



UCC

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

An Exploration Into the Potential Use of Gamification and Conversational Learning in AI Ethics Education

Name: Aidan Roche

Student Number: 121110118

Supervisors: John McCarthy, Sarah Robinson, John O'Mullane

Submitted: April 9th, 2024

Thesis presented as part of the requirements for the B.A. (Honours) Degree in Psychology and
Computing, University College Cork

DECLARATION OF ACADEMIC HONESTY: I declare that the content of this assignment is all my own work. It has not been submitted in respect of any other course/module. Where I have used the work of others it is acknowledged and referenced accordingly.

Contents

Abstract	4
Introduction	5
Literature Review	9
Education on AI	9
Gamification	10
Research Through Design	12
Methods	14
Description of Digital Education Game	14
Development of Digital Education Game	19
Evaluation	25
Ethical Measures	27
Results	28
Insights from RtD	28
Results From Interview Data	30
Results From Survey Data Analysis	33
Discussion	34
Gamification and Conversational Learning in AI Ethics Education	34
RtD as a Methodology in AI Education Research	37
Limitations	38
Conclusion	40

	4
References	42
Appendix 1: Digital Education Game Live Link & Code	47
Appendix 2: Interview Question and Coding Examples	48
Interview Questions	48
Example of Coded Transcript	49
Appendix 3: AI Literacy Test Questions	63

Abstract

As Artificial Intelligence (AI) continues to become ubiquitous in modern society it is having a profound impact on people's lives. This has led developers and consumers alike to raise a number of ethical concerns regarding AI. However, due to the complex, and often opaque, nature of AI, the public may have misconceptions on how the technology works, and the associated ethical concerns. These misconceptions may lead to ill-informed legislation, or direct harm to users. To mitigate these potential consequences, there is a need for more education on the topic. This study employs a research through design (RtD) approach to the development of a digital education game. The goal of this study is to explore the potential for gamification as a pedagogy in AI education, and to provide a reflection on the value of RtD in this context. This research draws its findings from three sources: the development process itself via RtD, interviews with undergraduate students, and a survey to compare the effectiveness of the game versus traditional text-based teaching. The game utilizes a combination of two pedagogies, gamification and conversational learning, and suggests that this approach may be valuable for future AI education. The interview findings indicate that the game is effective in keeping students engaged, motivated, and fostering critical thinking. The game was shown to have a statistically significant impact on people's scores on an AI literacy test, but was not more effective compared to text-based learning. The applied RtD approach was valuable in its ability to provide an in-depth understanding of the problem space, but in turn, it sacrifices certain quality assurance measures and is not suited for the development of a publishable product. However, further design iterations would be required to fully verify these findings. The aim of this paper is to provide proof of concept for the combination of gamification and conversational learning to teach AI ethics, and serve as a foundation for future research into this topic.

Introduction

The introduction of Artificial Intelligence (AI) into modern society has had an indisputably profound impact on people's lives. AI typically refers to technology that is able to learn and act autonomously, and is able to display, at least to some extent, human-like behavior such as reasoning and problem solving (Zhou & Chen, 2023; IBM, n.d.). This means that for the first time in history, a non-human entity is capable of generating content, such as text and art, that rivals human creations. This is a potentially pivotal moment in society, as there are now non-human stakeholders present in what were historically limited to intra-human conflicts, such as copyright infringement. The technology enabling AI systems has progressed at an exceptionally rapid rate in the past few years. This has led to a nearly ubiquitous presence of AI in the modern digital world (European Parliament, 2020). Unfortunately, this has led to serious concerns regarding the impact AI may have on people's lives.

The technology is extremely powerful, yet it is neither inherently good nor bad. For example, AI-enabled systems have already been proven to be capable of transforming the medical world, by expediting the diagnosis process for rare diseases and increasing the overall efficiency of hospitals (Lee & Don, 2021). However, these same systems could be modified to create dangerous bio-weapons, such as peptide toxins (Lee et al, 2022). These contrasting outcomes encapsulate the ambivalent nature of AI. This has created a unique situation in which hindering the development of AI could result in the loss of human lives, but the same is true for the advancement of AI. Therefore, it is vital that the development of these systems is conducted in an ethical manner, and that they are carefully monitored following implementation.

The field of study concerned with these topics is known as ethics of AI, or simply, AI ethics; it can be defined as “the moral behavior of humans as they design, construct, use and treat

artificially intelligent beings, as well as concerns with the moral behavior of AI agents” (Zhou & Chen, 2023, p. 3). Recent publications on the topic have examined issues such as deception (Masters et al., 2021), data privacy (Huang, 2023), biosecurity (Lee et al, 2022), and the concept of trust and AI (Ryan, 2020). Ethics is an incredibly important field, as it can, for example, inform government legislation and impact consumer behavior (Capgemini, n.d.; Madiega, 2021). Research has shown that how a company is perceived from an ethical perspective can impact people's trust and satisfaction regarding that company (Capgemini, n.d.). The European Union also recently passed the first piece of comprehensive legislation governing the implementation of AI (Madiega, 2021). The passing of this act, known as the AI Act, potentially sets the stage for AI ethics to become a topic of political relevance in the coming years.

However, due to the complex, and often opaque, nature of AI technologies public understanding of them is often limited (Burrell, 2016). Burrell (2016) argues that this opacity exists in three forms, intentional state or corporate secrecy, technical illiteracy, and characteristics and scale of modern machine learning algorithms. These forms of opacity, either independently or concurrently, may serve to limit public understanding of how AI systems function. These misconceptions may lead to unintended consequences, such as ill-informed legislation which could inhibit the development of potentially valuable AI systems, or alternatively, it could lead to unethical, and even harmful, implementations of AI (Long & Magerko, 2020).

To avoid these outcomes it is important that the public is educated on the underlying ethical principles guiding these decisions, as this will allow for more effective implementation of ethical practices into the development process (Zhou & Chen, 2023). While education is important on a larger legislative scale, it is equally so at an individual level, so that users are

aware of the associated risks with using AI systems, and can make informed decisions regarding their own use. This is especially true for students as AI continues to become more prominent in school settings (Huang, 2023). An analysis of 115 university level tech ethics classes found that AI was one of the most commonly taught topics (Fiesler et al., 2020), further highlighting the importance of education on the subject at a college level. However, there is a need for more research on how AI ethics should be taught, with a recent study stating “more studies on the education of AI ethics are still necessary to investigate approaches to educate various stakeholders of AI solutions effectively” (Zhou & Chen, 2023, p.12). In recent years, there has been significant growth in the amount of available research on these different approaches. While many of these publications focus on education within a college curriculum (i.e. Borenstein & Howard, 2021; Fiesler et al., 2020), there has also been some research into pedagogies that can exist both inside and outside the classroom setting.

One such pedagogy that has shown great potential is the use of gamification. A variety of studies have explored the use of digital learning games in fostering AI literacy, and have found this method to increase engagement and knowledge-retention (Pinski et al., 2024; Koravuna & Surepally, 2020). While these studies explore AI literacy at a higher level, as opposed to AI ethics, other studies have also found gamification to be a potentially viable method for teaching tech ethics (Brinkman & Miller, 2017; Canosa & Lucas, 2008). However, these studies are non-digital and are constrained to a classroom/group setting. This highlights another common theme in the current research on gamification in AI education, many of these studies employ partner/group games. Meaning an individual is not able to use the game on their own. There is also limited research exploring the use of digital education games in teaching AI ethics specifically. Therefore, this present paper takes a research through design (RtD) approach to

explore the potential for digital education games as a pedagogy for teaching undergraduate students about AI ethics.

The aim of this paper is two-fold. First, evaluate digital education games as a potential pedagogy for teaching ethics of AI. Second, evaluate RtD as a potential methodology for AI ethics education research. To accomplish these objectives a single-player digital education game is developed throughout this study, and is evaluated using feedback from a set of interviews with undergraduate students, and a survey that assesses AI literacy. To better situate the objectives of this research within the context of the existing literature this paper will proceed with a more in-depth literature review, followed by a methods section to describe the process taken to achieve the aforementioned objectives. Next, a results section will describe the outcomes of this research, and the discussion section will consider the implications of these findings. Finally, the conclusion will consider the significance of these findings, and the impact they may have on the field of AI ethics education.

Literature Review

Education on AI

Research on the topic of AI education has seen significant growth in recent years, with much of this research focused on the topic of AI literacy (Ng et al., 2023); a term which refers to a set of competencies relating to people's use and understanding of AI (Long & Magerko, 2020). Until 2021 the vast majority of AI related education took place at a university level, and it is only in recent years that literature has begun to emerge exploring AI education in K-12 classrooms (Ng et al., 2023). Since then, there have been a sizable number of studies investigating AI education in a variety of settings (i.e. Ng et al., 2022; Koravuna & Surepally, 2020).

Ng et al. (2023) recently published their paper *A review of AI teaching and learning from 2000 to 2020* which currently provides the most extensive analysis of AI teaching in the past two decades. They found that the most popular tools for teaching AI were software development platforms, with the second most popular being robotics. This may have limited AI education to students already comfortable with, and interested in, technology. However, the increased “availability of easy-to-use AI-empowered teaching tools” (Ng et al., 2023, p. 2) in recent years has led more educators to incorporate AI education into the curriculum of non-technical students. This shift in scope is a pivotal moment in AI education, as the subject is being brought to a larger audience. The importance of this has been stressed in a number of recent publications, due the prevalence of AI, and its potential to impact people's daily lives (i.e. Borenstein & Howard, 2021; Koravuna & Surepally, 2020).

As far as how these subjects are being taught, a variety of different pedagogies have been tested. These pedagogies refer to the methods of teaching, as opposed to the aforementioned tools, such as robotics, which can be used instrumentally within different pedagogies. Ng et al.

(2023) found that collaborative project-based learning and learning through game design were the two most popular. With the former of these fostering students' knowledge through social interaction and collaboration, and the latter providing easier access to AI concepts for beginners. It is important to note that learning through game design is different from gamification. Game design involves students creating their own games, whereas gamification involves playing a game created to teach something specific. Although, it has been theorized that there is significant overlap “in terms of enhancing [students] understanding and their levels of interest, motivation and engagement” (Ng et al., 2023). The use of gamification in AI education is an idea that has been suggested in other popular frameworks, such as *What is AI Literacy? Competencies and Design Considerations* by Long & Magerko (2020). It is within this sub-field of gamified AI education that this present paper aims to add to the literature.

Gamification

Gamification as a pedagogy involves the implementation of game elements into an educational experience (Aslan & Balci, 2015). The existing data on gamification suggests that it can lead to improvements in cognitive learning outcomes, motivation and emotion, and engagement and behavior (Rabah et al., 2018). These benefits have also been found to be true when gamification is applied to AI education (Ng et al., 2023). For example, Pinski et al. (2024) found that gamification could be used to develop students' metaknowledge of AI literacy. Another study used gamification to teach AI concepts through the development of related critical thinking skills, such as planning and adaptability (Du & Wang, 2023). Both of these studies found gamification to be an effective pedagogy within the context of AI education.

