Universiteit Leiden

Master Computer Science

Exploring algorithmic bias through games: A case study on interactive storytelling and its impact on students' understanding

Name:	Yue Li
Student ID:	4005317
Date:	26/09/2025
Specialisation:	Science Communication & Society
1st supervisor:	Jan van Rijn
2nd supervisor:	Giulio Barbero
3rd supervisor:	Marien Raat

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

Algorithmic bias has emerged as a pressing concern as artificial intelligence (AI) increasingly influences decisions in domains such as recruitment, finance, and criminal justice. While technical approaches to fairness have received much attention, less is known about how non-experts perceive and understand algorithmic bias. This study explores the use of interactive storytelling as a science communication tool to foster awareness of AI bias. We designed Echoes of Algorithm, a narrative-driven digital game that places players in the role of a journalist investigating hiring discrimination. Using a one-group pre- and post-test design with university students (n = 42), we evaluated changes in knowledge, fairness perceptions, and behavioral intentions. Results indicate that participants significantly improved their understanding of concepts such as historical and systemic bias, showed heightened sensitivity to the potential unfairness of algorithmic systems, and expressed greater willingness to question and challenge biased AI decisions. The findings suggest that interactive narrative games can serve as effective tools for communicating complex socio-technical issues, bridging knowledge gaps across disciplines, and encouraging more critical engagement with AI. Limitations and directions for future research include extending the game to additional scenarios, exploring long-term learning effects, and enhancing engagement through narrative and design improvements.

Contents

1. Introduction

Artificial intelligence (AI) has been employed in multiple fields such as recruitment, education, and financial services. However, current efforts to promote fairness and equity in AI are still in their early stages, and further institutional change and policy support are needed. [1]. As generative AI and large language models develop, these systems are being used as tools to support decision-making, with the expectation of improving efficiency and consistency. Research shows that public concerns about the risks of artificial intelligence exacerbating inequality and eroding trust are growing, driving increased attention and regulatory demands for responsible AI [2].

At the core of these concerns lies the problem of biased algorithms. A biased algorithm refers to one whose decisions are skewed toward a particular group of people and produces systematic and repeatable errors, even when the input data are unbiased [3, 4]. AI models are usually "black-boxes" for users and even for its developers. It is not always clear nor well understood why a certain AI algorithm made a certain prediction, or how trustworthy the predictions are [5]. Therefore, the increasing use of AI is often driven less by critical understanding and more by a near-blind trust in numbers and automation [6]. One way to mitigate algorithmic bias and protect the public from the commercial marketing of AI companies is to develop a more comprehensive understanding of how people perceive algorithmic decisions and to foster public awareness of the potential risks involved.

Our goal is to enhance public understanding of algorithmic bias and its societal implications. We use games as a tool to help players understand the concept of AI bias, explain the potential causes of bias, and highlight the potential impacts of these automated decisions. Therefore, the main research question is: *How can interactive storytelling game influence students' attitudes and perceptions of algorithmic bias?*

We develop an interactive narrative game that would expose players to the nature of algorithmic bias.[1,2] The game draws on a real example of recruitment discrimination. Interactive narrative games have been widely discussed as a means of increasing user engagement and facilitating the communication of complex concepts [7]. Therefore, we chose this format as a communication tool to help players interact with the situation and visually present abstract content. We also conduct pre- and post-responses surveys to assess the impact on players.

This study makes four main contributions. First, it provides a narrative-based, interactive tool for communicating AI bias to university students. Second, it shows the potential and role of game narratives in shaping a sense of fairness and stimulating critical reflection. Third, it empirically measures the effectiveness of game-based storytelling in communicating AI bias. Finally, it integrates the theory of algorithmic bias with the practice of science communication.

2. Literature Review and Position

This chapter reviews relevant literature on algorithmic bias, public perceptions of fairness, and the use of games and interactive media to communicate these issues, and concludes by positioning the present study within this interdisciplinary context.

2.1. Understanding Algorithmic Bias: Definitions, Sources, and Impacts

Algorithmic bias generally refers to the unfairness or advantage that an algorithm's results place on specific individuals or groups [8, 4]. However, the term "bias" itself has multiple meanings; it can be interpreted as a deviation from a certain standard (e.g., statistical, ethical, regulatory, legal, social, psychological, etc.) [9]. The existence of these multiple standards also comes from different sources at different levels.

