CHAPTER 7

Providing Alternative Ethical Perspectives Through Intelligent Agents in A Serious Game for Cybersecurity Ethical Training

MUHAMMAD HASSAN ALI BAJWA; DEBORAH RICHARDS; AND PAUL FORMOSA

ABSTRACT:

This article contributes to the development of virtual agents that act as non-player characters (NPCs) that offer different ethical viewpoints to assist the player to learn about ethical decision-making. To explore this, we developed a serious game designed to train users to consider five relevant ethical principles (Autonomy, Justice, Beneficence, Non-maleficence, and Explicability) when making cybersecurity decisions. After receiving interactive training in the game about these five ethical principles, the player participates in two cybersecurity scenarios involving three intelligent agents who play the role of virtual office employees. Each virtual agent has been allocated a different combination of personality and ethical principle priorities to present different viewpoints through their dialogues with the player. These dialogues are designed to represent each agent's personality (using the Big-5 personality model of Openness, Extraversion, Emotional

Stability, Conscientiousness, and Agreeableness) and the ethical principles that they consider to be of high or low importance. The scenarios conclude with the player making a choice, followed by a reflection statement to help them review their choice. Through comparison of pre- and post-game responses to other cyber-ethical scenarios, our initial analysis with first-year cybersecurity students shows that players' understanding of ethical principles in cybersecurity improved after playing our game.

KEYWORDS:

Cybersecurity ethics, Agents, ethical principles, serious game, ethical training, Agent personality

1. INTRODUCTION:

The key concept of cybersecurity technologies is to provide a range of tools, resources, and processes to protect the data, integrity, and confidentiality of online systems and end-users (Brey, 2007; Craigen et al., 2014). As the number of cloud-based services has rapidly increased in the last decade, the need for more secure services is required to protect these systems from the malicious attacks that may occur to steal important personal and financial data. Breaches due to the human element are also increasing (Ayyagari, 2012). As reported by Dunn (2014), 93% of cybersecurity breaches were caused by human error rather than any technical issue. System designers and users pose a key vulnerability if they are not aware of the human aspects that can impact computer systems. The designers and users of cyberspace need to understand the impact of their actions and decisions.

Despite the utmost importance of ethical decision-making in the domain of cybersecurity, this vulnerability gap has received comparatively less attention from users and researchers. This gap can be alleviated by training cybersecurity professionals to consider alternative ethical perspectives while making a decision in cyberspace. These decisions can range from implementing a new control protocol to setting the organization's policies or norms. Serious games provide pedagogical effectiveness (Kianpour et

al., 2019) in training by giving hands-on practical experience in an artificial, virtual but still familiar environment.

There are many serious games available in the literature that focus on cybersecurity training (e.g. (Ferro et al., 2022; Hale et al., 2015; Jordan et al., 2011). Nevertheless, there is a gap in the development of serious games whose primary focus is to consider the ethical aspects of cybersecurity decision-making. To cover this gap, we have designed a serious game populated with non-player characters (NPCs) who are agents with ethical priorities. Agent interactions with the player provides us with the chance to provide the player with *social learning* opportunities. In social learning, the player gets a chance to observe other agents and these observations create new behaviors in the observers. As argued by Reed et al. (2010) social learning occurs through an interaction between individuals and it changes the understanding of individuals involved in the interaction. Some of the time, observers might try to imitate others. In this case, providing multiple virtual agent perspectives aims to promote player reflection which may potentially lead to a change in ethical principles.

Experiential learning requires all the participants to play a specific role, participate, interact and apply the skills to create a replica of a real-time environment (Gentry, 1990). Serious role-play games are an effective tool to provide experiential learning. Designing intelligent agents for a serious role-playing game is vital as human players may not be available all the time and they may not be well trained to perform a specific role. Thus, agents acting as non-player characters in the game mimic the role of different humans. An artificial agent in place of a human in an artificial environment provides the opportunity for trainees to see the effects of their decision before implementing them in the real world.

We have developed a serious role-playing game that aims to provide a training medium for cybersecurity professionals in which they encounter different ethical perspectives through the design of intelligent virtual agents. The agents (non-player characters) have different personality behaviors and inclinations toward different ethical principles. Those agents will present different ethical perspectives during the game. The game provides a real-time working environment to the user in which they interact with other agents. The interaction helps an individual to apply and practice

ethical and social skills. At the start of the game, the player receives training on the five ethical principles in cybersecurity. These are from the principlist approach developed by Formosa et al. (2021) and comprise: beneficence, non-maleficence, autonomy, justice, and explicability.

To provide non-player characters (NPCs) in the game that model different ethical perspectives, we assign different combinations of ethical principles that are of high or low importance to each of our three NPCs. It is widely discussed in the literature that an individual's personality influences their ethical decision-making (Craft, 2013). For this reason, we also assign different combinations of personality traits to our NPCs using the Big Five personality model (Conscientiousness, Agreeableness, Extraversion, Neuroticism, and Openness) (Roccas et al., 2002). Due to their different personalities, agents provide different ethical perspectives, and these perspectives are used to improve the ethical awareness and sensitivity of the player during the game. This study aims to evaluate whether the player's awareness of ethical principles increased after playing the game. The following research question is addressed in this article.

Research question 1: What is the influence of a serious game on the awareness of ethical principles when making decisions in a cybersecurity context?

In the following section, we provide background literature. Section 3 describes our methodology, including a description of the implemented game. Section 4 provides the results which are discussed in Section 5. The paper ends in Section 6 with conclusions, limitations, and future work.

2. BACKGROUND LITERATURE

2.1 Decision-Making in Cybersecurity

Cybersecurity breaches can be minimized by using the latest cybersecurity technologies and understanding how these technologies are used by humans. The human aspect of implementing and using these technologies cannot be ignored as human error is a major contributing factor to cyber breaches (Streeter, 2013). One way these human errors occur is when

a cybersecurity professional is not able to make a suitable decision in cyberspace intentionally or unintentionally. Ethical decision-making provides an individual with a framework to decide between right and wrong. The ethical issues in cybersecurity arise when the ethical implications of a decision are ignored in cyberspace (Formosa et al., 2021; Vallor et al., 2018)). To minimize human error, cybersecurity professionals need to be trained to consider ethics in their decision-making in cyber ethical dilemmas (Blanken-Webb et al., 2018).

