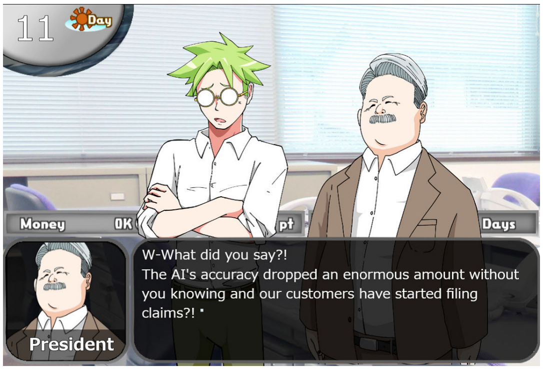
C. GAME SCENARIO

In REN-A.I., a user improves AI as a new employee of the company in the game by interacting with the game characters. AI deteriorates due to attacks, and the game characters are beginners to AI security, so they need to introduce countermeasures against attacks through the investigation of unknown phenomena. The user will be impressed with AI security and learn the necessity of AI security by experiencing this scenario.

Figure 2 shows a screenshot of REN-A.I., where the accuracy of AI in the user's company deteriorates due to an attack on it in the game. Through this experiment, the user can understand why AI security is necessary. It is identical to the gaining attention in the Gagné's nine events shown in Table 2.

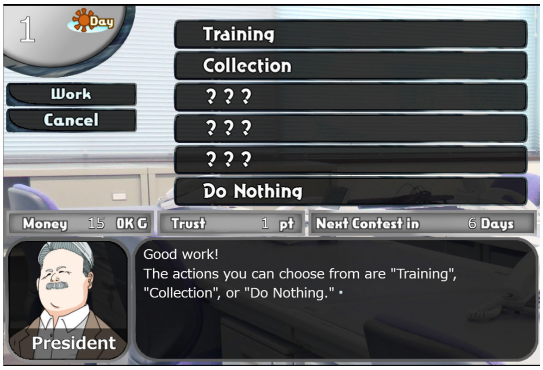
D. GAME FUNCTIONALITY

In REN-A.I., a user learns AI security through the game functionalities, i.e., *Training*, *Attack rivals*, *Receive an attack*, *Take countermeasures*, *Textbook*, *Quiz*. (Hereafter, we denote



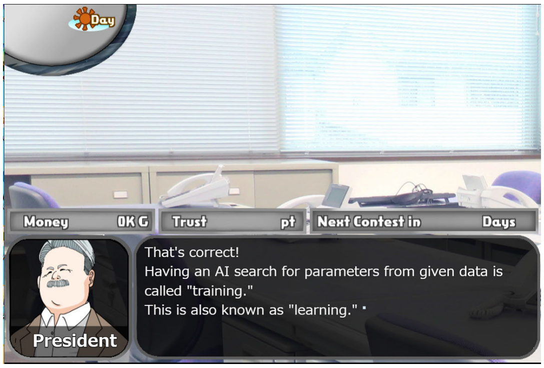
W-What did you say?!
The AI's accuracy dropped an enormous amount without you knowing and our customers have started filing claims?!

FIGURE 2. A screenshot of REN-A.I.: This figure shows that AI of the user's company is attacked in the game. It is located in step 1 of Gagné's nine events of instruction.



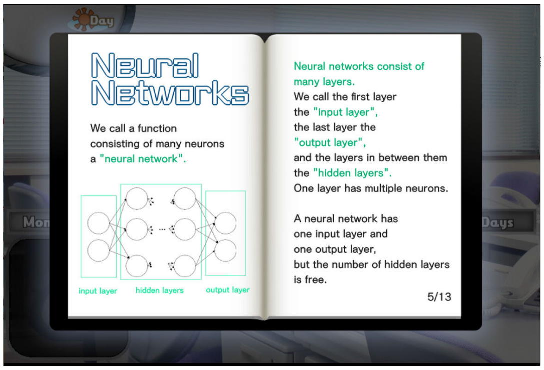
Good work!
The actions you can choose from are "Training", "Collection", or "Do Nothing."

FIGURE 3. A screenshot of REN-A.I. when a user chooses the functions. It is identical to the eliciting performance of the Gagné's nine events of instruction.



That's correct!
Having an AI search for parameters from given data is called "training."
This is also known as "learning."

FIGURE 4. A screenshot of REN-A.I. when a user is explained by a game character. It is located in step 7 of the Gagné's nine events of instruction.



Neural Networks

We call a function consisting of many neurons a "neural network".

Neural networks consist of many layers.
We call the first layer the "input layer",
the last layer the "output layer",
and the layers in between them the "hidden layers".
One layer has multiple neurons.

A neural network has one input layer and one output layer,
but the number of hidden layers is free.

5/13

FIGURE 5. A screenshot of REN-A.I. when a user reads *Textbook*. It is located in the fourth and seventh events of Gagné's nine events of instruction.

by the italic fonts the functions operable for a user in the game.) The user makes a choice, including *Training* and *Attacking rivals*, for playing REN-A.I.. The user improves AI by choosing them appropriately in accordance with a limited number of choices as shown in Figure 3. The user may obtain unexpected effects because the effects are probabilistic.

The user can learn the basic method to improve AI by *Training* and can learn attacks and their adversary's merits by *Attack rivals*. Since performing attacks is often unethical, the user who uses *Attack rivals* is required to read a scenario for the ethics.

AI in REN-A.I. is often attacked, and hence, several negative effects are caused, e.g., accuracy deterioration of AI. The user can reduce the probability of damage by introducing appropriate countermeasures. The user has to decide the appropriate countermeasures because they vary for each attack. When the user succeeds in protecting AI with countermeasures, he/she can learn each attack and its corresponding countermeasure.

To complete the game, the user should not only increase accuracy but also correct answers to quizzes about AI security, and then they can learn in semantic memory through quizzes. The game characters also support learning by providing quizzes randomly. Moreover, the user will learn basic knowledge about AI security by reading *Textbook* in the game. If the user answers a quiz incorrectly, the corresponding page of *Textbook* is displayed and game characters explain it, and then the user can obtain the knowledge correctly. *Textbook* can be referred to at any time while playing the game.