Lopez [10] argues that algorithmic bias is not simply a technical error but also a sociotechnical bias. Algorithmic Bias is not simply a problem with the data itself but rather reflects social structures of

[1]All code for the game is available at: https://github.com/Capybaraaa0/Echoes-of-Algorithm

[2]A playable version of the game can be accessed here: https://capybaraaa0.github.io/Echoes-of-Algorithm/

unfairness. For example, even if sensitive variables such as gender or race are avoided directly in modeling, the model can still indirectly infer these social attributes through proxy variables such as location, interests, and language style, perpetuating or exacerbating real-world inequalities. Some health risk prediction models, when using medical expenditures as a proxy for health status, may underestimate the actual needs of certain groups [10]. Furthermore, some argue that there is a lack of consensus on the concept of fairness in the public sphere, with different stakeholders holding differing views [11]. For example, while policies may avoid collecting sensitive attributes, technology requires these attributes to help eliminate discrimination.

In the technical process, algorithmic bias can arise from multiple stages of a machine learning system [12]. At the data level, it can come from incomplete, unbalanced, or unrepresentative training data. For example, if an image recognition system is trained primarily on lighter-skinned faces, its accuracy in identifying people of color will decrease [13]. Secondly, during model design, bias can be introduced through factors such as feature selection, proxy methods, and the deployment environment. Even if the algorithm itself is neutral, using variables that carry social meaning can still unfairly affect certain groups [10]. Furthermore, after the algorithm is deployed, in some cases user behavior can be fed back into the system, creating a feedback loop that can exacerbate existing biases [9]. For example, predictive policing systems often identify high-risk areas with high enforcement activity. This results in increased police concentration in those areas, which in turn reinforces the system to identify the area in a high-risk area.

The impact of these biases can be seen at both the individual and societal levels. Algorithmic bias can lead to unequal opportunities for individuals. Certain groups face unfair treatment based on features such as gender and race, limiting their access to recruitment, credit, and education [4, 14, 15]. In finance, for example, algorithmic bias can lead to unfair loan denials. For individuals, this means they cannot obtain the necessary credit, limiting their life choices and even potentially excluding them from the economic system for a long time [15]. Furthermore, empirical research suggests that interacting with biased AI can exacerbate human biases [16].

At the societal level, algorithmic bias further entrenches structural social inequalities and undermines public trust in AI systems [17, 18]. Furthermore, as the application of algorithmic systems continues to expand, the impact of bias is extending beyond recruitment and credit to criminal justice, medical resource allocation, and other areas. For example, Obermeyer et al. [19] analyzed a case of unfair medical resource allocation algorithms. Because historical medical expenditures were used as a proxy for health risk, patients of color received more than half as many medical resources as white patients with the same predicted risk. Since patients of color already receive lower medical expenditures due to structural issues, the algorithm mistakenly assumes they are healthier, resulting in fewer resources allocated. In the criminal justice field, the COMPAS system was widely used to assess recidivism risk, but investigations have shown that its predictions of "high risk" for African American defendants are systematically biased [20].

2.2. Algorithmic Fairness in Recruitment: Challenges and Critiques

Recruitment is a high-risk sector for algorithmic bias since it has a critical effect on candidates' career paths and social structures. Recruitment systems often rely on historical data, automated scoring mechanisms, and feature selection methods to evaluate candidates, all of which can inadvertently perpetuate or even amplify existing human biases [21]. Excessive representation bias occurs when certain groups are underrepresented in historical data [22]. Furthermore, information contained in resumes, often used for model training, can be correlated with sensitive variables. For example, accents and appearance can be linked to gender, and names can be associated with immigrant background. These indirect proxy features can easily pick up sensitive social traits without meaning, which can contribute to bias [23].