Past studies have also examined the use of gamification in teaching tech ethics; a field of study concerned with broader ethical concerns relating to the role of technology within

society (Canosa & Lucas, 2008). Examples include using mock-trials in a university classroom setting (Canosa & Lucas, 2008), and a quiz game to teach the ACM Code of Ethics (Brinkman & Miller, 2017). However, both of these studies implemented gamification in a non-digital context, and instead focused on facilitating a group discussion about various ethical dilemmas. A central theme throughout the aforementioned AI literacy and tech ethics studies is that they did not focus on teaching explicit knowledge, but instead fostered the necessary critical thinking skills to allow students to evaluate situations themselves. Another commonality, excluding the study by Pinski et al. (2024), was the use of collaboration and discussion amongst participants. The study by Du & Wang (2023) accomplished this via a multiplayer game, whereas both of the ethics focused studies emphasized the importance of conversation.

This style of conversational learning appears to be a recurring theme throughout ethical literature. For example, in his famous book *The Republic* Plato (2008) utilizes an interesting narrative style in which his protagonist, Socrates, engages in conversation with a variety of people. Throughout these conversations they share ideas, critically evaluate each other's thought processes, and find flaws in their thinking. Conversational learning has been explored as a pedagogy and has been shown to be an effective teaching methodology, especially in the context of business ethics (Sims, 2004; Thomson, 2011), further suggesting its viability in teaching AI ethics. Research into conversational learning has explored it from various perspectives, such as through experiential learning theory (Baker et al., 2005) and cybernetic systems theory (Sharples & Ferguson, 2019), and has been shown to improve individuals' ethical assessment and critical thinking skills (Thomson, 2011). This style of learning requires people to look beyond the surface of an issue, and instead focus on the underlying principles, or in other words, the deep structure of a problem. Learning to break problems down into the underlying deep structure is a

core pillar of teaching critical thinking (Willingham, 2019), and appears to be an important part of ethics literature and education. However, this style of learning almost always requires a group setting, and is therefore not available for someone learning alone. This is a key part of the digital education game being developed in this present study, it attempts to foster this style of conversational learning but in a single-player game.

The majority of studies on the topic of AI education also appear to focus primarily on the product, and not the process. They often follow a similar structure of developing a game, testing the game with participants, and using this data to determine the value of gamification. There is minimal reflection on the development process, and almost all the insights of these studies stem from the final participant centered evaluation. A goal of this study is to test an alternative approach to research on AI ethics education. This will be achieved through the utilization of a methodology known as RtD.

Research Through Design

RtD is a research methodology that exists at the intersection between research and design. It was first proposed by Frayling (1993), and suggests that knowledge can be gained through the process of designing a product. A key aspect of RtD is that the outcome of a study employing this methodology is knowledge, not the product. The design process is instrumental, and serves to help the researcher explore “how things will be”, and the preferred state of this reality (Jonas, 2006). The goal of this approach is to develop a dialogue with the product being developed. This ideally leads to a cycle of iterative development, and helps the researcher draw novel insights about the artifact being made (Godin & Zahedi, 2014).

In the context of this paper the goal of RtD is to better understand how digital education games could be used to teach undergraduate students about AI ethics. This will be achieved by

developing a digital education game, and using this development process as a learning experience to inform the proposed research question of how digital education games can be used to teach ethics of AI. Success in this aspect requires an in-depth understanding of the problem space and the various stakeholders (Zimmerman & Forlizzi, 2014). The developers must then use their own understanding of human-computer interaction (HCI) alongside insights extracted from existing literature to create a product that, ideally, exemplifies the aforementioned preferred state of reality, in other words, how things will be in the future of gamified AI ethics learning.

As discussed earlier in this paper one of the research objectives of this study is to evaluate RtD itself, and determine its viability as a research approach within the context of AI education. As research in this field continues to progress it is important to not only be evaluating the outcomes of these studies, but the process applied as well. Currently, there does not appear to be any research on this matter within AI education literature, and this present paper hopes to create a foundation for future research on the topic.

Methods

This study had two objectives: explore the potential of gamification as a pedagogy in teaching AI ethics, and evaluate RtD as a methodology in AI ethics education research. This paper is part of a larger research project, consisting of four studies, that aims to evaluate gamification as a pedagogy for fostering AI literacy. The four studies were divided between development and evaluation, with two studies focused on each. This present paper falls into the former category, and applies a RtD approach to the development of a digital education game. This game was designed to teach players about ethics of AI, explainable AI, and data literacy; however, this paper is focused specifically on ethics of AI. The theoretical framework behind this paper is that by applying a RtD approach to the development of a digital education game that teaches AI ethics, the developers would be able to learn about the intricacies of the problem space, and gain valuable information regarding how gamification could be used to teach AI ethics. Since there is currently minimal research on the use of RtD in the field of AI ethics education this paper also provides a reflective evaluation on the implications of using this methodology.

This section will proceed with a description of the game created, followed by an explanation of the development process. There will then be an outline of how the RtD approach allowed for relevant knowledge about the problem space to be extracted. Finally, there will be a description of the evaluation methods taken by the two aforementioned evaluative studies, and how their results were incorporated into this study.

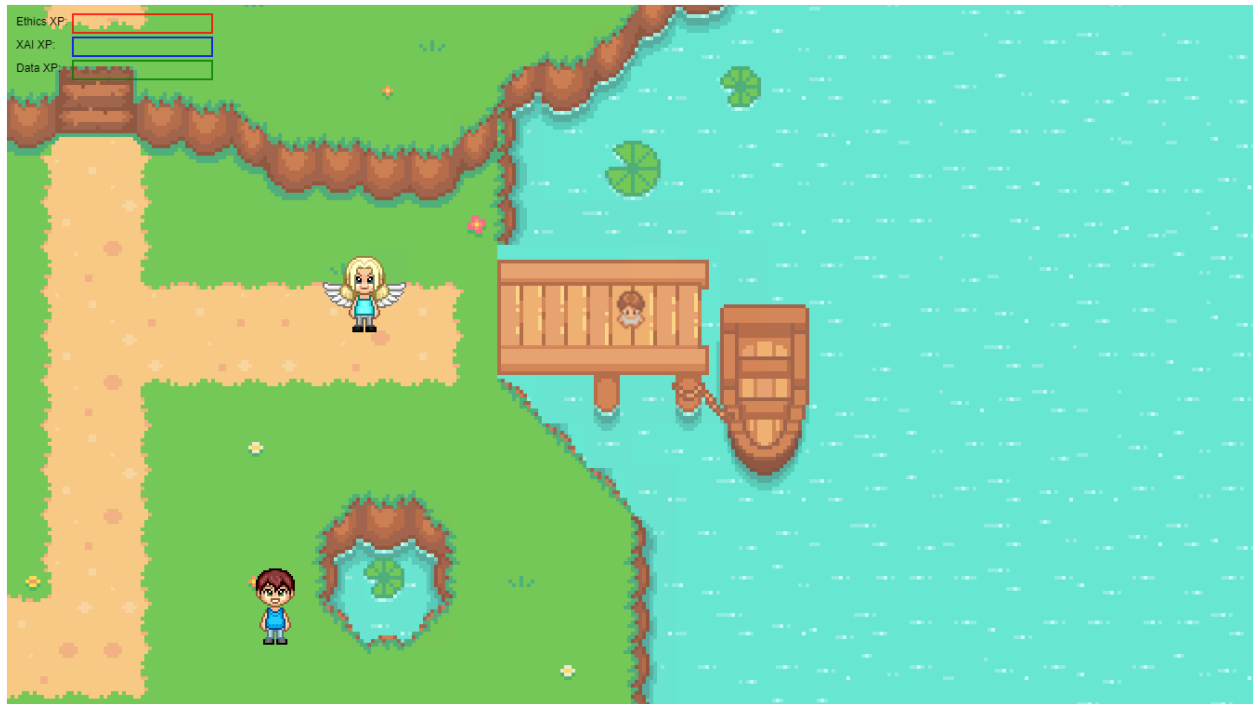
Description of Digital Education Game

The digital education game developed throughout this study was designed to teach players about certain relevant AI ethics topics, and give them the necessary critical thinking

skills to evaluate ethical dilemmas independently. The game is a web-based game designed for desktop. The player controls a sprite in the center of their screen using the arrow keys on their keyboard. When the game starts the player will see the following, shown in Figure 1.

Figure 1

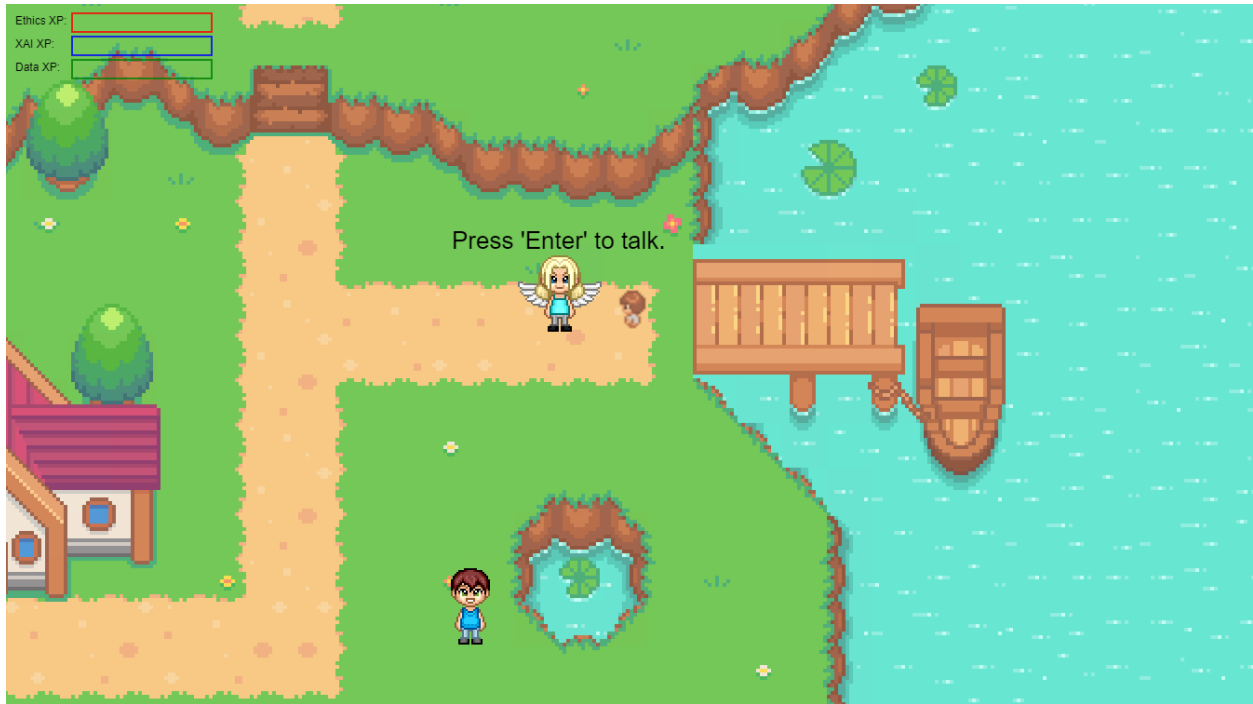
View of Game When First Loaded



The map for this game is an island, named AI-landia, that the player can explore freely. On this island they will find a number of non-playable characters (NPC), and as they approach these characters a prompt will appear stating “Press ‘Enter’ to Talk”, this can be seen below in Figure 2.