The actual screenshots of the game are shown in Figure 4, where the user is explained by the game characters. *Textbook* is shown in Figure 5.

E. LEARNING PRINCIPLES

We apply the Gagné's nine events of instruction [26] and the Merrill's principles of instruction [27], which are instructional design models to designing REN-A.I.. We describe these models below.

Table 2 shows the Gagné's nine events of instruction and how to apply them to REN-A.I.. Gagné proposed the nine events of instruction in order to address the conditions of learning.

The Merrill's principles of instruction are another instructional theory based on a broad review of many instructional models and theories. While the Gagné's nine events of instruction support a single learning objective, the Merrill's principles of instruction support multiple learning objectives. We designed REN-A.I. by using three principles in the

TABLE 2. Gagné's nine events of instruction. The left column is the original instructional events and the right column is how to apply it in REN-A.I.

Instructional Event	REN-A.I.
1. Gaining attention	Scenario
2. Informing learner of objectives	Scenario
3. Stimulate recall of prior learning	Scenario
4. Presenting stimulus	Scenario, *Textbook*
5. Providing learning guidance	Scenario, *Textbook*
6. Eliciting performance	*Quiz*, *Training*, *Receive an attack*, *Attack rivals*, *Take countermeasures*
7. Providing feedback	*Quiz*, *Training*, *Receive an attack*, *Attack rivals*, *Take countermeasures*
8. Assessing performance	Results of REN-A.I.
9. Enhancing retention and transfer	episodic memory

Merrill's principles of instruction, e.g., problem-centered, activation, and demonstration (see Table 3).

More specifically, we apply the above learning principles as follows. Taking an arbitrary choice, including *Take countermeasures* and *Quiz*, are identical to the sixth event, the eliciting performance, in the Gagné's nine events of instruction. The user can learn the essence of AI security by performing ideal actions repeatedly. Meanwhile, the following terms are identical to the seventh event, providing feedback, in the Gagné's nine events of instruction and the application in the Merrill's principles of instruction: effects by user's choice; a message displayed when a user succeeds in defense; game characters' explanations; and the corresponding page of *Textbook* is displayed when the user answers a quiz incorrectly.

Moreover, the game characters explain concrete examples after their explanations about general knowledge when the user answers a quiz incorrectly. It is identical to the demonstration in the Merrill's principles of instruction. To complete REN-A.I., the user is required to obtain an AI model with high accuracy and to answer the quizzes correctly. They are identical to the eighth event, the assessing performance, in the Gagné's nine events of instruction. We also leverage episodic memory for enhancing retention and transfer, which is identical to the ninth event of the Gagné's nine events of instruction.

IV. USER STUDY

In this section, we conduct a user study to evaluate REN-A.I.. We first describe the research questions in this paper and the study design of a questionnaire survey for answering the questions below. We then show the results of the user study.

A. RESEARCH QUESTIONS

We recall the hypothesis described in Section I: a user will create awareness of cybersecurity into episodic memory by experiencing damage caused by attacks and the advantages of their countermeasures throughout the game. We verify if it is true from two standpoints, i.e., game scenarios and game functionalities.

To this end, we try to tackle the following research questions: (RQ1) For the game scenarios, can a user learn the necessity of AI security and the damage through the scenarios related to AI security? (RQ2) For the game functionalities, can a user learn about AI security through two functionalities, attacks on others and countermeasures against the attacks? In particular, we reveal whether a user can learn (a) the damage caused by the attacks and (b) their countermeasures.

B. STUDY DESIGN

We conducted a questionnaire survey to answer the research questions described in Section IV-A.

In particular, we conducted the survey with SmartAnswer,[4] which is a questionnaire tool based on a smartphone application, and gathered responses from 48 non-experts. All of these non-experts are gathered in Japan, the authors' country. The details of these non-experts are eleven 1st/2nd-year undergraduate students of information science and 37 (under)graduate students of other areas. Their precise gender and age are unknown due to the anonymization of the system specification. We then computed the Mahalanobis' distance between users' responses in order to exclude outliers from these data. When we tried to exclude users whose median of the Mahalanobis' distance of all the responses is over 3, only one user was removed from the data.

We show the details of the questionnaire survey in Table 4, including questions, the degrees, and their average scores, where Table 5 describes the classification of the questions and their purpose. Prior to the questionnaire survey, we first screened users' traits in the survey: specifically, we asked the users about their gender, age, grade, major, and whether they like a simulation video game. We take the questionnaire survey from five perspectives as shown in Table 5, i.e., (i) confirmation of users' understanding, (ii) the impact of game scenario on users' understanding, (iii) the impact of game functionality on users' understanding, (iv) usability for obstacles to learning, and (v) overall rating of REN-A.I.

The response in the questionnaire is given by the five Likert scale. Specifically, a user chooses five levels from "1: I agree very much" to "5: I don't think so at all." For Q13-15, they can choose five levels from "1: too much" to "5: too little," where "3: appropriate" is the most ideal answer. For Q26, a user chooses five levels from "1: Extremely satisfied" to "5: Extremely dissatisfied." While they can choose multiple answers for Q9 and either "Yes" "or" No for Q12, we only described the questionnaire in Table 4.

We describe Q9 and Q20, which are independent of the five Lickert scale, below. The degrees for Q9 are as follows: 12 for *Textbook*, 27 for *Quiz*, 15 for *Receive an attack*, 18 for *take countermeasures*, 4 for *Attack rivals*, and 10 for *Training*. From the degrees for Q20, we can measure the number of

[4] https://smartanswer.folium-research.jp/

TABLE 3. The Merrill's principles of ID. The left columns are the original principles and the right two columns are how to apply it in REN-A.I.