Several empirical cases have clearly revealed the structural biases inherent in recruitment algorithms. For example, Amazon once used a resume screening system that was trained on the company's history data, which showed that it had mostly hired men over the past ten years. As a result, resumes with the

word ”women” were systematically downgraded, which kept many women from being able to compete for many positions [24]. LinkedIn’s advertising algorithm has also been shown to favor high-paying jobs for men, even when they have similar qualifications, thereby exacerbating gender occupational segregation [15]. Furthermore, multiple studies and field tests have shown that the photo, name, or language style used in resumes can significantly influence algorithmic scoring, reflecting implicit biases based on cultural background, race, or gender [25].

Another important concern is that when companies deploy AI recruitment tools, they often mask systemic biases in the name of fairness or efficiency. As Drage [26] criticizes, some AI recruitment companies claim their products can eliminate gender and racial bias. However, this claim actually obscures how technology operates within gendered and racialized power relations. By presenting algorithms as neutral tools, such narratives deflect attention from the mechanisms by which technology reproduces and even reinforces inequality, enabling organizations to blame the technology itself when faced with allegations of discrimination and even hindering responses to recruitment bias. Chen [18] also points out that the perception of high-tech systems as objective further obscures this reality.

In response, many countries and regions have begun to establish legal frameworks and ethical standards to oversee the use of recruitment algorithms. The European Union’s General Data Protection Regulation (GDPR) explicitly stipulates that automated decision-making must be transparent, and individuals have the right to be informed and explained, as well as the right to object to fully automated processing [27]. Relevant US laws, such as the Fair Employment Opportunities Act, require that all recruitment tools must not lead to direct or indirect discrimination; otherwise, employers will be held liable for compliance. For example, in New York City, local legislation directly mandates that recruitment AI undergo third-party bias audits and publicly disclose its core mechanisms to candidates [21].

2.3. Public Perceptions of Algorithmic Fairness

As algorithmic systems enter various industries, their acceptance and application will be hampered if they fail to gain public trust and recognition [28]. A growing body of research suggests that the public’s understanding of fairness extends beyond mathematical definitions and technical references to encompass multiple fairness concepts or social behaviors [29]. Starke et al. [28] note that public perceptions of fairness mainly focus on distributive fairness, procedural fairness, and informational fairness. Distributive fairness explores whether different groups are equitably affected by algorithmic outcomes. Procedural fairness considers the legitimacy, consistency, and social acceptability of the decision-making process. Informational fairness clarifies about whether the public gets sufficient explanatory information.

The public’s perception of algorithmic fairness varies across social groups. Mourali et al. [30] noticed that Democratic supporters are more likely to select algorithms that promote equal outcomes, while Republican supporters prefer algorithms with higher accuracy but with possibly unequal results. Educational level also plays a role. Individuals with higher education levels focus more on the fairness of the process than on whether the outcomes benefit them, whereas those with lower education levels tend to judge fairness by the results [31]. Furthermore, some research suggests that due to historical and ongoing racial discrimination, some social groups have a deep-seated distrust of the healthcare system, and this distrust also influences their perceptions of fairness in algorithmic systems [32].

There are differences in what the public expects and how algorithmic systems are built due to information asymmetry and cognitive paradigm differences. Algorithmic systems are typically treated as ”black boxes” whose inner workings are unknown to the public. This is likely to bring about too much trust or distrust [18]. Even if a system is statistically fair but non-explanatory, it will be felt to be unfair [33]. Therefore, explaining algorithmic choices, providing sufficient notification, ensuring the right to appeal, and maintaining human oversight can help build public trust and approval in the use of algorithmic systems [34].

2.4. Communicating through Games and Interactive Media

In recent years, serious games and interactive media have gradually become useful tools for science communication and education [35]. These media combine entertainment and education, effectively enhancing audience interest and engagement through their high degree of interactivity [36]. Furthermore, serious games transcend boundaries across multiple fields, including science, education, and media, playing a similar role as an intermediary in the dissemination of scientific knowledge [37, 38]. Research indicates that these media can provide platforms for audiences from diverse backgrounds to participate in scientific communication, broadening the audience for science, breaking down existing knowledge barriers, and promoting the spread of scientific literacy among a larger population [39]. Furthermore, these media make science participation more inclusive, engaging not only stakeholders within specific research fields but also hard-to-reach groups such as the general public [40].