Craft (2013) presented 16 individual factors that influence decision-making. From these factors, we found "Personality" to be the most influential factor as it is discussed by most of the studies. We use the most used personality model, Big Five Factor (Costa Jr & McCrae, 2008) also known as the Big Five, to illustrate personality in our virtual agents. The model provides five factors that are used to represent personality traits (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability). Openness to Experience describes a person's feeling about new changes and being open-minded. Conscientiousness describes a person's inclination towards following a plan and being self-disciplined. Extraversion describes a person's feelings about enjoying being with people, being outgoing and participating in social gathering. Agreeableness assesses a person's feeling of being generally helpful, warm, and getting along with others and is tied to a group interest. And lastly the emotional stability factor of the Big Five refers to a person's ability to be stable all or most of the time and not be easily changed emotionally.

2.2 Frameworks and Training for Ethical Decision-Making

2.2.1. Ethical Frameworks in cybersecurity

Ethical frameworks for cybersecurity provide us with the ability to analyze the ethical issues that arise in the context of cybersecurity (Loi & Christen, 2020) . For that purpose, we used a principlist framework that has been proposed for the cybersecurity domain by Formosa et al. (2021). While there are other principlist frameworks that could be applied to cybersecurity ethics (e.g. van de Poel & Christen, 2020; Loi & Christen, 2020;

Weber & Kleine, 2020; Morgan & Gordijn, 2020), we adopt this framework here because it clearly highlights the ethical issues raised by cybersecurity, builds upon previous work in this area, and avoids principle proliferation by re-using ethical principles that are widely used in cognate fields (Formosa, Wilson et al. 2021). The framework consists of five ethical principles. Those principles are beneficence (cybersecurity technologies should enhance human lives), non-maleficence (cybersecurity technologies should not be used to harm individuals' lives), justice (cybersecurity technologies should improve fairness and provide impartial access for all), autonomy (cybersecurity technologies should not limit users' choices of applications) and explicability (cybersecurity technologies should be both understandable and accountable clearly for their functioning). These principles are modeled on the five AI4People principles (Floridi et al., 2018) for ethical AI, which are in turn an extension of four well-accepted ethical principles from bio-ethics (Beauchamp & Childress, 2001). Our adopted framework provides a novel and relevant approach to understanding ethical issues in cybersecurity.

2.2.2. Serious game in cybersecurity

Serious games help to increase the effectiveness of training (De Freitas & Jarvis, 2007) by providing hands-on practical experience by replicating a real-time environment in an artificial virtual environment. According to Gino et al. (2009), reminding people about ethical behavior and/or observing others' ethical behaviour may change their own ethical behaviour, such as their honesty. Serious games can also help to achieve this goal.

There are several serious games developed to help the user to understand different aspects of cybersecurity. CounterMeasures (Jordan et al., 2011) is a text and command base game designed to teach about computer security. The game helps the user to learn and apply computer security skills through guided objectives, such as scanning a remote system given an IP address. CyberVR (Veneruso et al., 2020) is another attempt to increase user awareness of cybersecurity-related issues using virtual reality technology. CyberPhishing (Hale et al., 2015) is a serious game focused on raising awareness related to Phishing attacks. AWATO (Ferro et al., 2022) is

a role-playing serious game that teaches participants about cybersecurity vulnerabilities. They created a threat model for issues that arise due to human factors. The game focused on minimizing human error by identifying the factors influencing decision-making, including lack of knowledge, lack of resources, lack of awareness, norms, and complacency.

We see that games in cybersecurity have been created that offer a wide range of training to teach cybersecurity issues and challenges. There are also several games that have been developed for teaching ethics, such as Global Conflicts, Cooking Mama: Mama Kills Animals (Pereira et al., 2012). Global Conflicts is a series of games that focus on social awareness and ethics. Cooking Mama focuses on ethics by raising awareness about the cruelty involved in animal-based food production. Serious games are also used for corporate training (Larson, 2020) and ethics in IT design (Urquhart & Craigon, 2021). However, we did not find any game that focuses on teaching the ethical aspects of cybersecurity issues. To address this gap, we have developed a serious game designed to train users to consider the ethical aspects of decision-making in cybersecurity. The game and our evaluation of its ability to improve ethical awareness and decision-making are described in the next section.

3. METHODOLOGY

An online study called "V-Meet Cybersecurity" was conducted in week 13 of the first semester of 2022 at Macquarie University with the approval of the University's Human Research Ethics Committee. The main aim of this study was to raise participants' awareness of the ethical principles underlying common decision-making in cybersecurity contexts using a serious game. To measure the success of our game, we asked participants to respond to two cybersecurity-related scenarios: one before playing the game (pre-test) and one after playing the game (post-test). With the help of this approach, we were able to capture if the ethical knowledge of an individual increased after playing the game. The game provides a training medium where non-player characters (agents) present different ethical choices and issues through dialogue. The dialogues were created from the cues available in the literature. The player also had to respond to those agents by selecting choices that match the user's personality and ethical choice.

The study, including the game, was accessed via a survey developed in the Qualtrics research survey software.

3.1 Recruitment

The online study was conducted with undergraduate students enrolled in the course "COMP1300– Introduction to Cyber Security" at Macquarie University. Students participated during their scheduled tutorial class in the final week of the semester. Students were provided with information about the study and asked for voluntary consent to include their data for research purposes. A total of 272 students commenced the study.

3.2 Materials: V-Meet Cybersecurity ethical training Game

This is a short serious game developed to simulate a cybersecurity organization in which the player acts as Alex who is starting a new job as Lead Security Analyst at a cybersecurity firm, *Prescott and Kruger*, next week. The player (i.e. Alex) has agreed to sit in on a couple of video calls with Marielle, the CTO, and the current acting Lead Security Analyst, to meet some of the new team members and discuss some of the important decisions. The flow of the game is represented in Figure 1.