Figure 2

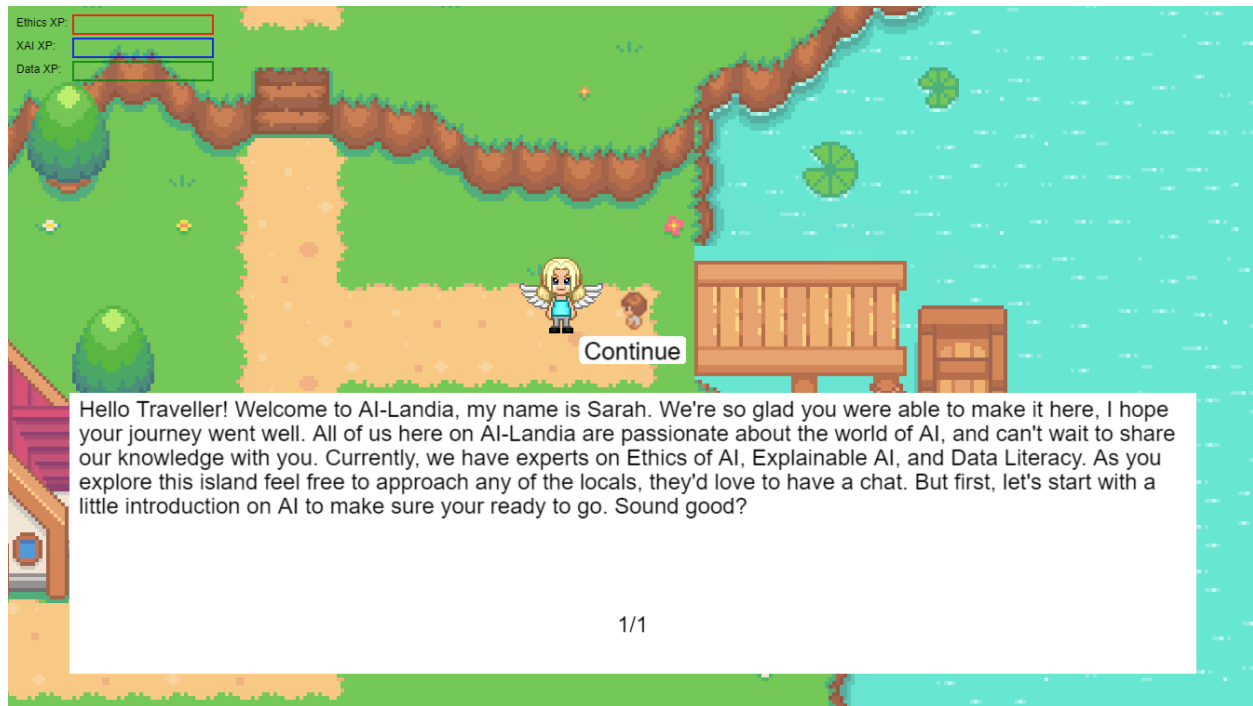
Approaching a NPC



At this point, the player can press the Enter key on their keyboard, and this will start a text-based dialogue interaction with the NPC. An example of this is shown in Figure 3.

Figure 3

Interaction with NPC



In Figure 3 a large white text box can be seen on the bottom half of the screen. This text box contains the dialogue from the NPC, and represents what they are saying. The player is able to interact with the NPCs through a multiple-choice system. Each choice represents what the player can say to the NPC, and the NPC's response will change accordingly. For example, once the player clicks the 'Continue' button in Figure 3 the NPC will continue with the following dialogue:

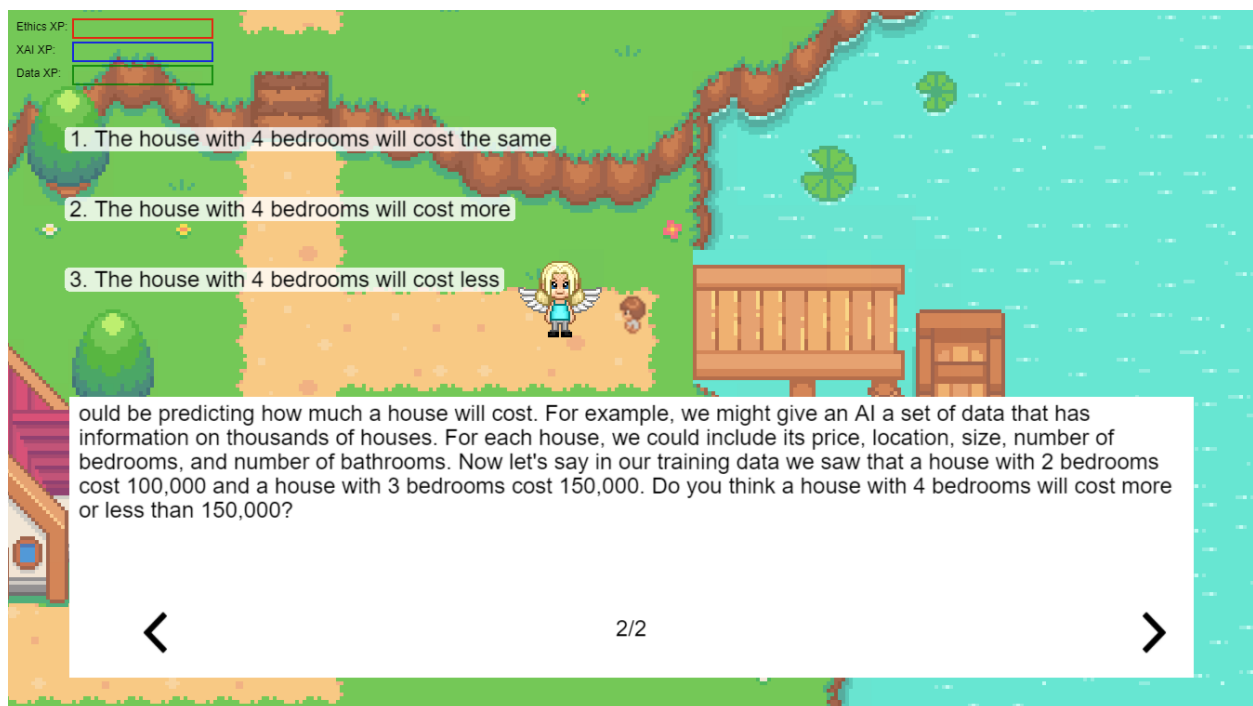
AI has been a pretty big buzzword, it feels like everyone is talking about it. In case you weren't sure, AI stands for Artificial Intelligence. Generally speaking, AI refers to the fields of Machine Learning and/or Deep Learning, but of which are subfields of computer science. Put simply, these fields are concerned with mathematically evaluating large sets of data, to consequently make predictions or classifications. An example of this could be predicting how much a house will cost. For example, we might give an AI a set of data that has information on thousands of houses. For each house, we could include its price,

location, size, number of bedrooms, and number of bathrooms. Now let's say in our training data we saw that a house with 2 bedrooms cost 100,000 and a house with 3 bedrooms cost 150,000. Do you think a house with 4 bedrooms will cost more or less than 150,000?

The player is then presented with the following three choices: “the house with 4 bedrooms will cost the same”, “the house with 4 bedrooms will cost more”, and “the house with 4 bedrooms will cost less”, this is shown in Figure 4.

Figure 4

Example of Multiple Choice Style Dialogue



Depending on which of these options the player clicks, the NPC's response will change. This style of interaction is the basis upon which this game was designed. There are currently six different NPCs spread across the map, and each of them has information on different topics. The player is able to engage with them in the same manner as shown above. By creating a dynamic

interaction, and giving the user control over their choices, the game attempts to emulate a conversation between the NPC and the player. Through these conversations the game is teaching the player about ethics of AI, and asks them questions that require careful consideration - with the intention of engaging relevant critical thinking skills.

Development of Digital Education Game

The development of this digital education game was guided by the GAMED framework. Proposed by Aslan & Balci (2015) it is intended to provide a guide for developing digital education games, and consists of an iterative cycle beginning with the identification of the education problem domain. Due to the breadth of AI ethics, and the time constraints of this study, it was important to narrow down what topics would be discussed within the game. The initial list of topics included accountability, biosecurity, fraud and deception, privacy, exacerbating discrimination, AI in education, AI in the workforce, and intellectual property. These were chosen by determining which topics appeared to be most relevant in current AI ethics literature. In the end, due to time constraints, accountability, intellectual property, and data privacy were chosen as the topics for the game. The goal of this study is an initial exploration on gamification in this context, not an exhaustive teaching of AI ethics; the focus is not on the topics themselves, but rather on the broader underlying ethical principles. The research team also deemed these three topics relevant for end users. Accountability and intellectual property are issues that may occur when using relatively common AI systems, such as self-driving cars and ChatGPT. Therefore, it seemed valuable for users to be informed on these topics as they could have a direct impact on their use. Data privacy is also a major concern in the digital age, and has been specifically cited in publications as a topic that should be taught to students (Huang, 2023), and was therefore included as well.

Once the problem space has been clearly identified the next step in the GAMED framework is game idea generation. This is a critical step in the process as the chosen game will provide a platform to facilitate the teaching of the aforementioned topics. An idea discussed earlier in this paper was the concept of conversational learning. The developers felt that based on existing ethics and gamification literature this style of learning had the potential to be effective, and could be implemented into a game. The idea was to create a game in which the player would interact with various NPCs, and that these interactions would simulate a conversation. These conversations would be the primary feature of the game, and they would help players strengthen the players' understanding of AI ethics and develop the related critical thinking skills. To create a more game-like feel, the developers wanted to create a map that players could explore freely and incorporate popular gamification elements such as experience points (XP).

GAMED suggests drawing inspiration from other sources to help the ideation process. This led to inspiration being taken from the popular video game franchise Pokémon. Pokémon features a large map for the player to explore. Players can also interact with various NPCs, often in the form of a battle. The idea was to transform these battles into conversations that help the player understand the deep structure of various ethical dilemmas. At this point, the developers had a fairly clear vision for what the game would look like, and how it would function. This marked the end of the preparation and ideation phase, and allowed the developers to move to the technical development of the game.