Principles	How to apply	Part of REN-A.I.
Problem-centred	Provide a worked example of the task that learners will complete.	Scenario
	Ensure learners are engaged at the problem and task levels, as well as the operation or action level.	Scenario
	Begin with a basic problem then build the complexity to scaffold learning.	Scenario
Activation	Tap into learners' existing knowledge and experiences.	*Quiz*
	Ensure tasks are engaging, interesting, and authentic.	-
	Begin with a basic problem then build the complexity to scaffold learning.	Scenario
Demonstration	Provide content with demonstrations and examples that reflect the learning outcomes.	*Quiz*, *Textbook*
	Provide multiple representations of ideas, concepts, and perspectives.	-
	Ensure media supports effective learning.	*Textbook*
Application	Align practice activities with learning outcomes.	Goal
	Diminishing Coaching: Gradually withdraw coaching to build learner independence.	-
	Provide opportunities for learners to apply their learning to different contexts	-
Integration	Provide opportunities for learners to demonstrate and share their learning.	-
	Include reflection activities to recognize progress.	-
	Encourage learners to transfer their learning to their own lives.	-

users who completed REN-A.I. as follows: 2 for 0-25%, 24 for 25-50%, 8 for 50-75%, and 13 for 75-100%. Thirteen users answered to play until the ending of REN-A.I. shown in Figure 1.

C. RESULTS

We analyze the questionnaire survey results and the degrees shown in Table 4. We define a ratio of users who answer "1" or "2" in the questionnaire as a two-top rate and a ratio of users who answer "4" or "5" as a two-bottom rate.

1) USER CONFIDENCE RATING

We identify whether the game scenario is helpful to learn AI security through answers to (ii). The game scenario is considered beneficial for understanding the technical details of AI security since the two-top rates for Q5 and Q6 are 72% and 74%, respectively. However, we could not confirm whether a user can learn the necessity of AI security in the real world because we ask the user only whether he/she can learn the knowledge of attacks and their countermeasures.

Next, we identify whether the game functionalities are helpful to learning AI security through answers to (iii). It is considered that using the game functionalities with respect to attacks on AI and their caused damage is beneficial for understanding the necessity of AI security since the two-top rate for Q7 is 81%. Likewise, using *Training* is beneficial for understanding the technical details of AI since the two-top rate for Q8 is 77%. Using *Quiz* is beneficial for learning since the two-top rate for Q10 is 88%. Therefore, we found that the game scenarios and the game functionalities are helpful for learning AI security for the results.

2) USER PERFORMANCE

We analyze factors of REN-A.I. in order to understand how users play REN-A.I. to learn AI security. The answers of Q20 show that 44% of users played half of the entire scenario in REN-A.I., and 27% of users completed it. We find that more than half of users read *Textbook* when it is shown at first because the two-top rate for Q11 is 55%. Likewise, one-over-three users do not read it because the two-bottom rate is 34%. We also found that two-over-three users read the scenario because the two-top rate for Q4 is 66%. Likewise, one-over-three users do not read it because the two-bottom rate for Q4 is 28%.

We identify whether there are any obstacles to learning. We consider that there is no obstacle in visibility and operability since the two-bottom rates for Q13 and Q14 are 8% and 10%. It is considered that the quizzes might be slightly difficult or appropriate because the rate of users who selected "appropriate" in Q15 is 62%, and those who selected "difficult" in Q15 is 34%, respectively. Meanwhile, the frequency of appropriate quizzes for Q16, the number of quizzes for Q16, and the time required to play the game for Q17 might be appropriate or slightly many: because the rates of users who selected "appropriate" in Q16 and Q17 are both 64% and those who selected "a bit more" in Q16 and Q17 are both 30%, respectively. We also consider that the description of the technical details of AI and AI security is reasonable to follow because the two-bottom rates for Q18 and Q19 are 4% and 2%. Consequently, it is considered that there is no obstacle to learning in REN-A.I..

V. DISCUSSION

In this section, to investigate factors of the results in the previous section for detail, we analyze the results from the standpoints of users. In particular, we discuss impacts on the user performance with respect to the flow of the game scenario played by the users in the study, reading the game scenario, and user traits such as gender. We test for the difference using the Wilcoxon's signed-rank test below, where p-value is 0.05.

A. IMPACT OF FLOW OF THE GAME SCENARIO PLAYED BY USERS

From the answers to Q20, we found several differences in the user performance. In particular, we focus on two groups of the users, i.e., ones who play more than half of the game scenario and ones who play the Ending of REN-A.I. in Figure 1.

TABLE 4. The questionnaire used in our survey. The right parts show the degrees and the average scores.

	Question	1	2	3	4	5	average
Q1	Could you learn the AI architecture?	9	27	10	1	0	2.06
Q2	Could you learn the damages from attacks on AI ?	8	29	9	1	0	2.064
Q3	Could you learn the countermeasures to attacks on AI?	5	31	7	4	0	2.212
Q4	Did you read the scenario?	5	26	3	13	0	2.510
Q5	Is the scenario helpful to understand AI architecture?	2	32	12	1	0	2.255
Q6	Did the scenario is helpful to understand the attacks, countermeasures, and damages?	2	33	11	1	0	2.234
Q7	Was coming under attack helpful to understand the necessity of countermeasures AI security?	12	26	7	2	0	1.979
Q8	Were Training AI and increasing accuracy helpful to understanding AI architecture?	12	24	10	1	0	2.000
Q9	Which were helpful to learn, *Textbook*, *Quiz*, *Receive an attack*, *Take countermeasures*, *Attack rivals* and *Training*? (Multiple choices allowed)	-	-	-	-	-	
Q10	Did the quiz help you learn?	14	25	8	0	0	1.872
Q11	Did you read *Textbook* when displayed first?	4	22	5	13	3	2.765
Q12	Did you use the "read the textbook" command other than the tutorial? (Yes/No)	-	-	-	-	-	-
Q13	Were the game screens easy to view?	9	25	9	3	1	2.191
Q14	Did you operate the game easily?	6	25	11	4	1	2.340
Q15	Were the quizzes appropriate difficulty?	1	16	29	1	0	2.638
Q16	Were the frequency and number of quizzes appropriate?	1	14	30	2	0	2.702
Q17	Was the time required to play the game appropriate?	2	14	30	1	0	2.638
Q18	Were the descriptions about AI easy?	4	33	8	2	0	2.170
Q19	Were the descriptions about AI security easy?	7	31	8	1	0	2.063
Q20	Which part of the game have you played?	-	-	-	-	-	-
Q21	Did you think of playing the game to see further scenarios, or to see the ending?	5	25	10	6	1	2.458
Q22	Did you want to play REN-A.I. because it is a video game?	7	22	14	4	0	2.354
Q23	Did you think of answering the quiz correctly, because you can get money by answering the quiz correctly in the game?	9	26	7	4	1	2.188
Q24	Did you want to play this game repeatedly?	3	20	14	9	1	2.708
Q25	Do you want to recommend this game to your acquaintances when they study AI security?	6	20	14	6	1	2.489
Q26	What is your overall satisfaction with the game?	5	33	6	2	1	2.170

TABLE 5. Classification and propose of each question.