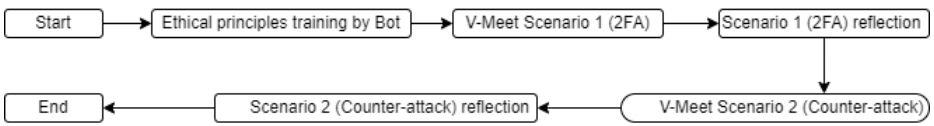

Figure 1: V-Meet Game Flow

The game starts with a conversation between Alex (the Player) and the player's assistant in a text messaging application (which can be viewed in Figure 2). The assistant guides the player through the initial training which is essential for every new employee. The ethical principles training is conducted by Ethbot (a virtual Ethical Training Bot) which guides players through learning about our five ethical principles (Beneficence, Non-Maleficence, Justice, Autonomy, Explicability) in cybersecurity. To complete

the training process, Ethbot tests the player's knowledge of the ethical principle they have learnt by providing an ethical scenario and asking the player to identify the principle that is most applicable. The player receives feedback and multiple chances to get the correct answer. After completing the ethical training, the player continues with routine official tasks in the game and takes part in the first video meeting using the V-Meet app. We discuss this in detail below.

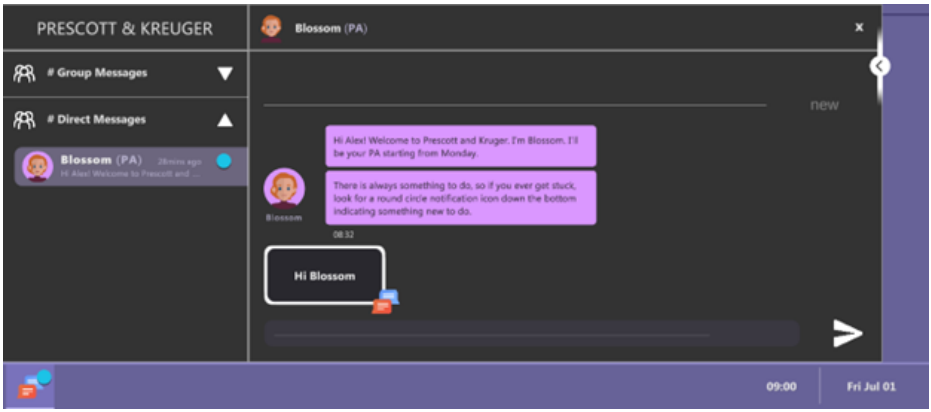

Figure 2: Text Messaging Application

3.2.1. V-Meet Scenarios and Agents

The game presents two scenarios to the player as video meetings: 2-factor authentication (2FA) and cyberattack. The 2FA scenario is presented in Figure 3. The outcome of this discussion is having to decide whether to implement the 2-factor authentication policy or not.

You are a customer of a popular cloud storage provider that provides you with an online space to save your data. The provider sends an email informing its users that in 3 weeks all users will have to nominate a mobile phone number to use with 2-factor authentication (2FA). The 2FA will use a combination of existing passwords and 1-time use codes delivered via SMS to improve the security of its authentication. The cloud storage provider updates their terms of use requiring all users to utilize 2FA or their accounts will be inaccessible for usage. Since you regularly need to access data on the cloud you face a choice. Providing a phone number will maintain access and improve the security of the service, but will also require you to share sensitive personal information (i.e., your mobile phone number) that raises privacy considerations. Not providing your Mobile phone number will avoid these privacy concerns, but result in you losing access to the shared documents from the storage provider.

Figure 3: 2FA scenario

Both the 2FA and cyberattack video meetings include three agents (non-player characters). Each agent is high or low in each of the five personality traits. Also, each agent considers each of the five ethical principles to be of low or high importance to them. One of the NPCs is Shiva who is a Coder/ Penetration Tester at Prescott & Kreuger. She is passionate about her work; her single-minded focus means she will often rush to complete tasks without planning. Though perfectly cheerful and open with her family, Shiva is a lot more distant with coworkers. Shiva's personality traits and ethical influences are provided below. Her avatar, personality scores, and ethical influences can be seen in Figure 4. Each character has a set of numerical values representing ethical principles. Ethical principle values range from -2 to +2. The characters are designed keeping in mind that each character should be strong (positive and negative) on some principles and relatively neutral on some principles. The same approach is used to define an agent's personality. Ranging from 0-100, a high and low score in any personality trait shows that the agent is high or low in that specific personality trait. For example, a high score in agreeableness means that the agent is highly agreeable and vice versa.

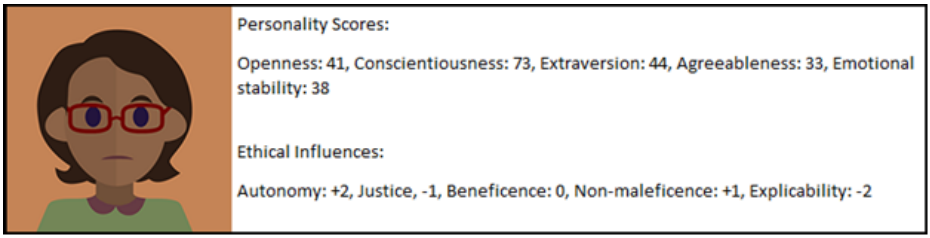

Figure 4: Shiva Avatar and Profile

To depict a human-like behavior in the agents, lip syncing of the dialogue and facial expressions were also embedded with the agent's interaction. For example, if something goes against the agent's principles, the agent makes a sad face. The game is particularly designed for a new employee in a cybersecurity organization who can then implement these principles in their subsequent decision making.

During their first video meeting in the 2FA scenario, the player and the other 3 agents (non-player characters) are discussing the implementation of a new 2-factor authentication policy in the organization. As programmed, the agents react to different situations as per their different personalities. The player participates in the dialogue by responding to questions asked by any agent in a variable period during dialogue. The player has a choice to select between two options for each question. Each option shows the opposite personality trait and/or ethical choice, so the selection of the answer depends upon the player's personality type and the player's ethical principle inclination. An example of a V-Meet application is found in Figure 5 which shows the characters talking to each other and the set of options which are provided to the player to provide their input.