At this point it is important to reiterate that the goal of this study was not actually the creation of a publishable product, but rather the knowledge gained through the development and evaluation of this game. Therefore, the development of this game does not always follow a standard software development approach, but instead favors a human-computer interaction (HCI)

style research approach. RtD is commonly implemented in the context of HCI research (Bardzell et al., 2015), and the outcome of RtD studies often differs from that of software development papers; this may stem from their respective focuses on knowledge gained versus the product created. While this choice may receive some criticism, it was a conscious choice by the developers to try an alternative approach to the development process, and the consequences of this decision, for better or for worse, will be evaluated in the discussion section at the end of this paper. Ideally, this approach will allow for a more nuanced understanding of the problem space, or in the words of Bardzell et al. (2015) it will help “clarify [the problem space’s] complex particularities”. This information could be used as a foundation for future iterations of this game.

This present paper describes the first iteration of what would ideally be a continuing development cycle. In this first iteration, the requirements for the software were quite limited. It was to be a web-based game, designed for desktop, and it must be single-player. The game was programmed in JavaScript (JS), ES6, and uses the PixiJS framework. JS is a light-weight language ideal for creating interactive web pages, and there are plenty of JS frameworks that can simplify the development process. A variety of these frameworks, such as React and Babylon, were considered, but PixiJS was chosen because it is designed specifically for 2d web based games. It simplifies the initialization and rendering of the game window, allows for easy sprite and map implementation, and provides a variety of ways to take user input, such as turning text boxes into buttons.

The system architecture of this game relies heavily on JS’s module functionality, which allows developers to create modules which can be imported and exported into different files. In the context of this game, most of the modules created contained objects. The result is that the game itself is run in the main.js file, which imports these different modules, and can then create

instances of the objects within them - allowing for access to the encapsulated methods. However, this also means that each time the site containing the game is opened all of these modules will need to be loaded, which may cause performance issues. To avoid this, a module bundler, called Webpack, was utilized. Webpack automatically bundles and compiles JS modules, alongside any dependencies. It then outputs this bundle into a single JS file, which can be referenced in index.html to display the game on the site. This reduces the number of requests made to the server when the page is first opened, which allows for faster load times.

These modules are used to enable the player-NPC interactions featured in this game. This functionality is achieved through the use of three modules: AppInfo.js, interaction.js, and NPC.js. AppInfo.js contains an object holding information about the game window, which is then imported into both interaction.js and NPC.js. The NPC.js module is what handles the logic for these interactions. The developers created a file for each NPC containing all their text, and the associated dialogue options for the player. A method within NPC.js will then extract the necessary information from this file, and display the choices available to the player. Interaction.js will then handle the displaying of the NPCs dialogue, in the form of a white text box on the bottom half of the player's screen. This is achieved by using data from the imported AppInfo object to calculate the size and position of the textbox based on the user's window size. This textbox is then filled with the NPCs dialogue, and if necessary, will be split into sections that the user can cycle through by pressing the arrows in the bottom corners of the box. The above figures three and four depict what this interaction looks like to the user.

The way this is implemented in the game is by creating a new instance of the NPC.js object for every NPC. These objects contain a number of methods that allow the programmer to assign different scripts, sprites, and map locations for each NPC. Each of these sprites are then

‘anchored’ to their respective map locations. Movement in the game is then achieved by moving the map around behind the player, and by anchoring the NPCs to specific points they are moved along with the map, creating the illusion of movement. Distance between the player and the various NPCs is calculated using euclidean distance, and once the player is close enough text will appear above the NPC prompting them to “Press ‘Enter to Talk’”. This same distance calculation is used for the collision detection system that prevents the player from walking off the island, or through trees and houses. When the map was initially created a series of boxes were drawn representing boundaries that the player cannot walk past. These boxes are programmatically converted into a list that holds their respective locations on the map, and then redraws them as invisible boxes. When a player gets too close to a box the game will disable movement in that direction, depending on the directional key currently being pressed, until the player is no longer adjacent to the box. All these systems work in tandem to create a game in which the player is able to explore the map freely, and engage in ‘conversation’ with different NPCs.

Currently, this game is considered a prototype. There are incomplete features, and some bugs. For example, in the upper left-hand corner of the screen there are three bars that track the XP for the AI ethics, XAI, and data literacy. Players earn XP based on certain metrics recorded during their interactions with NPCs. Each NPC has a predetermined amount of XP available, and the player will receive a percentage of this XP corresponding to the percentage of questions they answered correctly during the interaction. While this system does work its functionality is currently limited. It is useful for giving a visual gauge of the player's success, but in the future it could be used in more ways, such as restricting access to certain NPCs until the player gets more XP in that subject.

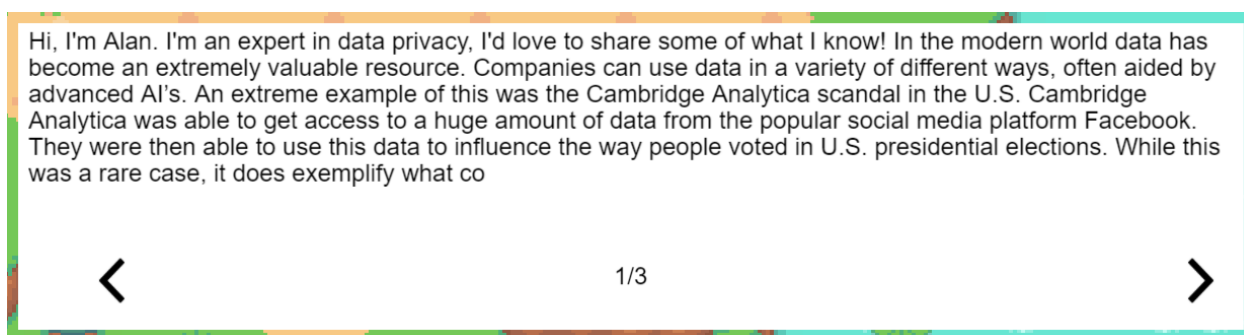
Otherwise, there are some small, but not game breaking, bugs in the system. There are imperfections in the collision system which can lead to the players movement being stopped even when they are not colliding with anything, however, pressing a different directional key will terminate the false collision and allow them to move in the initial direction. Another issue pertains to the displaying of the users dialogue choices. Sometimes, the system will overlay choices on top of each other. Although, all the choices are still displayed, and can be pressed. An example of this can be seen in Figure 3.

Figure 3



Currently, it is unclear what the cause of these glitches is. Finally, when the interaction module displays the NPCs text words are sometimes cut off in the middle. This can be seen in Figure 4.

Figure 4



mpanies are able to do with people's data in the modern age. AI systems require vast amounts of data to be effective, and this has led to various approaches to data collection. Every time someone visits a website, that site has access to a large amount of data, including information such as IP addresses, which show the user's current location. Various laws exist to protect users and their data, unfortunately, they rely on traditional informed consent methods. Except these methods are not effective in the digital world. Almost every time a user opens a new website they are faced



2/3



Note. In this example the word companies has been cut into 'co' and 'mpanies'.

Evaluation

This paper is part of a larger project consisting of four studies. Two of these are focused on the evaluation, and this paper draws on their results to determine the potential value of gamification in the context of AI ethics education. Of these two studies, one is an interview based study focused on qualitative research, and the other is a quantitative analysis of survey data. It is important to note that this present study did not conduct any data collection, and only received fully anonymized data from these two studies. The interview questions and the survey can be found in the appendix (Appendix 2 and Appendix 3, respectively), but the information and debrief sheets are not included as this study did not play any role in the collection of this data.

The focus of the interview based study is on how participants' understanding and attitudes towards AI change once they play the digital education game that was developed. This is achieved through a two stage interview process, taking place before and after their play through of the game. The interviews were conducted with five undergraduate students. The developers of this game created it with the intention of providing an enjoyable and engaging learning experience that taught players about ethics of AI, and helped them build relevant critical thinking skills. By reviewing the transcripts from the post-playthrough interviews this study will gauge

the games success in these aspects, and will also be reviewing what criticisms that participants suggested. These interviews will be analyzed using a theoretical thematic analysis as outlined by Braun & Clarke (2006), and will be taking a semantic approach in the identification of themes.

To provide a more concrete measure of the effectiveness of gamification compared to more traditional, text-based, learning the second study used surveys to conduct A-B testing. One group of participants was asked to read a document that contains the same information in the game, and the other group was asked to play the game. Both groups then completed the same survey. This survey contains a questionnaire that is a validated measure used to assess people's AI literacy, and includes AI ethics questions (Hornberger et al., 2023). The before and after scores of each group will be analyzed with a paired samples t-test to see if the learning interventions had a significant impact on the scores. Then an independent t-test will be used to determine if there is a statistically significant difference in performance between the two groups. This will provide a clearer picture on the effectiveness of gamification in teaching about AI literacy, and consequently AI ethics as its a sub-field of AI literacy. Having this data is important because it provides a numerical measure of the outcomes of this education game, and can be used to determine if gamification is providing any measurable benefit compared to more traditional pedagogies in terms of learning outcomes.

This present study will be collating the results of both studies to better understand the potential role of gamification in AI ethics education. The qualitative data gathered from the interviews will also be valuable in assessing RtD as a methodology within the present context. One of the major critiques of RtD is the difficulty verifying any knowledge gained. However, Bardzell et al. (2015) argued that an important aspect of RtD is the reaction of others when presented with the product. Therefore, it is possible, at least to an extent, to gauge the validity of

the findings through the response of the interview participants. With this in mind, the final stage of the evaluation for this study will be a reflection on RtD as a methodology. This will involve considering the merits and obstacles of RtD throughout this study, and a discussion on the potential role of RtD in future AI ethics education research. The conclusions drawn from this reflection will be more akin to a series of hypotheses, and should be considered as a theoretical basis for future research - these conclusions are not empirically validated and may have been influenced by the developers' biases.

Ethical Measures

This study received ethical approval from University College Cork's Research Ethics Committee for the School of Applied Psychology on the 8th of February, 2024. This study took all the necessary precautions to ensure the well-being of the participants. This included providing participants with information sheets prior to the interview or survey, and debrief sheets following their completion. The debrief sheets provided them with contact details for the relevant parties, such as the data controller and study facilitator. An extensive Data Management Plan (DMP) was detailed prior to any data collection, and this ensured that all participant data was handled with proper care. All data was anonymized prior to analysis, and stored on a secure platform. Non-anonymized data was deleted if participants did not choose to withdraw within the two weeks following their participation in the study.

Results

This study presents a multi-faceted approach to answering its two proposed research questions: what is the potential of gamification as pedagogy in teaching AI ethics, and what is the value of a RtD approach to AI education research? To answer these questions this study draws on three sources, information gained from development via RtD, a thematic analysis of interviews, and a quantitative analysis of surveys. The results of each will be reported independently in this section. The interpretation of these results within the context of one another, with the goal of answering the two research questions, will take place in the discussion section.

Insights from RtD

Through the process of designing and developing this game the researchers drew three main insights about gamification, and its role in the context of teaching AI ethics.

1. A central component of gamification is enabling the player to have an active role in their own learning.
2. The game created is used to engage and motivate players, but the desired learning outcomes are actually achieved through conversational learning. The learning experience created is a hybrid of these two pedagogies (gamification and conversational learning).
3. The conversational learning in this game features a feedback loop that engages critical thinking and reflection on one's own conclusions, the game is only instrumental in facilitating this loop.