	Perspective of questions	Purpose of questions
Q1-Q3	(i) Confirmation of understanding	User confidence
Q4-Q6	(ii) Game scenario	Testing hypothesis
Q7-Q12	(iii) Game functionality	Testing hypothesis
Q13-Q20	(iv) Usability	User performance
Q21-Q24	(v) Motivation	User performance
Q25-Q26	(vi) Overall rating	User performance

TABLE 6. Difference between the users who played the Ending in Figure 1 and the other users. For Q20, the users in the former case chose "1," "2," or "3," and those in the latter case are "4", respectively. Each score in this table represents the average score of the users' choice in each case except for Q9. For Q9, it represents the rate of the users who chose "Quiz" in each case.

	Average or rate		
Question number	Played the ending	The other users	p-value
Q4	2.70	2.00	0.034
Q7	2.117	1.613	0.046
Q9 (Quiz)	84.6%	47.0%	0.021
Q10	2.02	1.46	0.010
Q14	2.14	2.84	0.025
Q24	2.5	3.153	0.046

1) PLAYING MORE THAN HALF OF THE GAME SCENARIO

According to the results described in Section IV-C2, 44% of the users played half of the entire scenario in REN-A.I.. We discuss whether there is a difference between users who played more than half of the game scenario in Figure 1 and the other users, i.e., users whose choices are "1" or "2" for Q20 and those whose choices are "3," "4," or "5" for Q20.

We then found the significant difference for Q14 and Q24 as follows, where the following three scores represent the average score of users whose choices are "1" or "2" for Q20, that of users whose choices are "3", "4," or "5" for Q20, and p-value, respectively: for Q14, the scores are 2.07, 2.66, and p=0.019; and for Q24, the scores are 2.38, 3.04, and p=0.019.

These scores indicate that there is no difference in the user confidence between the above two kinds of users. Meanwhile, the reason why the users whose choices are "3," "4," or "5" gave low scores for the operability, i.e., Q14, is due to the game functionality for AI security, which becomes playable after the 2nd competition in Figure 1.

2) PLAYING THE ENDING OF REN-A.I.

Next, we discuss whether there is a difference between users who played the Ending in Figure 1 and the other users, i.e., for Q20, users whose choices are "1," "2," or "3," and those whose choices are "4" for Q20.

We then found the significant difference for Q4, Q7, Q9, Q10, Q14 and Q24. (See Table 6.) Based on the results of Q7 and Q9, we confirm that repeatedly experiencing incidents and quizzes is effective for learning AI security. We also identify that the users who play the Ending in Figure 1 read the game scenario in more detail than the other users.

TABLE 7. Difference between the users who read the game scenario and the other users. For Q4, the users in the former case chose "1" or "2," and those in the latter case chose "3," "4," and "5," respectively. Each score in this table represents the average score of the users' choice in each case except for Q9. For Q9, it represents the rate of the users who chose "Textbook" and "Quiz" in each case, respectively.

	Average or rate		
Question number	Read the game scenario	The other users	p-value
Q1	1.80	2.56	0.001
Q2	1.77	2.62	0.001
Q3	2.00	2.625	0.006
Q5	2.03	2.68	0.001
Q6	2.10	2.50	0.024
Q8	1.70	2.56	0.001
Q9 (Textbook)	35.4%	6.3%	0.031
Q9 (Quiz)	74.2%	25%	0.001
Q10	1.709	2.187	0.019
Q11	2.54	3.18	0.045
Q18	1.97	2.56	0.003
Q19	1.90	2.38	0.020
Q20	2.93	2.18	0.011
Q21	2.12	3.00	0.025
Q23	2.00	2.50	0.037
Q26	2.00	2.50	0.009

B. IMPACT OF READING THE GAME SCENARIO

According to the results described in Section IV-C2, one-over-three users did not read the game scenario. We discuss whether there is a difference between users who read the game scenario and the other users, i.e., for Q4, users whose choices are "1" or "2," and those whose choices are "3," "4," and "5."

We then found the significant difference for Q1-3, Q5, Q6, Q8-Q11, Q18-Q21, Q23, and Q26. (see Table 7.) We believe that users who read the game scenario aim to engage more seriously in the game and, consequently, can learn AI security more effectively than the other users.

Meanwhile, there is no significant difference for Q7. Since the average score of users who read the game scenario is 1.84 while that of the other users is 2.25, it is considered that the functionality for accuracy deterioration due to attacks is effective even for users who did not read the game scenario.

The above result may give us a beneficial insight into a generalized game for security education. For instance, by protecting assets for the goal of a game and by thwarting the protection due to security incidents, a user can learn security regardless of reading a game scenario.

C. IMPACT OF USER TRAIT

We consider that user traits might have an impact on the results in the previous section. To shed light on the impact of user traits, we analyze the results by separating the users into two kinds of groups with respect to genders, area of expertise, and game genre preferences below.

1) IMPACT OF GENDERS

First, we discuss whether there is a difference in genders, i.e., male or female. There were 26 male users and 21 female users in our user study.

TABLE 8. Difference between the users who prefer simulation game, i.e., dating simulation, and training simulation and the other users. Each score in this table represents the average score of the users' choice except for Q12. For Q12, it represents the rate of the users who chose "Yes" among the former users and those among the latter users, respectively.

	Average or rate		
Question number	Prefer	The other users	p-value
Q5	2.07	2.78	0.004
Q6	2.07	2.78	0.008
Q12	67.0%	0.33%	0.018
Q13	1.96	2.78	0.039
Q18	2.03	2.44	0.038
Q21	2.21	3.22	0.008
Q24	2.34	3.44	0.001
Q25	2.14	3.22	0.001
Q26	2.00	2.77	0.016

We then found a significant difference between the genders for Q9. For instance, 42.3% of the male users and 76.1% of the female users selected "Quiz" in Q9, and then the p-value is 0.02. Besides, 46.1% of the male users and 14.2% of the female users selected "Receive an attack" in Q9, and the p-value is 0.021. We feel strongly that there are differences between the genders in learning AI security because 14.2% of the female users selected "Countermeasures against Attacks" as effective for learning AI security, even though 76.1% of the female users selected "Quiz" as effective. The above result is in line with several existing works [68], [69] in the context that female users are more attracted to immersive and relaxed games, such as quizzes rather than competition.