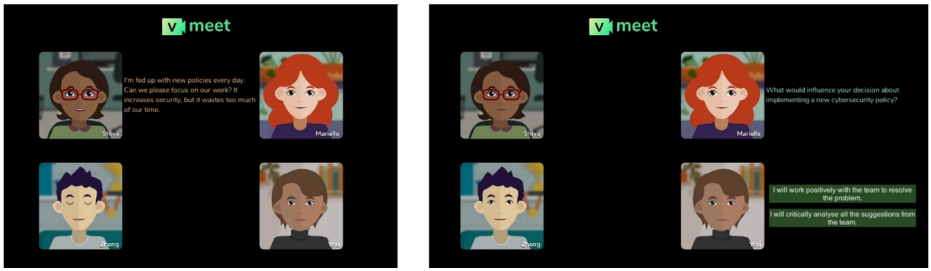

Figure 5: V-Meet Dialogue

A snippet of the agents' dialogue discussing the 2FA scenario is shown in Figure 6. In the given code snippet, it can be observed that Zheng is agreeing that 2-factor authentication is a good policy from his sentence "Seems like a good initiative", which shows that he has a high score in agreeableness. The blue highlighted area of the dialogue shows that Zheng is open to new experiences. Zheng regards the beneficence ethical principle as highly important, and this can be seen from the underline with 2 as a superscript that clearly shows that Zheng is oriented towards the principle of Beneficence.

Zheng:

The 2-factor policy seems like a good initiative. It is good idea to try implement this new policy in our company.
I hope it will benefit us and our clients[2] by increasing our security. As a Security Analyst, I believe this is a great step toward increasing the security of employees'[2] accounts.

Marielle:

Shiva, I've given careful consideration to what you are talking about. But I think, eventually we will have to strictly follow the company's policies. The email I received stated that failing to provide a personal mobile phone number will result in losing access to shared documents from our storage provider. And providing your phone number will also improve security.

Figure 6: 2FA Dialogue Snippet

3.2.2.Reflection Statements

After participating in each V-Meet meeting with other agents, the player

responds to a set of ten reflection statements. The reflection statements were used to analyze the player's ethical perspective on the decision taken by them in that dialogue. There were a total of ten reflection statements after each dialogue, two statements for each ethical principle. Five out of the ten statements were opposite to the decision the player took in the scenario and five statements were in favor of their decision. For example, the player will see two statements for the ethical principle 'Autonomy' in the 2-factor implementation scenario, with one statement in favour of implementing the policy and the other autonomy statement not in favour of implementing the policy. The player was asked to give thumbs up for a statement if the statement supports their decision and thumbs down if the statement is against their decision. The player also had to label each statement with the relevant ethical principle. A star is used to indicate the player's most important statement in support of their decision, and the bin is used if the player thinks that a statement was not relevant to their decision-making (as shown in Figure 7). The player can also click on Ethbot (bottom left) to remind them about the five ethical principles and to help them provide better reflections. The reflection statement was also used to answer our research question. From these results, we were interested to know if the player could correctly identify which ethical principle applied to each reflection statement.

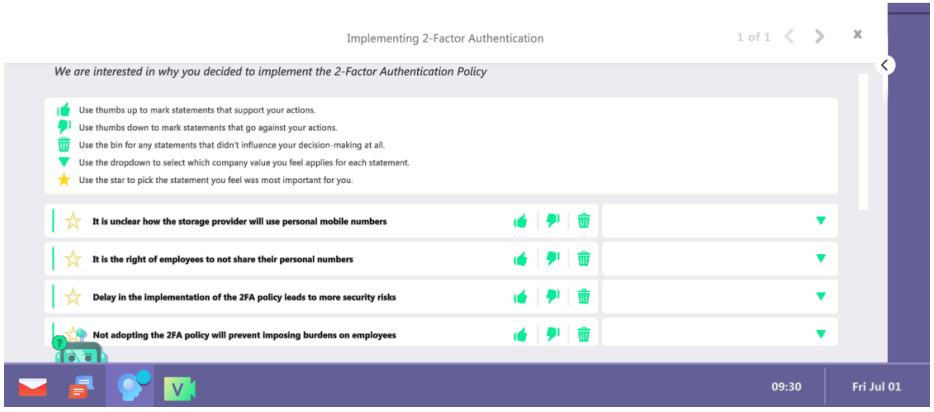

Figure 7: Reflection Statements to apply the ethical principle in different scenarios

3.2.3.Pre and Post scenarios

We designed two written cybersecurity ethical scenarios other than the two scenarios played in the game. The player had to respond to one scenario before playing the game and respond to the other scenario after playing the game. The player was provided with two options to choose from in each ethical dilemma. One scenario was about releasing an ethical worm to employees' computers to force automatic updates (Figure 8) and the second scenario was about undertaking a password hack to identify employees with poor passwords (Figure 9). The participants were divided randomly and evenly into two groups. Group 1 received scenario 2 first and group 2 received scenario 1 first for their pre-test. The reason behind randomizing the scenarios was to ensure that the participants' responses had nothing to do with which scenario was provided first. After taking the decision, the player was given the option to justify their answers. From their justifications, we aimed to answer our research question by determining if their ethical awareness increased after playing the game. This also allowed us to know whether the participant took into consideration our five ethical principles, or ethics more generally, when making their decision.

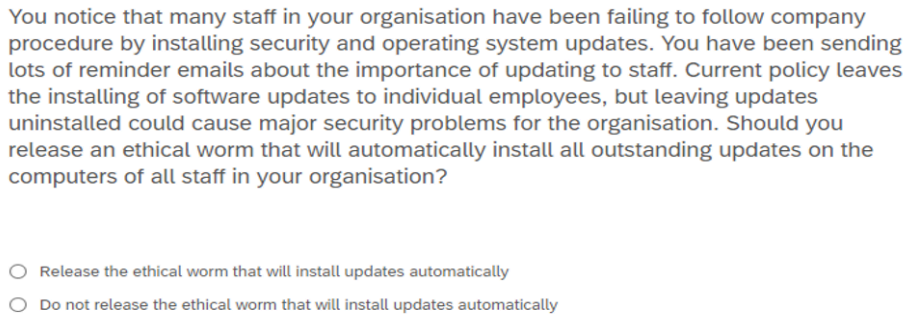

Figure 8: Scenario 1 (Pre/Post Test Scenarios)