A central component of gamification is that it shifts users' learning experience from passive to active, and allows the player to be involved in their own learning. Throughout development it became clear that the major focus was creating an experience that the user could directly

influence, and would allow them to express their own thoughts on different situations. This type of active learning has been cited as an important part of teaching critical thinking (Willingham, 2019), and appears to be one of the major benefits afforded by gamified learning experiences.

As the team pursued this goal of creating an active learning experience they decided that this could be achieved through a multiple-choice style dialogue system, in which the player has a ‘conversation’ with a NPC. Throughout the development of the game it became clear to the developers that the game was a platform to enable these conversations. For most of this paper gamification has been referred to as the pedagogy being used, but as the game continued to take shape it became clear that the conversational learning was actually accomplishing the teaching. Content was being delivered to the players through these conversations, and the gamified aspects were utilized to increase engagement and motivation. The game created in this paper is a hybrid of two pedagogies, conversational learning and gamification. At the onset of this project the developers had not fully understood the critical role of the conversations in this game, but throughout the development the conversations became the central pillar of this learning experience.

This led to the team to analyze these conversations more deeply to determine how they could be used to facilitate learning and critical thinking. The researchers found that these conversations were creating a feedback loop. The player would be presented with an ethical dilemma, and would be given multiple options from which they could choose what they considered to be the ‘right’ one. This would force the player to critically evaluate each option, and come to a conclusion independently. The player can then choose what they personally feel is right, and the NPC will respond accordingly. The NPC will provide feedback about the player's choice, by pointing out potential flaws and possibly unconsidered implications. Then the NPC

will continue and ask more questions about other situations. This creates a loop consisting of critically evaluating a situation, sharing an answer based on this, receiving feedback, and then trying again. This represents the underlying basis upon which this game functions, the game is simply a way of facilitating this loop. If this loop were created through other means the outcome should remain the same. However, these insights are purely theoretical, and need to be verified by feedback from users to see if this cycle of critical thinking and consequent learning is actually being achieved.

Results From Interview Data

A review of the transcripts from the post-playthrough interviews provided insights into what users experienced during their playthrough, and how this aligned with insights gathered during development. The goal of this analysis is to review the players' experience in terms of enjoyment and engagement, information learned, use of critical thinking, and criticisms of the game.

Most of the players did mention that they enjoyed playing the game, at least to some extent, and all five participants agreed that it provided a more engaging experience compared to just reading the information from a document. Participant four shared their thoughts on this by saying “it was really fun i really liked the the layout, I liked that you could go in any order you wanted yeah no it was... I thought it was really engaging”. It seems that this primarily stemmed from the ability to explore the map, and that breaking up the text and involving the player made a big difference. Participant two even stated they would not have played the game had it been too much plain text:

Yeah I did feel more engaged and it was definitely.... I mean, if it was just text, I probably would have closed the game. But yeah, because there was the element of oh I can continue here and I can answer questions and try to get them right was rewarding alright. The gamification elements of the experience appeared to not only engage the participants, but actually motivated them. Participant five said it made them want to talk to more NPCs and correctly answer any question:

I do think like the XP like that was like a motivator for me to like, talk to other people and like to explore the map more and talk to more people and like, try and get the answers correct.

This is promising feedback, and indicates the game does provide certain benefits over more standard teaching methods.

However, researchers also wanted to verify that the participants were actually learning information through these conversations. The analysis of the transcripts revealed that the main learning outcome for players was a new perspective on AI ethics, or as participant one put it “it sort of broadened my horizons on it”. In general, it seemed that the participants had not considered these issues before, or if they had, only briefly. The game helped them consider these issues more in-depth. When participant five was asked if they had given any thought to these issues previously they replied:

Well no, not like that sort of like... blatantly, you know, like I hadn't, like, had a question like proposed to me like that. Like it's been in the back of my mind, you know? But like, when it was actually like, you know, proposed to me, it was like it did make me think a lot, and it made me think like, ohh. Like, this is actually a really strange one. Like, who gets the money for this, you know?

Other participants responded similarly when asked if their knowledge on the topic of AI had changed. They would often cite an ethical issue discussed in the game, and say that the game had helped them understand and think about the issue more clearly.

This also indicated that players were being required to critically think about these situations. A major goal for this game was to facilitate conversations that required some level of critical thought. Based on the responses from the participants it does seem that the game was helping them think more deeply about certain issues. Two participants even explicitly stated that the game facilitated critical thinking. With participant three saying “it gave good examples and stuff you know and made you choose which one do you think so it kind of definitely got you thinking more critically about it you know”, and participant four said “...I think it would help people think a little bit differently about AI. A little bit more critical thinking maybe, help them think more outside the box”. This shows that the game was successful in its attempt to make players critically evaluate situations.

While the game was successful in many aspects, researchers also wanted to analyze the major criticisms of the game. There appeared to be two major problems with the game, there was too much text, and it did not always feel like a game. Four of the five participants mentioned at least one of these two issues when asked their thoughts on the gamified aspects of the experience. One participant stated “there was very little about it that made it feel like a game to me”. This feedback made it clear that while the game seems to be accomplishing most of its major goals, it still has significant flaws. This highlights an important point for the findings of this research. A large part of this study focused on facilitating conversational learning, and while it appears that this did result in the desired learning outcomes, this style of learning may not actually feel like a game. This is important to consider for future research on the topic.

Results From Survey Data Analysis

The final evaluation of this game was a quantitative comparison of its effectiveness versus a more traditional text-based approach to teaching. A validated AI literacy test was used to assess participants' knowledge (Hornberger et al., 2023). Participants ($N = 62$) were asked to fill out a survey containing this test. Half of the group was then asked to play the game, and the other half was presented with a text-only document containing the same content taught in the game. Upon completion of these tasks they were then asked to fill out the AI literacy test once more. The average score on the AI literacy test was as follows. For the gamification group ($N = 31$), the average prior to playing the game was 45.01%, and after playing the game the average was 54.06%. For the traditional text based group ($N = 31$), the average before reading the document was 44.58%, and the average after reading the document was 50.83%. A paired samples t-test was used to compare the before and after scores for each group, and found a $p < 0.01$ for both groups. This shows that both groups had a statistically significant increase in score on the AI literacy test, following their respective learning experiences. However, an independent t-test comparing the scores between the two groups resulted in $p = 0.278$. This means there is no statistically significant difference in performance between the two groups.

Discussion

Gamification and Conversational Learning in AI Ethics Education

This study employed a RtD approach to the development of a digital education game. The aim of this research was two-fold: explore the potential of gamification as a pedagogy in teaching ethics of AI, and better understand the value of RtD in this field. The development team created a single-player online education game in which players could interact with a series of NPCs. These interactions were designed to emulate a conversation, and teach players about ethics of AI while developing their critical thinking skills. Throughout the development of this game a RtD approach was maintained, and allowed the developers to better understand how gamification could be used to teach ethics of AI. This resulted in three main findings (described in more detail in the results section): it is important that people have an active role in their own learning, the game could be used as a platform to facilitate conversational learning, and this conversational learning utilized a feedback loop to teach players about ethics of AI and engage their critical thinking skills.

However, these findings stemmed from RtD, a methodology often criticized for its epistemological founding (Godin & Zahedi, 2014), so they needed to be verified with further data. Bardzell et al. (2015) suggested that the findings from RtD can be validated through the reaction of others. In the case of this study, five undergraduate students were asked to play the game, and were then interviewed about their experience. An analysis of these interviews confirmed that the findings from RtD were reflected by the players. The participants stated the game was effective in keeping them engaged and motivated, and felt more comfortable discussing AI ethics following a play-through of the digital education game. Two participants

explicitly stated that they felt it required critical thinking, and the other participants indicated that the game changed the way they thought about AI ethics.

The initial goal of this study was to assess gamification as a pedagogy, but these findings found that the product created was actually a hybrid of two pedagogies, gamification and conversational learning. This was a major shift in the focus of this study, as initially the scope was limited to gamification, and the introduction of conversational learning occurred later during the sociotechnical design of the game. Both of these pedagogies have a long standing history, and have been shown to be valid and effective forms of teaching (Sims, 2004; Zeybek & Saygı, 2024). They have also independently been used in ethics education, conversational learning has been shown to be successful in teaching business ethics (Thomson, 2011), and gamification has had success in tech ethics (Canosa & Lucas, 2008). This present paper suggests that the combination of these pedagogies results in an educational tool that is able to effectively teach ethics of AI and foster the associated critical thinking skills. The digital education game created in this paper provides a preliminary prototype for how this could be achieved. The game presented in this paper provides a platform to facilitate conversational learning. This results in a unique experience that draws on the benefits of both gamification and conversational learning. Gamification has been shown to increase engagement and motivation (Rabah et al., 2018), and conversational learning can improve ethical assessment and critical thinking skills (Thomson, 2011).

In the aforementioned interviews it was found that the game was effective in engaging and motivating players, teaching them about ethics of AI, and required critical thinking. A validated AI literacy test, proposed by Hornberger et al. (2023), was also used to test for changes in AI literacy after playing the game. It was found that after playing the game participants

showed a statistically significant improvement in scores. However, when compared to traditional text-based learning there was not a statistically significant difference in improvement on the AI literacy test. All participants, regardless of whether they had a text-based learning intervention or played the digital education game, showed statistically significant improvements on the AI literacy test. This reveals that the game is effective in teaching AI literacy, but is not necessarily better than traditional methods. Although, participants that took part in the game based learning intervention had a slightly higher average score improvement (~3%) in the pre- and post-intervention AI literacy test, compared to the text-based learning intervention, the difference was not statistically significant. This indicates that the digital education game proposed in this study is an effective way of teaching, but requires further improvements to be considered more valuable than traditional text-based teaching. However, this does still show promise for the combination of gamification and conversational learning in AI education.

There is some research on the combination of these two pedagogies, but according to a recent publication by Benner et al. (2022) there are only 13 papers on the topic which “may signify that research on this topic is still in a rather early stage, thus requiring further research” (Benner et al., 2022, p. 13). While the current literature is limited, the findings thus far do align with those proposed in this paper (Fadhil & Villafiorita, 2017; Benner et al., 2022). However, none of the existing publications are focused on teaching ethics or AI literacy. Therefore, at this time, this study appears to be the first one to explore combining gamification and conversational learning to teach AI ethics. While the findings presented from this research are promising, they are intended to serve as a foundation for future research on the topic, and should not be considered as conclusive findings.

This study was the first iteration in what would ideally be a continuing development cycle. RtD can provide valuable insights into a problem space (Bardzell et al., 2015), but to create a more complete, and theoretically grounded, product more development cycles are required. While the response to the game was mostly positive, there was some valuable criticism from the interview participants. Currently, it lacks enough gamified elements to make it feel like a game. Certain aspects, such as the XP and the ability to explore a map freely, were found to make the experience enjoyable and engaging, but participants still said it did not always feel like a game. Another issue is the lack of theoretical grounding for the NPC interactions. Future versions could draw on experiential learning and cybernetics systems theory to improve the effectiveness of these conversations, in terms of learning outcomes and the development of critical thinking skills. The digital education game developed in this paper provides proof of concept for the combination of gamification and conversational learning in teaching AI ethics, and lays the groundwork for future research in this area.