2) IMPACT OF AREA OF EXPERTISE

Next, we discuss whether there is a difference between the area of expertise of the users, i.e., users who have expertise in information science or not.

We then found a significant difference for Q8. Specifically, the average score of users who have expertise in information science is 1.55 and the other users are 2.11, and then the p-value is 0.02. However, we could not find any significant difference in the user confidence rating. The above result is in line with several existing works [70] in the context that there is no difference in assessing the effectiveness of video games between IT students and non-IT students despite that the IT students are more exposed to technology. Meanwhile, We consider that the helpful functionalities differ from users since they have different mental models because the average score of Q8 differ.

3) IMPACT OF GAME GENRE PREFERENCE

We discuss whether there is a difference between users who prefer simulation games (i.e., dating simulation and training simulation) or not. We then found a significant difference for Q5, Q6, Q12, Q18, Q21, Q24-26. (see Table 8.) Playing many times helps enhance learning effects. It is important to design video games for education in cybersecurity in various genres because users who prefer simulation games answered that they want to play REN-A.I. repeatedly.

D. THREATS TO VALIDITY

In this section, we describe two threats to the validity of our results. First, there might exist users who may only give priority to answering the questionnaire due to the specifications of the underlying system of our survey. Specifically, according to the results, only thirteen users among 48 users completed the scenarios and three users among them did not read the scenarios. We need to conduct statistically stable surveys with a larger number of users to overcome the above concern in the future.

Second, we conducted a questionnaire survey with the five Likert scale, and most of the users are limited to Japanese. Japanese often chooses the middle option [71], which is identical to "3" in Table 4. Further studies, which take the diversity of users into account, will need to be undertaken. Specifically, we plan to conduct questionnaire surveys in many other countries.

E. LIMITATION

We have five limitations in our results and describe them below. Solving these limitations is in future work.

First, our questionnaire survey focused on only the confirmation of understanding for users. Some users might not understand AI security accurately in contrast to their answers. It will be more useful if more objective measures of learning outcomes, such as game assessments of knowledge, are investigated. Consequently, we are in the process of investigating learning outcomes, i.e., whether a user can accurately learn knowledge of AI security through examinations after playing REN-A.I.

Second, our questionnaire survey did not investigate the long-term retention of knowledge. Confirming the long-term retention of acquired knowledge is important for further demonstrating the effectiveness of the game.

Third, although we designed REN-A.I. and conducted a questionnaire survey, we did not implement it in an actual educational environment in the real world. We plan to implement REN-A.I. in an actual educational environment for further study.

Fourth, we assume that users initiate learning of their own accord, and we do not consider the presence of a different educator. In cases where the user and educator are different entities, such as in a classroom setting, considerations regarding the user experience will be addressed as a future task.

Fifth, although we described the novelty of REN-A.I. in Section II-A, it is desirable to compare its learning effectiveness with those of the existing games. The design of a video game for education should be dependent on its learning topic as described in Section II-A, so converting the existing games into games for AI security may be insufficient. We plan to conduct further research on a helpful design of games for each topic.

Finally, unfortunately, we were unable to investigate the significant relationships between learning cybersecurity and episodic memory. Although we confirmed that experiencing both attacks and countermeasures through game scenarios and game functionalities is helpful for learning security, we have not conducted a precise investigation to confirm whether it is by virtue of episodic memory. Whereas we could find plausible results regarding episodic memory, further studies that take the detailed analysis of episodic memory into account will need to be undertaken.

VI. CONCLUSION

In this paper, we proposed REN-A.I., a video game as an educational tool for AI security. REN-A.I. is publicly available (https://www-infosec.ist.osaka-u.ac.jp/software/ren-ai/REN-AI(EN).html). For the design of REN-A.I., we focused on episodic memory for education in cybersecurity and utilized game scenarios and game functionalities to instantiate episodic memory. Moreover, we conducted a questionnaire survey with 48 users who played REN-A.I.. From the fact that 74% of users consider game scenarios effective and 81% of users consider game functionalities effective, we confirmed that the game scenario and the game functionalities are effective for education in AI security.

Furthermore, we analyze the impacts on the user performance for the flow of the game scenario played by users, reading the game scenario, and user traits. We also found two suggestions for beneficial design aspects through statistical testing. First, users who read game scenarios in REN-A.I. can learn AI security more effectively than the other users by playing the game. Second, the functionality for accuracy deterioration due to attacks in REN-A.I. is effective even for users who do not read the game scenario. In the future, we will conduct a further questionnaire survey to consider the diversity of users. We also plan to investigate users' knowledge after playing REN-A.I. and shed light on game scenarios and game functionalities to realize episodic memory rigorously.

CONTRIBUTION FOR EACH AUTHOR

Each author contributed to this paper as follows:

- Mine Arai mainly proposed REN-A.I. and contributed to the entire parts of this paper, including implementation, conducting a questionnaire survey, analyzing the results, and writing this paper.
- Koki Tejima, Yuya Yamada, Chihiro Kado, Rei Shimizu, and Masataka Tatsumi contributed to implementing REN-A.I.. Yuya Yamada contributed to this work when he was at Osaka University.
- Kyosuke Yamashita contributed to a part of a questionnaire survey.
- Takayuki Miura contributed to improving explanations of AI in REN-A.I..
- Naoto Yanai contributed to organizing the project, analyzing the results, and writing this paper.
- Goichiro Hanaoka contributed to the paper presentation.

TABLE 9. Example for contents.