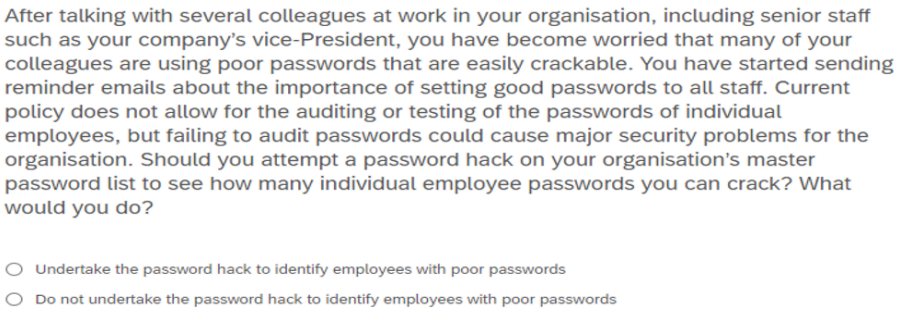

Figure 9: Scenario 1 (Pre/Post Test Scenarios)

3.3. Data Collection and Procedure

After providing consent, participants answered a set of demographic questions (age, gender, cultural background, and area of study). Then they answered questions concerning their knowledge of ethics using a 5-point Likert Scale (1=Terrible, 2=Poor, 3=Average, 4=Good, 5=Excellent), followed by completion of the Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003) to capture the individual's personality traits. The scale has ten items to measure the five personality traits of Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. The scale uses a 7-point Likert Scale (1=strongly disagree, 7=strongly agree).

To measure the change in ethical awareness before and after playing the game, we conducted a pre- and post-test involving the two scenarios described in Section 3.2.3. After completing the pre-test scenario, participants took a link in the survey to the game which passed an anonymous user ID to the game. After playing the game, participants completed the post-test scenario. Finally, we asked players whether their understanding of ethical principles increased by playing the game, and questions about their competency in the game and the intuitiveness of the game. The player responded to these questions using a 7-point Likert Scale (1-Strongly Disagree, 7-Strongly Agree).

3.4. Data Analysis

We calculated descriptive statistics (mean and standard deviation) and performed other statistical tests, such as the T-test to find any significant difference between pre- and post-test results. We used $p<0.05$ to determine the statistical significance.

To answer our research question, we aimed to identify whether the player was able to identify the ethical principle associated with the given ethical scenario. For that purpose, we used the reflection statements described in section 3.2.2. Correctly identifying the ethical statement shows that the player can identify the ethical principles in different scenarios and thus can apply those principles outside the training context. We excluded the statements nominated as irrelevant by the player and then found the average correctly identified from the remaining statements.

To determine a change in the ethical reasoning of participants, we coded their pre and post-test responses to two different ethical cybersecurity scenarios before and after playing the game. We capture how participants responded to those scenarios and analyzed the results using the t-test to determine any significant differences in the mean responses in their pre and post-test results.

To analyze the players' justifications used in the pre- and post-test scenarios, we encoded their justifications using the following coding scheme. Initially, six codes were created to categorize participants' responses. This coding scheme allowed us to categorize the given participants' text justifications about the decision they had taken in pre- and post-test scenarios. For example, they justified why they decided to release the ethical worm in an ethical worm scenario. We were interested to see if participants considered ethics in their decision-making or not. We grouped codes 2, 3 and 4 as ethical responses and 1, 5, 6 as others. This coding scheme was then further simplified to three codes (0,1,2). In order to calculate paired t-test we needed response to both the pre and post-test.

1. No answer or irrelevant answer was given – unrelated to the scenario or rubbish

2. They used at least one of our ethical principle terms (Autonomy, Justice, Beneficence, non-maleficence, explicability)
3. They used the word "ethics" or derivatives (e.g., ethically)
4. They discussed ethics in general, without using one of our particular words or terms
5. They chose to act in that way due to policy
6. Relevant answer was given, but did not refer to ethics

Simplified Coding Scheme:

1. No answer or no irrelevant answer – unrelated to the scenario or rubbish
2. Discussed ethics in general or used any ethical principle term [Combining 2,3 and 4]
3. Relevant answer, but did not discuss ethics at all [Combining 5 and 6].

4. RESULTS

A total of 272 participants took part in the online study conducted through Qualtrics. Out of the 272 participants, 40 participants didn't give consent to use their data. We therefore excluded their data from further analysis. After careful consideration of all the responses, we excluded all the invalid attempts. We considered responses to be invalid if the data in the survey response was too low, such as if they had only completed 10 percent of the survey, which means they had only answered the demographic questions. Moreover, a response was also invalid if we found a similar pattern in choice selection, such as the user always choosing option number 4 in all the questions. After excluding such responses, the total number of valid responses left for further analysis was 219.

Among those 219 responses, 153 (69.86%) responses were 100 percent complete and answered all the questions. 49 (22.38%) responses completed less than 50 percent of the survey, and 17 responses (7.76 %) completed the survey between 50 to 94 percent. 200 participants played

the game that was embedded inside the survey questionnaire. 19 participants did not play the game at all. Of the 200 participants who played the game, 92 (46%) participants completed the game, and 35 participants (17.5%) played the game but left the game before the start of the first V-meet scenario. Player logs reveal 102 attempts to play both scenarios in the game, which is 51% of the players who played the game. 25 players (12.5%) completed only the first scenario and left the game afterward. 19 players (9.5%) started responding to the first scenario but left before completing it, and the same number of participants completed the first scenario but did not complete the second scenario and left before completing it. We retained the data for those participants who played the first scenario and responded to the reflection statement as this data provides a full scenario cycle that was useful for analysis.

4.1. Participant's demographic information:

152 participants had Computing as their major area of study, which was 69.41% of the total population. 31 participants (14.16%) had Business, 15 (6.85%) had Arts, and two (0.91%) had Psychology as their major area of study. The remaining 19 (8.68%) included people having other areas of studies or combining two major areas such as Computing and Arts, Business and Computing, Cybersecurity, Software engineering, Information Technologies, Clinical Science, and Actuarial.

There were 62 female and 153 male participants, which were 28.31% and 69.90% of the total participants respectively. Only one participant selected the other option and identified as "Gender Fluid" and three participants didn't identify themselves with any gender. Participants were aged from 17 to 50 with an average age of 20.05 years. There were nine cultural groups to which the participants indicated they belonged. Participants were able to select more than one cultural group if they identifed themselves with more than one cultural group. Table 1 presents the cultural group with which participants identified themselves. Computing is the major area of study of the participants as shown in Table 2. Participants were asked to rank their knowledge about ethics in IT from values 1 to 5 (1=Terrible, 2=Poor, 3=Average, 4=Good, 5=Excellent) (mean 3.59). On average, participants

reported playing video games 2.75 hours per week. Personality scores are given in Table 3.