RtD as a Methodology in AI Education Research

This paper had two primary objectives, explore the potential for gamification in AI ethics, and evaluate RtD as a methodology for conducting research in this field. This paper provides a critical reflection on the use of RtD throughout this study, and hypothesizes about the potential value of RtD in the space of AI ethics education research. One of the major benefits of RtD is that it allows for a deeper understanding of the problem space (Bardzell et al., 2015). This can be incredibly useful when exploring new approaches to a problem. In the case of this paper, it led the developers to the unique combination of gamification and conversational learning. RtD can help the developers better understand the end-goal of the product they are creating, and provides a new perspective on the problem being solved. This creates a strong foundation for the

development, and can help the designers align the products functionality more closely with the users needs.

However, this paper finds that the products created through RtD should be viewed as prototypes, and are not suitable as final products. This is consistent with long-standing criticisms of RtDs validity (Godin & Zahedi, 2014). Therefore, this paper argues that RtDs value is dependent on the study's goal. If the objective is to gain knowledge about a particular problem space, RtD is valuable, but if the objective is the development of a publishable product, RtD is not suitable due to the lack of quality assurance measures. Ideally, RtD would be implemented in the early stages of a multi-stage development process. Once the problem space has been understood, and an initial prototype has been developed, there should be a shift from RtD to more standardized development approaches. For example, in the case of this study, the next iterations could more closely follow standard software development guidelines, such as stricter requirements, to ensure that the final product is of the necessary quality to be implemented into an educational setting.

Limitations

This study was subject to a number of limitations. Primarily, sample size and time constraints. The qualitative data used in this study stemmed from interviews with only five participants. A strict time schedule also limited the study to a single design iteration, and may have impacted the quality of the RtD approach implemented. RtD is an evolving process which requires continuous reshaping and realigning of the product being created. This study also did not have significant documentation in regards to the evolution of the product being created. It is recommended that RtD studies should strictly document the entire design process, so that more detailed and insightful inferences can be made. Other evaluative methods should also be

considered to better assess the learning outcomes of this digital education game. For example, there may be better ways to assess critical thinking, which could provide valuable insights into the effectiveness of the proposed game. Future studies should consider incorporating alternative methods of measuring learning outcomes.

Conclusion

By utilizing a RtD approach to the development of a digital education game this study was able to explore the potential for gamification in teaching ethics of AI, and through a critical reflection hypothesized on the role of RtD in this field. This study found that there is potential to use a combination of gamification and conversational learning to teach AI ethics. By combining these two pedagogies it is possible to create an experience that engages and motivates the player, through elements of gamification, and also fosters ethical assessment and critical thinking skills, through the use of conversational learning. The game developed in this study provides proof of concept for this idea, and provides an example of how these two pedagogies could be combined. A quantitative analysis did show that this game was effective in increasing players' AI literacy. However, there is a need for more design iterations and improvements to the current game. Currently, it does not seem to provide a significant benefit compared to traditional text-based teaching methods.

A core aspect of RtD is the continual redesign and realignment of the artifact. In this case, the game created would ideally undergo further development iterations. These could involve adding more game elements, and more concretely grounding the applied conversational learning in existing theories, such as experiential learning and cybernetic systems. The utilization of the RtD approach helped this study reach this unique combination of pedagogies, and appears to be a valuable research methodology for AI ethics education, but should be used to help researchers explore new approaches to problems. It is not suited to studies that intend to create a publishable product, as RtD lacks certain quality assurance measures, such as development of software requirements.

As research into the field of AI ethics education continues it is important to continually critique not only the products created, but the methods applied. It is vital that researchers continue to keep up with the rapid pace of technological advancement in the modern world. As the field of AI ethics continues to evolve, so should the field of AI ethics education. Researchers should remain proactive in the exploration of new pedagogies and research methodologies. This present study aims to lay the groundwork for further research into the combination of gamification and conversational learning in AI ethics education, and hopes to encourage further research into other novel pedagogies.

References

- Aslan, S., & Balci, O. (2015). GAMED: digital educational game development methodology. *Simulation*, 91(4), 307-319.
- Baker, A. C., Jensen, P. J., & Kolb, D. A. (2005). Conversation as experiential learning. *Management learning*, 36(4), 411-427.
- Bardzell, J., Bardzell, S., & Koefoed Hansen, L. (2015, April). Immodest proposals: Research through design and knowledge. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 2093-2102).
- Benner, D., Schöbel, S., Süess, C., Baechle, V., & Janson, A. (2022). Level-up your learning—Introducing a framework for gamified educational conversational agents. *Wirtschaftsinformatik 2022 Proceedings*. 5.
- Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1, 61-65.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Brinkman, B., & Miller, K. W. (2017, March). The code of ethics quiz show. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (pp. 679-680).
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big data & society*, 3(1), 2053951715622512.

- Canosa, R. L., & Lucas, J. M. (2008, March). Mock trials and role-playing in computer ethics courses. In *Proceedings of the 39th SIGCSE technical symposium on Computer science education* (pp. 148-152).
- Capgemini. (n.d.). Why Addressing Ethical Questions In AI Will Benefit Organizations
<https://www.capgemini.com/insights/research-library/why-addressing-ethical-questions-in-ai-will-benefit-organizations/>
- Du, X., & Wang, X. (2023). Play by Design: Developing Artificial Intelligence Literacy through Game-based Learning. *Journal of Computer Science Research*, 5(4), 1-12.
- European Parliament. (2020, September 4). *What is artificial intelligence and how is it used?*
<https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used>
- Fadhil, A., & Villafiorita, A. (2017, July). An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 408-412).
- Fiesler, C., Garrett, N., & Beard, N. (2020, February). What do we teach when we teach tech ethics? A syllabi analysis. In *Proceedings of the 51st ACM technical symposium on computer science education* (pp. 289-295).
- Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers series*, 1(1).
- Godin, D., & Zahedi, M. (2014). Aspects of research through design: a literature review.
Design's Big Debates - DRS International Conference 2014, 16-19 June, Umeå, Sweden.

- Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5, 100165.
- Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, 16(2), 2577-2587.
- IBM. (n.d.). *What is AI ethics?* <https://www.ibm.com/topics/ai-ethics>
- Jonas, W. (2006). Research through DESIGN through research - a problem statement and a conceptual sketch. *Paper presented at the Design Research Society International Conference, Lisbon, Portugal.*
- Koravuna, S., & Surepally, U. K. (2020, September). Educational gamification and artificial intelligence for promoting digital literacy. *In Proceedings of the 2nd International Conference on Intelligent and Innovative Computing Applications* (pp. 1-6).
- Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, 18(1), 271.
- Lee, Y. C. J., Cowan, A., & Tankard, A. (2022). Peptide Toxins as Biothreats and the Potential for AI Systems to Enhance Biosecurity. *Frontiers in Bioengineering and Biotechnology*, 10, 860390.
- Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. *In Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).

- Madiega, T. (2021). Artificial intelligence act. *European Parliament: European Parliamentary Research Service*.
- Masters, P., Smith, W., Sonenberg, L., & Kirley, M. (2021). Characterising deception in AI: A survey. In *Deceptive AI: First International Workshop, DeceptECAI 2020, Santiago de Compostela, Spain, August 30, 2020 and Second International Workshop, DeceptAI 2021, Montreal, Canada, August 19, 2021, Proceedings 1* (pp. 3-16). Springer International Publishing.
- Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2023). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, 28(7), 8445-8501.
- Pinski, M., Haas, M. J., & Benlian, A. (2024). Building Metaknowledge in AI Literacy—The Effect of Gamified vs. Text-based Learning on AI Literacy Metaknowledge. In *Proceedings of the 57th Hawaii International Conference on System Sciences*.
- Plato. (2007). *The Republic*. (D. Lee, Trans.; 2nd ed.). Penguin.
- Rabah, J., Cassidy, R., & Beauchemin, R. (2018, November). Gamification in education: Real benefits or edutainment. In *17th European Conference on e-Learning, Athens, Greece* (pp. 489-497).
- Ryan, M. (2020). In AI we trust: ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, 26(5), 2749-2767.
- Sharples, M., & Ferguson, R. (2019). Pedagogy-informed design of conversational learning at scale. In *CEUR Workshop Proceedings* (Vol. 2437).

- Sims, R. R. (2004). Business ethics teaching: Using conversational learning to build an effective classroom learning environment. *Journal of Business Ethics*, 49, 201-211.
- Thomson, G. S. (2011). Good conversations: An enhanced model to teach business ethics. *Journal of International Education Research (JIER)*, 7(1).
- Willingham, D. (2019). How to teach critical thinking. *Education: Future Frontiers Occasional Paper Series*.
- Zhou, J., & Chen, F. (2023). AI ethics: From principles to practice. *AI & SOCIETY*, 38(6), 2693-2703.
- Zimmerman, J., & Forlizzi, J. (2014). Research through design in HCI. In *Ways of Knowing in HCI* (pp. 167-189). New York, NY: Springer New York.

Appendix 1: Digital Education Game Live Link & Code

A live version of the digital education game presented in this paper is available to play at:

aidanr56.github.io

All the code for the game can be found in the following GitHub Repository:

<https://github.com/aidanr56/aidanr56.github.io/tree/main>

Appendix 2: Interview Question and Coding Examples

Interview Questions

The questions below served as an outline for the semi-structured interviews conducted. It is important to note that this present study did not conduct these interviews, and simply used the data from another study that was a part of this project. Therefore, the information and debrief sheets have not been included.

1. When you think of the term Artificial Intelligence, what comes to mind?
2. In what ways, if any, do you feel your knowledge/understanding of AI has changed?
 - *Can you pinpoint any areas of the tool that contributed to this change?*
 - *How do you feel about this change?*
3. Would you say you feel more knowledgeable about AI?
 - *Any specific concepts?*
4. Do you think differently about AI compared to before? If yes, how so?
 - *Have your thoughts on the role of AI in society changed?*
 - *Have your thoughts about the ethical implications of AI in society changed?*
5. Do you feel that your attitudes/feelings/opinions towards AI have shifted? If so, how? If not, why do you think that is?
6. Do you think your experience of using this tool will affect your use of AI in everyday life?
 - *Do you think you will use it more than before, or maybe use new tools?*
 - *Have your thoughts changed surrounding how important it is for users to understand how the AI tools they interact with work?*
7. Did the tool address any misconceptions about AI that you may have had?

- *How did your awareness of these misconceptions change or evolve throughout the interaction?*

8. Do you think that this tool has had an effect on how confident you feel about discussing AI?

Example of Coded Transcript

The transcripts from these interviews were thematically analyzed, and coded for the following themes: enjoyment and engagement (yellow), ethics of AI learning (green), critical thinking (blue), criticism (red). Here is an example of this coding:

Interviewer

OK, I just started recording there. First of all, thank you again for taking the time for another interview, the last one. I guess for the record this is just after you've interacted with the tool, yes. So no demographic questions or anything like that. This time, I'm happy to just get straight into. It if you are?