<table>
<tr><th colspan="3">AI architecture</th></tr>
<tr><td colspan="3">· AI is an artificial reproduction of human intelligence.
· AI consists of many "neurons", and We call a function consisting of many neurons a "neural network".
· Neurons compute outputs by weighting each component of inputs and using some functions.
· We call the process in which AI finds an appropriate weight "training"
· The way to train AI is classified into three types., e.g., Supervised learning, Unsupervised learning, and Reinforcement learning.</td></tr>
<tr><th colspan="3">AI security</th></tr>
<tr><th>attack</th><th>Threat to AI</th><th>countermeasure</th></tr>
<tr><td>Data Reconstruction attacks</td><td>Leakage of training face images from face recognition systems.</td><td>Processing learning data before training AI</td></tr>
<tr><td>Model extraction attacks</td><td>Decrease in sales due to a stolen knockoff AI.</td><td>limiting the number of queries</td></tr>
<tr><td>Adversarial examples</td><td>Accidents caused by automated vehicles misrecognizing signs.</td><td>Processing inputs before inference</td></tr>
<tr><td>Poisoning attacks</td><td>The authentication of an unauthorized person against a facial recognition system.</td><td>Using reliable data</td></tr>
</table>

CODE AVAILABILITY

REN-A.I. is publicly available (https://www-infosec.ist.osaka-u.ac.jp/software/ren-ai/REN-AI(EN).html).

APPENDIX

CONTENTS OF REN-A.I.

The contents that a user can learn in REN-A.I. are shown in Table 9.

REFERENCES

[1] M. Barrett, "Framework for improving critical infrastructure cybersecurity version 1.1," NIST Cybersecur. Framework, Tech. Rep., 2018, doi: 10.6028/NIST.CSWP.04162018.

[2] Y. Hu, W. Kuang, Z. Qin, K. Li, J. Zhang, Y. Gao, W. Li, and K. Li, "Artificial intelligence security: Threats and countermeasures," *ACM Comput. Surv.*, vol. 55, no. 1, pp. 1–36, Jan. 2023.

[3] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102470.

[4] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–35, Sep. 2023.

[5] L. Zhang-Kennedy and S. Chiasson, "A systematic review of multimedia tools for cybersecurity awareness and education," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–39, Jan. 2022.

[6] S. Hart, A. Margheri, F. Paci, and V. Sassone, "Riskio: A serious game for cyber security awareness and education," *Comput. Secur.*, vol. 95, Aug. 2020, Art. no. 101827.

[7] C. Linehan, B. Kirman, S. Lawson, and G. Chan, "Practical, appropriate, empirically-validated guidelines for designing educational games," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2011, pp. 1979–1988.

[8] R. D. M. Charles, M. Reigeluth, and B. J. Beatty, *Instructional-Design Theories and Models, the Learner-Centered Paradigm of Educatio*, vol. 4. Evanston, IL, USA: Routledge, 2017.

[9] S. Gaw, E. W. Felten, and P. Fernandez-Kelly, "Secrecy, flagging, and paranoia: Adoption criteria in encrypted email," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2006, pp. 591–600.

[10] J. Wu and D. Zappala, "When is a tree really a truck? exploring mental models of encryption," in *Proc. SOUPS*, 2018, pp. 395–409.

[11] M. A. Khan, A. Merabet, S. Alkaabi, and H. E. Sayed, "Game-based learning platform to enhance cybersecurity education," *Educ. Inf. Technol.*, vol. 27, no. 4, pp. 5153–5177, May 2022.

[12] Z. Kilhoffer, Z. Zhou, F. Wang, F. Tamton, Y. Huang, P. Kim, T. Yeh, and Y. Wang, "'How technical do you get? I'm an english teacher': Teaching and learning cybersecurity and AI ethics in high school," in *Proc. IEEE S&P*, May 2023, pp. 2032–2049.

[13] E. Tulving, "What is episodic memory?" *Current directions Psychol. Sci.*, vol. 2, no. 3, pp. 67–70, 1993.

[14] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Proc. ICLR*, 2014, pp. 1–10.

[15] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart, "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *Proc. 23rd Secur. Symp. (USENIX Security)*. San Diego, CA, USA: USENIX Association, Aug. 2014, pp. 17–32.

[16] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction APIs," in *Proc. 25th USENIX Conf. Security Symp.*, Austin, TX, USA, 2016, pp. 601–618.

[17] T. Gu, K. Liu, B. Dolan-Gavitt, and S. Garg, "BadNets: Evaluating backdooring attacks on deep neural networks," *IEEE Access*, vol. 7, pp. 47230–47244, 2019.

[18] S. Li, H. Liu, T. Dong, B. Z. H. Zhao, M. Xue, H. Zhu, and J. Lu, "Hidden backdoors in human-centric language models," 2021, *arXiv:2105.00164*.

[19] M. Xue, C. He, J. Wang, and W. Liu, "Backdoors hidden in facial features: A novel invisible backdoor attack against face recognition systems," *Peer Peer Netw. Appl.*, vol. 14, no. 3, pp. 1458–1474, May 2021.

[20] C.-S. Lin, C.-Y. Hsu, P.-Y. Chen, and C.-M. Yu, "Real-world adversarial examples via makeup," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 2854–2858.

[21] N. Carlini, F. Tramèr, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. Brown, D. Song, Ú. Erlingsson, A. Oprea, and C. Raffel, "Extracting training data from large language models," in *Proc. USENIX Secur.*, 2021, pp. 2633–2650.

[22] Q. Xu, X. He, L. Lyu, L. Qu, and G. Haffari, "Student surpasses teacher: Imitation attack for black-box NLP APIs," in *Proc. COLING*, 2022, pp. 2849–2860.

[23] J. Singh, M. Wazid, A. K. Das, V. Chamola, and M. Guizani, "Machine learning security attacks and defense approaches for emerging cyber physical applications: A comprehensive survey," *Comput. Commun.*, vol. 192, pp. 316–331, Aug. 2022.

[24] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–36, Mar. 2022.

[25] G. R. Machado, E. Silva, and R. R. Goldschmidt, "Adversarial machine learning in image classification: A survey toward the Defender's perspective," *ACM Comput. Surv.*, vol. 55, no. 1, pp. 1–38, Jan. 2023.

[26] R. M. Gagn and K. Medsker, *conditions of Learning: Training Applications*. San Diego, CA, USA: Harcourt Brace College Pub., 1996.

[27] M. D. Merrill, "First principles of instruction," *Educ. Technol. Res. Develop.*, vol. 50, no. 3, pp. 43–59, Sep. 2002.