Cultural groups	Total	Percentage
Oceania (including Australian)	88	40.18%
South-East Asian	42	19.18%
Southern and Central Asian	22	10.05%
North-East Asian	12	5.48%
Northern-Western European	6	2.74%
Southern-Eastern European	7	3.20%
North African and Middle Eastern	15	6.85%
Sub-Saharan African	2	0.91%
People of the Americas	1	0.46%
Oceania & Northern-Western European	1	0.46%
Oceania & Southern-Eastern European	1	0.46%
Oceania & North African and Middle Eastern	1	0.46%
Oceania, South East & Central Asian	1	0.46%
No answer or do not identify	20	9.1%
Sum	**219**	**100%**

Table 1: Cultural Group Distribution

Main area of study	Total	Percentage
Computing	152	69.41%
Psychology	2	0.91%
Arts	15	6.85%
Business	31	14.16%
Others	19	8.68%
Sum	**219**	**100%**

Table 2: Participant's Area of Study

Personality traits (scale 1-7)	μ	SD
Extraversion	3.69	1.43
Agreeableness	4.51	1.00
Conscientiousness	4.70	1.18
Emotional Stability	4.43	1.37
Openness to Experiences	4.98	1.01

Table 3: Mean and Standard Deviation for TIPI

4.2. Pre and Post-test results

Table 4 provides the results comparing the pre- and post-test coded responses. This test is comparing each individual player's justification for their decision to the text-based scenarios provided before and after playing the game to see if their response has changed or not in terms of the use of ethical considerations.

	Pre-game			Post-game			
Scenario order	Ethical	Other	Total	Ethical	Other	Total	*P-value
Password Hack First	46	25	71	52	19	71	0.112
Ethical Worm First	44	27	71	56	15	71	0.007
Sum	90	52	142	108	34	142	0.004

Table 4: Pre and Post-Game Test Analysis using Paired T-test

4.3. Identification of the in-game ethical scenarios

A summary of correct responses to the reflection statements is shown in Table 5. Each player had to complete a reflection statement twice, once after the 2-factor authentication and again after the counter-attack scenarios. 92 participants (71.32 %) responded to both reflection statements and 37 participants (28.68%) only took part in the first set of reflection statements. Of the 129 participants who took part in the reflection part of the game, we found that 19 participants (14.73%) correctly identified less than 25 percent of the reflection statements, 72 participants (55.81%) identified less than half and more than 25 percent of the statements, and 38 (29.46%) correctly identified more than half of the reflection statements. We exclude those statements that were not answered by the participants.

Total no of participants	Percentage	% Of correctly identified statements
19	14.73	<25 %
72	55.81	25-50 %
38	29.46	> 50 %

Table 5: Correct identification of the Reflection Statements

Table 6 shows the number of times each ethical reflection statement had been correctly identified by the players. More players were able to correctly identify the statement related to Beneficence. On average, statements related to Beneficence and statements related to Justice were mostly identified correctly in the 2FA and counterattack scenarios, respectively. We saw some improvement in identifying the statements related to Justice. This can be found by taking the average of Justice-1 and Justice-2 in 2-FA (35.27%) and Justice-1 and Justice-2 from the counterattack scenario (51.09%). The improvement was 15.82 percent. Statements that were least identified correctly by the players were about Autonomy in the counterattack scenario as these statements were identified correctly only 16.85% times.

2-FA			
Ethical Principle Statement	N	Correct identification	
		N	%
Non-maleficence – 1	129	28	21.71
Non-maleficence – 2	129	43	33.33
Beneficence - 1	129	81	62.79
Beneficence - 2	129	37	28.68
Autonomy - 1	129	60	46.51
Autonomy - 2	129	28	21.71
Justice – 1	129	54	41.86
Justice – 2	129	37	28.68
Explicability - 1	129	56	43.41
Explicability - 2	129	39	30.23

Counterattack			
Ethical Principle Statement	N	Correct identification	
		N	%
Non-maleficence - 1	92	31	33.70
Non-maleficence - 2	92	20	21.74
Beneficence - 1	92	50	54.35
Beneficence - 2	92	23	25.00
Autonomy - 1	92	11	11.96
Autonomy - 2	92	20	21.74
Justice - 1	92	49	53.26
Justice - 2	92	45	48.91
Explicability - 1	92	29	31.52
Explicability - 2	92	25	27.17

Table 6: Frequency of correct identification of each reflection statement

4.4. Player experience

Mean and Standard Deviations for the player experience questions can be seen in Table 7. The results show that a greater number of participants thought that their understanding of the ethical issues in cybersecurity was improved after playing the game.

Reflections	**M**	**SD**
I felt competent at playing the game	5.31	1.43
The game was intuitive and easy to play	5.16	1.55
This game improved my understanding of the range of ethical issues raised by cybersecurity.	5.21	1.41

Table 7: Player Experience

5. DISCUSSION

As discussed above, 93% of cybersecurity breaches were caused by human error (Dunn, 2014). Not considering ethics in decision making is one of the causes of human error (Mohamad et al., 2005). Thus, providing training about considering ethics in decision making will reduce the cybersecurity breaches and this is the main contribution of the proposed game. We aimed to raise awareness of ethical reasoning in cybersecurity decision-making. To provide NPCs that helped players consider different viewpoints based on the importance of the five ethical principles and personality factors, we designed agents that had different personality behaviors based upon the Big Five personality model (Extraversion, Openness to experience, Consciousness, Emotional stability, and Agreeableness) and could provide different ethical perspective based upon five ethical principles (Beneficence, Non-Maleficence, Justice, Autonomy, Explicability) in the context of cybersecurity. The agents were embedded in a serious game to provide alternate ethical perspectives to increase the ethical awareness of the players. The goal of our study was to determine if interacting with the agents increased the player's knowledge about and awareness of ethical principles and how they impacted on cybersecurity issues and decision-making.