Interviewee

Yep, sounds good to me.

Interviewer

Good. Great. So let's start this off with a question which was actually in the first interview also. When you think of the term artificial intelligence, what comes to mind?

Interviewee

I suppose..... I'm trying to think back to what. I said before. I think if you... my guess would have been to what I said before was just. Straight away, something like. ChatGPT. Some form of

generative AI? I think actually now after. I suppose freshly just after using the tool that you made, I think maybe more. Along the lines of. Extracting data. Learning patterns providing. Information based on that analysed data set, stuff like that. I think that's what comes to mind straightaway when someone says AI now.

Interviewer

Okay. Do you think that those are some areas that you kind of knew about before or are they new to you after you've interacted with the tool?

Interviewee

I think it's. Not something I didn't know about, but it's it wouldn't be something that I knew a whole lot of. I think if someone. Explained it to me a bit like the tool. I'd I'd say, Oh yeah, that makes sense. I. Kind of would have had an inkling about it. But not something I would have been able to tell you. A whole lot About beforehand.

Interviewer

Great. So in what ways, if any, do you feel that your knowledge or your understanding of AI has changed all?

Interviewee

Yeah, I think.... I suppose I am.... Maybe accountability was something I I I noticed from the tool. You know if there was an issue that that arose from someone using some form of AI... You know, maybe if there is to. Be some form of. I think punishment is what was said maybe? that, you know, you'd say, OK, if if this AI inflicts any sort of harm on a user. You know that there's there should be. There should be some form of punishment to potentially the developers, but then

that in itself. Would form some sort of paradox because you know, then you're. Inflicting punishment on the developers, you've now inflicted harm so there should be reprimands for yourself. That's interesting

Interviewer

Is that something you've thought about before in terms of... I guess accountability is a good term.

Interviewee

Yeah, it's. It's something I think if it happened to myself. You'd... I'd probably say, oh, you know that this is all the developers fault. The whole paradox itself, if someone then.... You know you use that as a as reasoning to. To punish the developers, that kind of think that is something that was new to me, I hadn't quite thought about that side of it. So yeah, in that in that way it sort of broadened my horizons on it.

Interviewer

Great. Yeah. Would you say you've any particular feelings about this change, positive or negative?

Interviewee

I think it's a positive. I mean, at the bare minimum I'd like to say that I, you know, have gained a bit more knowledge on the area itself, which I'll always take as a positive. Particularly the fact that it's around, you know, the ethics of AI and. You know. Being in the know from an ethical standpoint. Is always a. Is always a. You know a positive. So for yeah, definitely I would, I would say it had a a positive effect on myself.

Interviewer

Right. Just briefly there before I move to the next question, you said it sort of like broadened your horizons. So you do you think like when you're using AI now you might be thinking about it a little bit differently or considering these issues?

Interviewee

Of yeah, I'd say I would. I'm trying to think again. There was one point, I wouldn't be able to tell you the characters name, my memory on that isn't the greatest. It was along the lines of... Getting AI to, you know, to show its sources, we'll say for the information it's providing you.

Interviewer

Yeah.

Interviewee

Yeah, I think for me anyway, it's definitely something that sort of, you know, comes into the back of the mind when. I used it. Before. Mm-hmm. But never really would have. You know, never really would have thought about, you know.... wouldn't have given it a second thought. I think now, that's definitely something I could foresee myself doing, you know? Making sure that it is credible sources that the AI is actually getting its information from. And just putting a bit more thought into how it works while I'm using it, and thinking a bit more in general.

Interviewer

Great, we've covered a few there now, but just out of interest, is there any sort of applications of AI in society which the tool spoke about that you were not aware of prior to using the tool? I

know generative AI was the most talked about form in the first interviews, I'm just curious if any others come to mind now?

Interviewee

Yeah. Just trying to think there. I think there was 1 character mentioning giving people diagnoses for medical issues? Again, that's something I probably could have thought about, sure, but that definitely sort of, I don't know what the word is.... Definitely sort of shone a light on that area which, once highlighted made sense to me. Definitely not something I've thought about prior to that.

Interviewer

Great. Yeah. And in terms of, you mentioned it briefly earlier there, in terms of like data. I guess how it's being collected, the way it's being used and that kind of the thing. Do you feel like your knowledge of that has changed.

Interviewee

Yeah you know, I suppose most people I would, I would assume, know that these bigger companies are, you know extracting a lot of data from all its users. I hadn't really thought of it before the scenario in the tool where maybe a user felt, you know, that they'd like to withdraw their data, and particularly if it's anonymized. I know for myself and I mean, even yourself, you know, like interviews and stuff.... A lot of the data is anonymous, and if you're doing something like an interview, you know people are given the right to withdraw their information, stuff like that. But when you're dealing with such a vast amount of data and it's all, it's all anonymous. I hadn't really thought about the sort of the issues that would arise if someone got in contact with

the company and said, you know, I'd like all my data removed and stuff like that. Yeah, that kind of thing sort of piqued an interest for myself.

Interviewer

And in terms of the ethical issues with data that's collected and used, the character that talked about the medical diagnosis had a final question. They said something along the lines of “before I leave you, of the four options we gave you for the data that was collected, which one do you think could have ethical implications? The answer was the user's ethnicity, and so I'm wondering is that something you were aware of, had thought of, you know, the fact that it could lead to biases potentially, depending on I guess how the AI is used and applied and programmed?

Interviewee

Yeah to be honest, I hadn't thought of it before it was pointed out to me. Even then, I'm not sure I would have known the answer if I didn't have a list of options presented to me. It makes a lot of sense though when I think about it now.

Interviewer

Okay interesting, so do you think, if that particular character had left it there. You would have even considered one of those four options as being an ethical issue?

Interviewee

No, I don't think I would have

Interviewer

Yeah, like you said, it's easy once they say to you. “Hey, one of one of these four is an ethical risk. Guess which one?”

Interviewee

Exactly. Yeah, you're like, oh, I can see it now.

Interviewer

Sure, yeah. So do you think, if you're using AI in the future let's say, you might be considering what kind of data is being collected and the implications behind it?

Interviewee

Oh yeah definitely, for sure. Yeah, like I said, it's sort of... the the tool itself has just sort of improved my knowledge I suppose. I know for myself, a bit of repetition helps me learn so I'd like to use it again, and it would help me remember things a bit better. But yeah even now, I definitely think I'd be considering more things that I maybe hadn't thought of before

Interviewer

Yeah. Yeah, it was a brief interaction to be fair. So I don't fault you on that.

Interviewee

But yeah, no, I think. Like like I said it's it's improved my knowledge on there so. It could be something that that I would that I would notice it with future AI use.

Interviewer

Great. Now we've kind of stayed on this area for a while now but just one last question on it, but just so you know, we've actually covered quite a few questions from it. So it's don't, don't worry about it. Last one, it's just in terms of, I guess these these different terms or buzzwords could be one way of saying it. So I guess by that I mean stuff like the ethics of AI, the responsibility that

comes with AI, data literacy like you know, understanding your data and ethics of data, all this stuff, all these terms. Do you think knowing those terms has had any impact on you? Or even, could have an impact going forward?

Interviewee

Yeah. No, definitely. Yeah, I think knowing these buzzwords and stuff could. You know, allow me to get more specific and clear cut answer if I'm looking stuff up in the future. Especially if it's something that I'm not quite sure how to explain or I guess...phrase it if I want to Google it or whatever. Yeah I think it would make things easier going forward for sure, yeah.

Interviewer

Great, great. We kind of just covered a lot just now naturally as we were talking, so I'm going to skip a few questions. Just give me one minute there. Do you think that like your, I guess your attitudes towards your feelings towards AI and its use in society, do you think they've changed at all after using the tool?

Interviewee

I'm not sure if my feelings have changed towards it. I definitely think I know more about this, you know. I know more about the stuff we've kind of talked about, like the accountability and the ethical issues that can be raised from AI, and things like that. I still think it's.... I still think it's something that is useful but shouldn't be relied on entirely. I'm hoping that's what I said before

Interviewer

Ah yeah, don't worry too much about that.

Interviewee

But yeah, I definitely I think I think I still sort of firmly lie there. So I probably wouldn't say that my my feelings towards have changed, but my obviously my knowledge on this has.

Interviewer

In terms of being used as a tool but not relied on, could you elaborate a little bit more on what you mean by that?

Interviewee

Uhhh, I think like... It should be used to assist in something as opposed to just doing a job for you. You know, I'm not knocking it's usefulness, like I've used it many a time myself, particularly generative AI. But I would.... you know, I would urge people to use this as a form of assistance rather than substitution for their own work.

Interviewer

Sure, and is there a particular reason that you have for thinking that?

Interviewee

Yeah, I think. You know, as as we were kind of just saying the past 10 or 15 minutes. You know, we don't really know... if these are credible sources that that we're getting the information from. Like we mentioned before, potential biases in the in the data and the effect that can have on the results people are given. See, you know it's... just it's still a bit of a grey area I think. If people are using it without knowing about these biases and they're taking everything it says to be um.... completely the truth and like gospel then there you know...there are concerns there with how it's affecting people's opinions. And then just, you know, like relying on tools like Chat GPT and other generative AI entirely kind of.... I don't think that would have a great impact, a positive

impact even, on ummm.... on how often people think about the task they're doing, or like... evaluate it in their head you know?

Interviewer

Sure okay, yeah interesting. Just to touch on that last point there, could you explain what you mean by that? So like, over relying on a tool like Chat GPT and the... I guess the concern you have for how that affects peoples thinking?

Interviewee

Yeah sure no problem, I mean I guess... I guess for myself personally, I know when I first started using things like Chat GPT, I kinda treated it like.... like a google search, you know? I never really double checked if the information I'd been giving was.... you know....like 100% factual or true or like, where it even came from? And then you know, if you're asking it to like...perform a task for you it just does it so well and its so easy, you know? But if you just have it do absolutely everything for you, I feel like it could ummm... it could, you know, make it harder for you to do these kind of tasks on your own in future cos you've relied on something doing it for you for so long? And it's just so easy to like, ask Chat GPT to do something for you you know. But I think.... I think after using the tool I could see myself being more aware of these things as I'm using AI, and I think it's ummm.....I think it's a good first step towards making people think a little bit more, I could see it helping people in....in that regard I suppose.

Interviewer

Yeah I agree, that's great. Alright, we're almost done here now by the way. Do you think the tool has had an effect on how confident you might feel discussing AI? Talking about it with you know friends, family, just in any sort of setting really?

Interviewee

Ummmmm, yes...I'd say yes. But I wouldn't say, you know.... I wouldn't be totally confident you know. I definitely think I could sort of, you know, explain sort of the basics around it, sure something like that. I think if there was anyone who sort of had a bit of an inkling around the topic and had a question then ummmm.... that's probably an area where it's kind of beyond the realms of what I'd know.