[28] C. E. Irvine, M. F. Thompson, and K. Allen, "CyberCIEGE: Gaming for information assurance," *IEEE Secur. Privacy Mag.*, vol. 3, no. 3, pp. 61–64, May 2005.

[29] M. Mostafa and O. S. Faragallah, "Development of serious games for teaching information security courses," *IEEE Access*, vol. 7, pp. 169293–169305, 2019.

[30] D. Twitchell, "SecurityCom: A multi-player game for researching and teaching information security teams," *J. Digit. Forensics, Secur. Law*, vol. 2, no. 4, pp. 9–18, 2007.

[31] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge, "Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish," in *Proc. 3rd Symp. Usable privacy Secur.*, Jul. 2007, pp. 88–99.

[32] T. Denning, A. Lerner, A. Shostack, and T. Kohno, "Control-Alt-hack: The design and evaluation of a card game for computer security awareness and education," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2013, pp. 915–928.

[33] Y. Kido, N. P. Tou, N. Yanai, and S. Shimojo, "sD&D: Design and implementation of cybersecurity educational game with highly extensible functionality," in *Proc. FICC*, vol. 1129. Cham, Switzerland: Springer, 2020, pp. 857–873.

[34] R. Reid and J. Van Niekerk, "Snakes and ladders for digital natives: Information security education for the youth," *Inf. Manage. Comput. Secur.*, vol. 22, no. 2, pp. 179–190, Jun. 2014.

[35] J. Yerby, S. Hollifield, M. Kwak, and K. Floyd, "Development of serious games for teaching digital forensics," *Issues Inf. Syst.*, vol. 15, no. 2, pp. 335–343, 2014.

[36] S. Chaudron, R. Di Gioia, M. Gemo, and K. Lagae, "Happy onlife-a video game to support mediation on internet risks and opportunities," *Commun. Papers*, vol. 4, no. 6, pp. 47–62, 2015.

[37] E. Trickel, F. Disperati, E. Gustafson, F. Kalantari, M. Mabey, N. Tiwari, Y. Safaei, A. Doupé, and G. Vigna, "Shell we play a game? CTF-as-a-service for security education," in *Proc. ASE*, 2017, pp. 1–10.

[38] M. Thompson and C. Irvine, "Active learning with the cyberciege video game," in *Proc. CSET*, 2011, pp. 1–10.

[39] B. Carrier and E. Spafford, "An event-based digital forensic investigation framework," in *Proc. Digital Forensic Res. Workshop (DFRWS)*, Jan. 2004, pp. 1-12.

[40] M. K. Thomas, A. Shyjka, S. Kumm, and R. Gjomemo, "Educational design research for the development of a collectible card game for cybersecurity learning," *J. Formative Design Learn.*, vol. 3, no. 1, pp. 27–38, Jun. 2019.

[41] A. Yasin, L. Liu, T. Li, R. Fatima, and W. Jianmin, "Improving software security awareness using a serious game," *IET Softw.*, vol. 13, no. 2, pp. 159–169, Apr. 2019.

[42] S. Frey, A. Rashid, P. Anthonysamy, M. Pinto-Albuquerque, and S. A. Naqvi, "The good, the bad and the ugly: A study of security decisions in a cyber-physical systems game," *IEEE Trans. Softw. Eng.*, vol. 45, no. 5, pp. 521–536, May 2019.

[43] G. C. Jayakrishnan, G. R. Sirigireddy, S. Vaddepalli, V. Banahatti, S. P. Lodha, and S. Pandit, "Passworld: A serious game to promote password awareness and diversity in an enterprise," in *Proc. SOUPS*, 2020, pp. 1–18.

[44] A. Jaffray, C. Finn, and J. R. C. Nurse, "Sherlocked: A detective-themed serious game for cyber security education," in *Proc. HAISA*. Cham, Switzerland: Springer, 2021, pp. 35–45.

[45] G. Tondello, H. Premsukh, and L. Nacke, "A theory of gamification principles through goal-setting theory," in *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, 2018, pp. 1118–1127.

[46] M. M. Saeed, "Designing scenarios for in-organization training using the CyberCIEGE game," *Comput. J.*, vol. 67, no. 1, pp. 338–346, Jan. 2024.

[47] S. Maqsood, C. Mekhail, and S. Chiasson, "A day in the life of jos: A web-based game to increase children's digital literacy," in *Proc. 17th ACM Conf. Interact. Design Children*, Jun. 2018, pp. 241–252.

[48] N. M. Katsantonis, I. Kotini, P. Fouliras, and I. Mavridis, "Conceptual framework for developing cyber security serious games," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2019, pp. 872–881.

[49] F. G. M. Silva, "Practical methodology for the design of educational serious games," *Information*, vol. 11, no. 1, p. 14, Dec. 2019.

[50] J. Mink, H. Benkraouda, L. Yang, A. Ciptadi, A. Ahmadzadeh, D. Votipka, and G. Wang, "Everybody's got ML, tell me what else you have: Practitioners' perception of ML-based security tools and explanations," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2023, pp. 2068–2085.

[51] A. Abbey, S. E. Pegram, J. Woerner, and R. Wegner, "Men's responses to women's sexual refusals: Development and construct validity of a virtual dating simulation of sexual aggression," *Psychol. Violence*, vol. 8, no. 1, pp. 87–99, Jan. 2018.

[52] E. Dere, B. M. Pause, and R. Pietrowsky, "Emotion and episodic memory in neuropsychiatric disorders," *Behavioural Brain Res.*, vol. 215, no. 2, pp. 162–171, Dec. 2010.

[53] E. Dere, A. Zlomuzica, J. P. Huston, and M. A. De Souza Silva, "Chapter 2.2 animal episodic memory," in *Handbook Episodic Memory* (Handbook of Behavioral Neuroscience), E. Dere, A. Easton, L. Nadel, and J. P. Huston, Eds. Amsterdam, The Netherlands: Elsevier, 2008, vol. 18, pp. 155–184.

[54] V. Garg, L. J. Camp, K. Connelly, and L. Lorenzen-Huber, "Risk communication design: Video vs. text," in *Proc. PETS*, vol. 7834. Berlin, Germany: Springer, 2012, pp. 279–298.