We recruited students enrolled in a first year cybersecurity unit, as these students will become IT professionals in the future, and we want to sensitise them to ethical thinking from the outset of their education and training. We observed that the gender imbalance in our study 69.90% (male) and 28.31% (female) corresponded with the gender imbalance in the technology industry, as reported by CompTIA (2020) the gender ratio in technology is 68% male to 32% female, which is very close to our gender ratio. The personality profiles of the cohort were not widely diverse. Extraversion has 1.43 SD, SD of agreeableness is 1.00, Conscientiousness has 1.18, emotional stability has 1.37 and openness has a standard deviation of 1.01. In general, we could characterize the cohort as tending towards introversion, agreeableness, conscientiousness, emotional stability, and openness to new ideas.

Answering our research question and study aim, the pre- and post-scenario

results show that, overall, there was a significant ($p < 0.01$) increase in ethical reasoning (12.68%) after playing the game. The results in Table 4 reveal that 63.38% of participants who responded to the pre-test scenario mentioned ethical issues, whereas 76.06% of participants mentioned ethical issues in justifying their response to the post-test scenario. Further, there were 33 (23.24%) participants who went from mentioning no ethical issues in the pre-test to mentioning ethical issues in the post-test. We provided two different scenarios to avoid learning effects (Georgiev, 2018; Wright, 1936) or other bias from using the same scenario twice. We note that regardless of which scenario was received first, the number of players expressing ethical reasons for their decision increased for both scenarios, but only the group that received the ethical worm scenario first significantly changed their reasoning. Possibly participants found it easier to identify ethical concerns with the password hack scenario than in the ethical worm scenario (this is supported by a higher incidence of ethical responses whether it was received first or second). The scenario differences demonstrate the difficulty of designing equivalent scenarios and the contextual nature of ethical reasoning. Further, according to (MOSHER, 2018) 90% of the small business owners and employees are worried about password hacks and 96% are concerned about viruses such as malware, worms, etc. But according to Spafford (1991), some people claim that some viruses are beneficial, such as ethical worms, but more people consider them as dangerous. According to these numbers, we can say that our results confirm that people seem to be more familiar with the concept of password hacking than ethical worms.

We further evaluate the knowledge of the player by providing reflection statements. The results of the reflection statements show that a large number of participants were able to recognize the ethical dilemmas and which ethical principle is best applied in this dilemma. Our analysis showed that nearly 30% could correctly identify half of the statements and principles correctly. This indicates that there is low familiarity with the application of ethical principle in cybersecurity decision-making. This encourages us to develop more effective training for ethical principles in our future studies and also identifies a need for cyber security professionals and researchers to focus on this area. Lastly, the participants acknowledged that the game was easy to play, and increased their

knowledge. Analysis of the 97 out of 142 participants who reported an increase in knowledge with their correct identification of the principles in the reflection statement and the post-test reported that 18 (18.56%) responses shifted from "other" to ethical, which supports their claim of increasing ethical knowledge. Moreover 44 (45.36%) of those participants were able to correctly identify more than 30 percent of reflection statements.

6. CONCLUSION, LIMITATIONS AND FUTURE WORK

We developed a serious role-playing game to help users to consider the ethical features of their cybersecurity decisions. The awareness of ethical principles was provided with the help of artificial agents that acted as cybersecurity professionals and provided their differing ethical perspectives. We conducted a study to analyze the impact of our game on the ethical awareness of participants. Our results verify the literature finding that observing the acts of other individuals can change one's behavior. According to our results, 12.68% of individuals shifted from a non-ethical to an ethical response when responding to our pre- and post-test scenarios after playing the game. This shows that our game has potential to educate cybersecurity professionals to become more aware of ethical principles and their application in their decision-making. The participants also acknowledged that the game increased their knowledge of cybersecurity ethics.

We acknowledge some limitations of our game and our study and we plan to address them in future studies. Firstly, the difference between participants pre- and post- test results and their player experience results might be because of the Acquiescence bias (Ross & Mirowsky, 1984), as the question was very straightforward and shows the aim of the study conducted so people tend to agree with the statement as this is easy way out. So, the player experience question could be improved to allow the participants to choose their own answer without any bias.Currently the agents designed in our game depict the personality and ethical inclinations separately. We aim to find the relation between personality and ethical inclinations so that more realistic agents can be designed. Moreover, the agents need to be dynamic and change their mental state according to

the needs of the environment or other agents. For example, an agent could observe if a player is low on a specific ethical principle and then provide tailored training or perspectives to that agent on that specific ethical principle. To create more realistic non-player characters (agents) in serious games, agents should be able to respond to organizational norms and policies (Dignum et al., 2009). Multi-agent interaction and communication (Dignum et al., 2009) is thus another milestone we aim to achieve to design more intelligent agents in serious games in a cybersecurity ethics context.

As far as our study is concerned, the types of participants recruited for the study was a limitation. All the participants were university students enrolled in the course of "Introduction to Cybersecurity". This is justified since these students were future cybersecurity professionals and providing them with training in ethics was vital. However, a wider range of participants should be recruited, ranging from early-stage students to experts in the domain of cybersecurity. Moreover, as reported in the results section, 92 participants were able to complete the game. It is unclear why other participants left the game before completing. This might be because they had less time to play the game and they wanted to return to the remaining part of the survey before the time ended as they only had 30 minutes in class to complete the game and study survey.

This is our first step in including agents in a serious cybersecurity game and we aim to make our agents more intelligent so that they can adopt the dynamic behaviors and reflect other agents accordingly. This should make the training even more effective training through using more intelligent agents. We plan to explore whether longer gameplay and more scenarios and reflection or multiple play sessions could make the game even more effective.

REFERENCES

Ayyagari, R. (2012). An exploratory analysis of data breaches from 2005-2011: Trends and insights. *Journal of Information Privacy and Security*, *8*(2), 33-56. https://doi.org/10.1080/15536548.2012.10845654

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics*. Oxford University Press, USA.