Interviewer

I know, sure. And look like I said, it was a brief interaction, but I guess even in terms of like maybe. putting you on the Right track so?

Interviewee

Yeah, exactly. I definitely think I, you know, I know sort of... a bit more of what I need to know from using the tool. I certainly don't feel like I wouldn't be able to talk about it, you know, but I'm just not sure how.... how much in depth I could go I guess

Interviewer

Perfect. So just finishing up now, just a brief bit on the actual tool itself. Any parts in particular that you liked or that you disliked about the tool?

Interviewee

Yes, I. I did like the.... I know it was sort of based sort of a Pokémon style. Which I did like. Even just being able to sort of control your own character, you know, just adds that sort of... game feel? I'm sure it's probably been mentioned before but there obviously is a lot of text. Which, you know, after you've met the sort of 3rd, 4th character can be a bit tedious. So if there

was. Some sort of. Maybe like some sort of like mini game that could be done with.... with the sort of like , the knowledge that you gain from each character? Something to do with the XP maybe?

Interviewer

Did you find when you kind of.... so the the general theme of it was to, you know, speak to a character, kind of get the information on the topic and then you get multiple choices. You answer the question and they kind of.... they tell you whether you're right or wrong on your answer. Did you enjoy that aspect, or have any thoughts or issues on that?

Interviewee

Yeah I think that.... there were a couple of times where I got the answer wrong I think, but it wasn't quite clear, if you get me? Like, I think maybe if you get brought back to the choices again if you end up ummm...if you end up getting a question wrong? Cos yeah, there was times when I think that would have.... I guess helped me understand the situation a bit more.

I also think when answering the questions, I mean this this is kind of minor stuff. I think at the very start it wasn't clear how I was to answer it. I know now that you click on the answer, but because they were numbered I was kind of thinking.... can I press 1 or 2, you know? There was one point also where someone mentioned a character, I'm trying to think.... Joseph maybe? Anyway, they said "Oh Joseph will tell you more about that", but I couldn't find anyone named Joseph? I don't know, maybe an error or maybe not but.... yeah I couldn't find them anyway

Interviewer

Okay great. So you said there, and look you're absolutely right. It is quite a lot of text which I think for anyone reading a lot of text it is.... it could be quite a difficult task to stay focused. But from the gamification elements that were there, so you know like you've got, you've got an island, a colourful island with interactable characters, you can control your character and all that. Do you think that that had any effect on your ability to kind of stay engaged while reading all this text?

Interviewee

I think it definitely kept me more engaged, yeah. If it was just text with.... you know without all of those elements I think it would have been more difficult. Also with the text, I think... You know, having even something as simple as just a bit of colour or something, you know? I think that could be a good change cos like.... you know the last thing you wanna see is just black text on a white background.

But yeah I definitely think it kept me engaged. You know, like you said, being able to control your own character. You know, walking around on the island, that kind of thing. Yeah, it it definitely ummm.... I definitely didn't get to a stage of boredom where you know, I sort of just wanted to stop, stop playing.

Interviewer

Okay great. So look that's everything I have for you, unless you have any questions for me?

Interviewee

No, we're good.

Interviewer

OK, we'll leave it there then. Thank you again.

Interviewee

Thank you very much.

.

Appendix 3: AI Literacy Test Questions

The below questions are from Hornberger et al. (2023), and is a validated measure for assessing AI literacy. It is important to note that this present study did not conduct the surveys, and simply used the data from another study that was a part of this project. Therefore, the information and debrief sheets have not been included.

1. In which of these areas is AI typically applied?
 - a. - Detecting credit card fraud
 - b. - Cryptocurrency mining
 - c. - Web tracking
 - d. - Encryption for instant messaging services
2. Imagine you are chatting with an assistant on the Internet. How could you proceed to find out whether you are interacting with a human or with an AI?
 - a. - you can save yourself the effort, since you can no longer distinguish between a human and an AI in written communication
 - b. - ask a difficult factual question, since only a human can answer it
 - c. - make a few typing errors in your text, then the AI can no longer understand you, but a human can
 - d. - make an ironic remark, because this is better understood by humans
3. Which of the following systems often use AI?
 - a. -Flight surveillance systems
 - b. -Geopositioning systems
 - c. -3D printing systems
 - d. -Inventory management systems
4. Which of the following interdisciplinary research fields is also a subfield of AI?
 - a. - Blockchain
 - b. - Natural Language Processing
 - c. - Psychology of Learning
 - d. - Bioinformatics
5. What makes AI intelligent?
 - a. - AI can walk and talk
 - b. - AI has an artificial brain
 - c. - AI is at least as intelligent as humans
 - d. - AI acts rationally to achieve a particular goal as well as possible
6. Why do AI systems behave intelligently?
 - a. - they have no feelings that could distract them from their task
 - b. - they think autonomously and pursue their own goals
 - c. - they have been programmed to try to achieve a given goal as well as possible
 - d. - they are built similar to the human brain and therefore have a similar intelligence
7. In AI, a distinction can be made between "weak" and "strong" AI. "Weak AI" refers to AI systems that have capabilities in a limited area. "Strong AI", on the other hand, is said to be capable of a very broad range of tasks, similar to humans. Which of these

examples could be considered strong AI?

- a. - intelligent virtual assistant (e.g. Alexa)
 - b. - fully self-driving car
 - c. - powerful search engine (e.g. Google)
 - d. - strong AI does not exist at the moment
8. What can weak AI NOT do?
- a. - make decisions under uncertainty
 - b. - solve a wide range of tasks
 - c. - solve a task better than a human
 - d. - learn from unstructured data
9. In which task is AI already superior to humans?
- a. - Detecting tumors
 - b. - Programming software
 - c. - Translating novels
 - d. - Designing cancer therapies
10. In which areas are humans still superior to AI?
- a. - Predict extreme weather events from weather data
 - b. - Find a proof for a mathematical theorem
 - c. - Answer quiz questions
 - d. - Play poker
11. What are knowledge representations in the field of AI?
- a. - particular representation of knowledge in order to communicate it to humans through AI
 - b. - sensors that capture information from the environment
 - c. - information about the world that can be processed by a computer
 - d. - an algorithm that generates knowledge from data
12. An AI is supposed to divide images into meaningful parts. Input is always an image. Which input-output pair is **NOT** meaningful:
- a. - Image -> number
 - b. - Image -> vector
 - c. - Image -> image
 - d. - Image -> video
13. How do AI systems make decisions?
- a. - based on mathematical-logical principles
 - b. - based on links defined by programmers
 - c. - based on quantum entanglement
 - d. - based on artificial intuition
14. What is a key criterion for the quality of a model in machine learning?
- a. - it can predict the output values of the test data as well as possible
 - b. - it contains as few variables as possible
 - c. - it is as well adapted as possible to the training data
 - d. - the predictions are as unambiguous as possible
15. How does supervised learning differ from unsupervised learning?
- a. - in supervised learning, the output values of the training data are known
 - b. - in supervised learning, humans must supervise the AI during learning and intervene if necessary
 - c. - in supervised learning all computational steps are documented
 - d. - in supervised learning, stricter legal regulations apply

16. Which statement about the steps in the machine learning process is correct?
 - a. - the steps of the process are based on behaviorist learning theories
 - b. - the development of machine learning is in parts an iterative process
 - c. - the steps of the process of supervised and unsupervised learning are basically the same
 - d. - the steps of the process can be run backwards to generate an (artificial) data set
17. Sort the process steps in supervised learning into the correct order using drag & drop:
 - a. Train model with training data
 - b. Predict test data with the model
 - c. Collect and prepare data
 - d. Divide data into training and test data Calculate accuracy of prediction
18. What should be considered in machine learning when dividing the data into training and test data?
 - a. - the data should be divided into parts of as equal size as possible
 - b. - the data should be randomly divided into training and test data sets
 - c. - the test data should be of higher quality than the training data
 - d. - the training and test data should be as different from each other as possible
19. How can humans influence the outcome of machine learning?
 - a. - calculation of the accuracy of the prediction
 - b. - randomized division into test and training data
 - c. - selection of the model
 - d. - abstraction of the model
20. To what extent can humans influence the outcome of machine learning?
 - a. - Humans can hardly influence the result of machine learning, since it runs automatically
 - b. - The results can only be influenced when selecting the data the model learns with
 - c. - Humans can influence the result during development at several process steps
 - d. - Humans can only influence the interpretation of the results
21. What determines the behavior of AI systems?
 - a. - AI systems strive for autonomy
 - b. - AI systems pursue a goal that has been given to them by humans
 - c. - AI systems perform behaviors randomly
 - d. - AI systems seek out goals independently and pursue them
22. What is the benefit of data visualizations?
 - a. - Maintaining transparency
 - b. - Preparation of training for image recognition
 - c. - Communication of results
 - d. - Conducting statistical tests
23. Why can systems based on machine learning obtain good results?
 - a. - their work is often observed by humans and corrected if necessary ("supervised learning")
 - b. - they think similarly to humans, but faster
 - c. - they can draw conclusions from large amounts of data and thus improve their model
 - d. - they are derived from expert systems in which expert knowledge is stored

24. What data do AI-based recommender systems of streaming services use?
 - a. - the data of every user when using the service
 - b. - all data that a user of the service leaves on the Internet
 - c. - the data of other users, but not one's own
 - d. - only one's own data when using the service
25. You are testing a machine learning model that is supposed to classify photos of animals. You notice that the model is better at recognizing cats than dogs. What could be the reason for this?
 - a. - dogs are more difficult to recognize than cats, since there are fewer images of dogs on the internet
 - b. - small objects (cats) are better recognized than large ones (dogs)
 - c. - most models are generally better at recognizing cats than dogs
 - d. - the training data of the dogs were not representative for all dog breeds
26. Which ethical principles should be considered when developing AI?
 - a. - Holism, fairness, transparency
 - b. - Prevention of harm, transparency, fairness
 - c. - Fairness, holism, prevention of harm
 - d. - Transparency, prevention of harm, holism
27. What is the black box problem?
 - a. - AI entails a residual risk that is hard to calculate
 - b. - it is often difficult to determine how AI makes decisions
 - c. - users are often not informed about the application of AI
 - d. - many users have little knowledge about AI
28. Which societal challenge is frequently mentioned in the context of AI?
 - a. - lack of investment incentives in the educational system
 - b. - chip shortage in industry due to the high computational cost of AI
 - c. - high error rate in AI-enabled manufacturing
 - d. - replacement of human workforce by AI
29. What are central risks in using AI for predictive policing?
 - a. - Vulnerability to hacking
 - b. - Discrimination against suspects based on origin and status
 - c. - Lack of legal certainty in the event of AI breakdown
 - d. - Undermining the authority of police officers
30. Which legal challenges do AI applications entail?
 - a. - Users of AI have no option for legal protection
 - b. - Protecting the rights of AI itself
 - c. - Lawyers do not understand the importance of AI
 - d. - Limited control of AI because of its autonomy