[55] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 3–18.

[56] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang, "Membership inference attacks on machine learning: A survey," *ACM Comput. Surv.*, vol. 54, no. 11s, pp. 1–37, Jan. 2022.

[57] M. R. Alam and C. M. Ward, "Adversarial examples in self-driving: A review of available datasets and attacks," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2022, pp. 1–6.

[58] T. Ristenpart. (2017). *Confidentiality and Privacy Threats in Machine Learning*. [Online]. Available: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/03/Ristenpart.pdf

[59] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proc. ACM Asia Conf. Comput. Commun. Secur.*, Apr. 2017, pp. 506–519.

[60] L. Lyu, X. He, F. Wu, and L. Sun, "Killing two birds with one stone: Stealing model and inferring attribute from BERT-based APIs," 2021, *arXiv:2105.10909*.

[61] Y. Li, Y. Li, B. Wu, L. Li, R. He, and S. Lyu, "Invisible backdoor attack with sample-specific triggers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16443–16452.

[62] R. Ning, J. Li, C. Xin, and H. Wu, "Invisible poison: A blackbox clean label backdoor attack to deep neural networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, May 2021, pp. 1–10.

[63] F. Tramèr, R. Shokri, A. San Joaquin, H. Le, M. Jagielski, S. Hong, and N. Carlini, "Truth serum: Poisoning machine learning models to reveal their secrets," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2022, pp. 2779–2792.

[64] S. Mahloujifar, E. Ghosh, and M. Chase, "Property inference from poisoning," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2022, pp. 1120–1137.

[65] Y. Chen, C. Shen, Y. Shen, C. Wang, and Y. Zhang, "Amplifying membership exposure via data poisoning," in *Proc. Adv. Neural Inf. Process. Syst.*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Red Hook, NY, USA: Curran Associates, 2022, pp. 29830–29844.

[66] *HackThisAI*. Accessed: Feb. 6, 2024. [Online]. Available: https://github.com/JosephTLucas/HackThisAI

[67] R. M. Gagne, W. W. Wager, K. C. Golas, J. M. Keller, and J. D. Russell, "Principles of instructional design, 5th edition," *Perform. Improvement*, vol. 44, no. 2, pp. 44–46, Feb. 2005.

[68] G. Fortes Tondello, D. Valtchanov, A. Reetz, R. R. Wehbe, R. Orji, and L. E. Nacke, "Towards a trait model of video game preferences," *Int. J. Hum.–Comput. Interact.*, vol. 34, no. 8, pp. 732–748, Aug. 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:51882952

[69] G. F. Tondello and L. E. Nacke, "Player characteristics and video game preferences," in *Proc. Annu. Symp. Comput.-Hum. Interact. Play*, Oct. 2019, pp. 365–378, doi: 10.1145/3311350.3347185.

[70] N. T. H. Giang and L. H. Cuong, "Evaluating feasibility and effectiveness of digital game-based instructional technology," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 16, no. 16, p. 4, Aug. 2021. [Online]. Available: https://online-journals.org/index.php/i-jet/article/view/23829

[71] K. Tasaki and J. Shin, "Japanese response bias: Cross-level and cross-national comparisons on response styles," *Jpn. J. Psychol.*, vol. 88, no. 1, pp. 32–42, 2017.

MINE ARAI received the B.Eng. and M.Eng. degrees from Osaka University, Japan, in 2021 and 2023, respectively, where she is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology. Her research interest includes education in information security.

KOKI TEJIMA received the B.Eng. degree in engineering science from Osaka University, Japan, in 2022, where he is currently pursuing the M.S. degree with the Graduate School of Information Science and Technology. His research interests include machine learning and cryptography.

YUYA YAMADA received the B.Eng. degree in engineering science from Osaka University, Japan, in 2022. He is currently pursuing the M.S. degree with the Graduate School of Science and Technology, Nara Institute of Science and Technology, Japan. His research interests include machine learning and information security.

TAKAYUKI MIURA received the B.S. and M.S. degrees from The University of Tokyo, Japan, in 2017 and in 2019, respectively. He is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include machine learning security and differential privacy.

KYOSUKE YAMASHITA received the B.E., M.E., and Ph.D. degrees from Kyoto University, in 2013, 2015, and 2021, respectively. From 2021 to 2022, he was a Postdoctoral Fellow with the Cryptography Platform Research Team, National Institute of Advanced Industrial Science and Technology (AIST). Currently, he is an Assistant Professor with Osaka University and a Collaborative Researcher with AIST. He received SCIS Paper Prize from IEICE, in 2019.

CHIHIRO KADO received the B.Eng. degree in engineering science from Osaka University, Japan, in 2022, where she is currently pursuing the M.S. degree with the Graduate School of Information Science and Technology. Her research interest includes blockchain.

REI SHIMIZU received the B.Eng. and M.Eng. degrees from Osaka University, in 2021 and 2023, respectively. His research interest includes blockchains.

MASATAKA TATSUMI received the B.Eng. and M.Eng. degrees from Osaka University, in 2021 and 2023, respectively. His research interest includes machine learning.

NAOTO YANAI (Member, IEEE) received the B.Eng. degree from the National Institution of Academic Degrees and University Evaluation, Japan, in 2009, and the M.S.Eng. and Dr.E. degrees from the Graduate School of Systems and Information Engineering, University of Tsukuba, Japan, in 2011 and 2014, respectively. He was an Assistant Professor with Osaka University, Japan, until 2021, where he is currently an Associate Professor. His research interest includes information security.

GOICHIRO HANAOKA (Member, IEEE) received the Graduate degree from the Department of Engineering, The University of Tokyo, in 1997, and the Ph.D. degree from The University of Tokyo, in 2002. In 2005, he joined AIST, where he is currently the Leader of the Advanced Cryptosystems Research Group, Information Technology Research Institute. He engages in the research and development for encryption and information security technologies, including the efficient design and security evaluation of public key cryptosystems. He received numerous awards, including the DoCoMo Mobile Science Award from the Mobile Communication Fund, in 2016, the Wilkes Award from the British Computer Society, in 2007, the Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE), in 2008, and the Innovative Paper Awards at the Symposium on Cryptography & Information Security (SCIS), IEICE, in 2012 and 2014.