Blanken-Webb, J., Palmer, I., Deshaies, S.-E., Burbules, N. C., Campbell, R. H., & Bashir, M. (2018). A case study-based cybersecurity ethics curriculum. 2018 USENIX Workshop on Advances in Security Education (ASE 18),

Brey, P. (2007). Ethical aspects of information security and privacy. *Security, privacy, and trust in modern data management*, 21-36. https://doi.org/DOI: 10.1007/978-3-540-69861-6_3

CompTIA. (2020). *The definitive guide to the U.S. tech industry and tech workforce.* https://comptiacdn.azureedge.net/webcontent/docs/default-source/research-reports/comptia-cyberstates-2020.pdf?sfvrsn=39494164_0

Costa Jr, P. T., & McCrae, R. R. (2008). *The Revised Neo Personality Inventory (neo-pi-r)*. Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Craft, J. L. (2013). A review of the empirical ethical decision-making literature: 2004–2011. *Journal of business ethics*, *117*(2), 221-259. https://doi.org/10.1007/s10551-012-1518-9

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, *4*(10). https://doi.org/10.22215/timreview/835

De Freitas, S., & Jarvis, S. (2007). Serious games-engaging training solutions: A research and development project for supporting training needs. *British Journal of Educational Technology*, *38*(3), 523. https://doi.org/10.1111/j.1467-8535.2007.00716.x

Dignum, F., Westra, J., van Doesburg, W. A., & Harbers, M. (2009). Games and agents: Designing intelligent gameplay. *International Journal of Computer Games Technology*, *2009*. https://doi.org/10.1155/2009/837095

Dunn, J. (2014). Data breaches in UK healthcare sector double since 2013, ICO numbers show. *Computerworlduk. com, available at: www.*

computerworlduk. com/news/security/data-breaches-inuk-healthcare-sector-double-since-2013-ico-numbers-show-3589814/(accessed 12 November 2017).

Ferro, L. S., Marrella, A., Catarci, T., Sapio, F., Parenti, A., & De Santis, M. (2022). AWATO: A Serious Game to Improve Cybersecurity Awareness. In *International Conference on Human-Computer Interaction* (pp. 508-529). Springer, Cham.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, *28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Formosa, P., Wilson, M., & Richards, D. (2021). A principlist framework for cybersecurity ethics. *Computers & Security*, *109*, 102382. https://doi.org/10.1016/j.cose.2021.102382

Gentry, J. W. (1990). What is experiential learning. *Guide to business gaming and experiential learning*, *9*, 20.

Georgiev, G. (2018). Representative samples and generalizability of A/B testing results. http://blog.analytics-toolkit.com/2018/representative-samplesgeneralizability-a-b-testing-results/.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological science*, *20*(3), 393-398. https://doi.org/10.1111%2Fj.1467-9280.2009.02306.x

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528. https://doi.org/10.1016/S0092-6566(03)00046-1

Hale, M. L., Gamble, R. F., & Gamble, P. (2015, January). CyberPhishing: a game-based platform for phishing awareness testing. In *2015 48th Hawaii International Conference on System Sciences* (pp. 5260-5269). IEEE.

Jordan, C., Knapp, M., Mitchell, D., Claypool, M., & Fisler, K. (2011).

CounterMeasures: A game for teaching computer security. in 2011 *10th Annual Workshop on Network and Systems Support for Games*, (pp. 1-6). IEEE.

Kianpour, M., Kowalski, S., Zoto, E., Frantz, C., & Øverby, H. (2019). Designing serious games for cyber ranges: a socio-technical approach. 2019 IEEE European symposium on security and privacy workshops (EuroS&PW),

Larson, K. (2020). Serious games and gamification in the corporate training environment: A literature review. *TechTrends*, *64*(2), 319-328. https://doi.org/10.1007/s11528-019-00446-7

Loi, M., & Christen, M. (2020). *Ethical frameworks for cybersecurity* (Vol. 21). Springer Cham, Switzerland. https://doi.org/10.1007/978-3-030-29053-5_4

Mohamad, S., Aliandrina, D., & Feng, Y. (2005). Human Errors in Decision Making. https://mpra.ub.uni-muenchen.de/8171/

Morgan, G., & Gordijn, B. (2020). A care-based stakeholder approach to ethics of cybersecurity in business. *The Ethics of Cybersecurity*, 119. https://doi.org/10.1007/978-3-030-29053-5_6

MOSHER, G. (2018). Cybersecurity: Reality Check. https://blog.avast.com/cybersecurity-reality-check

Pereira, G., Brisson, A., Prada, R., Paiva, A., Bellotti, F., Kravcik, M., & Klamma, R. (2012). Serious games for personal and social learning & ethics: status and trends. *Procedia Computer Science*, *15*, 53-65. https://doi.org/10.1016/j.procs.2012.10.058

Reed, M. S., Evely, A. C., Cundill, G., Fazey, I., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., & Raymond, C. (2010). What is social learning? *Ecology and society*, *15*(4). http://www.ecologyandsociety.org/volXX/issYY/artZZ/

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The big five personality factors and personal values. *Personality and social psychology bulletin*, *28*(6), 789-801. https://doi.org/10.1177/0146167202289008

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and

acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, 189-197. https://doi.org/10.2307/2136668

Spafford, E. H. (1991). Computer viruses and ethics. https://docs.lib.purdue.edu/cstech/901

Streeter, D. C. (2013). The effect of human error on modern security breaches. *Strategic Informer: Student Publication of the Strategic Intelligence Society*, *1*(3), 2. https://digitalcommons.liberty.edu/si/vol1/iss3/2

Urquhart, L. D., & Craigon, P. J. (2021). The Moral-IT Deck: a tool for ethics by design. *Journal of Responsible Innovation*, *8*(1), 94-126. https://doi.org/10.1080/23299460.2021.1880112

Vallor, S., Green, B., & Raicu, I. (2018). Ethics in technology practice. *The Markkula Center for Applied Ethics at Santa Clara University. https://www.scu.edu/ethics*

van de Poel, I., & Christen, M. (2020). Core values and value conflicts in cybersecurity: beyond privacy versus security. *The Ethics of Cybersecurity*, 45. https://doi.org/10.1007/978-3-030-29053-5_3

Veneruso, S. V., Ferro, L. S., Marrella, A., Mecella, M., & Catarci, T. (2020). CyberVR: an interactive learning experience in virtual reality for cybersecurity related issues. In *Proceedings of the International Conference on Advanced Visual Interfaces*, (pp. 1-8)

Weber, K., & Kleine, N. (2020). Cybersecurity in Health Care. In *The Ethics of Cybersecurity* (pp. 139-156). Springer, Cham. https://doi.org/10.1007/978-3-030-29053-5_7

Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the aeronautical sciences*, *3*(4), 122-128.