Compared to other traditional teaching models or scientific communication methods, games are more engaging [41]. Research has found that games empower players with control and decision-making power, which can increase their learning motivation and encourage them to explore and solve problems proactively [42]. A large-scale survey of 705 MOOC learners further demonstrated that autonomy, curiosity, and self-directed exploration play a crucial role in maintaining learning motivation and engagement [43]. Furthermore, games enhance motivation through goal-driven learning and immediate rewards through task feedback [44]. These immediate feedback mechanisms (such as scores, hints, and milestone achievements) can reinforce correct behaviors, correct errors, and provide players with a sense of accomplishment [45].

Games can induce deep emotional involvement, thus improving the effectiveness of science communication [46]. Emotional elements can influence science communication, particularly when addressing socially controversial or ethical scientific issues [47, 48]. Through character design, situational dilemmas, and narrative techniques, games can evoke players' empathy for scientific issues, fostering a resonance of values [49]. For example, in games exploring issues such as climate change, public health, and environmental governance, players are often placed in specific roles and must consider the consequences of decisions made within limited resources [50, 51, 52]. During the COVID-19 epidemic, games such as Plague Inc. have been used to help players to think about the connection between games and the real world because of their immersive nature [53, 54].

Games also have limitations in science communication. Serious games may face challenges in design, resources, and teacher training during implementation [55]. For example, some games do not fit well with the course content, leading to inconsistencies between the learning paths and the assessment objectives, thus weakening the systematic nature of knowledge structures. Furthermore, educators' acceptance of game tools, their ability to use them, and their time commitment can also impact the effectiveness of games [56].

Furthermore, balancing the scientific and entertainment aspects of games is a core challenge in their communication effectiveness. Some research suggests that the term "serious game" itself is an oxymoron [57]. Some games, when overly educational, undermine the principles of freedom and fun. However, games that neglect aesthetics and narrative can also lead to a dull and tedious experience [58].

2.5. Position

As summarized in the literature review, existing studies on algorithmic bias have explored its definitions, sources, and social consequences across domains such as recruitment and credit scoring. Research has also examined how the public perceives fairness in algorithmic systems and how interactive media can support science communication. Although such studies highlight the systemic and multi-layered origins of bias, they scarcely engage with how lay individuals encounter algorithmic injustices or how awareness can be promoted outside expert circles. Research has found that while companies claim to prioritize fairness in recruitment, they often retain some structural discrimination or inequality in practice [26, 18]. This highlights the need for communication strategies that can expose these biases.

At the same time, research on science communication has pointed to the potential for games and

interactive media to engage people, elicit emotional responses, and model rich systems [7, 35]. There have been experiments with serious games in areas such as climate change, health, and environmental governance [50, 51, 52], and their use for communicating algorithmic fairness and bias has been limited. Quite a lot of research on them has been speculative and few experimental studies have tested their impact on knowledge, attitudes, and behavioral intention.

The present project attempts to balance the scales by designing and testing an interactive story game, called "Echoes of Algorithm." The game immerses players in a scenario of hiring bias. Building on previous research that has highlighted the educational and communicative potential of serious games, the evaluation design adapts fairness perception measures [5, 59] and applies a mixed-method pre-/post-test framework to assess not only knowledge gains but also changes in attitudes and behavioral intentions.

Thus, this research is situated at the intersection of computer science, social science, and science communication. It seeks to explore how interactive narratives can help people reflect on the social and ethical aspects of algorithmic bias. Instead of aiming to solve technical problems, the project focuses on how such issues can be communicated and experienced through game play. This study hopes to encourage more inclusive and reflective discussions about the role of algorithms in society.

3. Method

3.1. Research Design and Participants

This study adopted a mixed-method experimental design. It combines an interactive digital game with pre- and post-game surveys. It aims to evaluate changes in participants' knowledge, perceptions of fairness, and behavioral intentions of algorithmic bias. This approach focuses on the behavioral impact of algorithmic bias in information systems research, which means developing not only technical correction methods but also examining people's perceptions and behavioral responses.

To achieve this goal, we used a one-group pre-test/post-test quasi-experimental design, a method for evaluating data in real-world settings where random assignment of participants is not feasible [60]. Specifically, the design used within-subject comparisons, measuring and comparing the same participants before and after a single intervention involving the game. This approach allows for a degree of control for confounding by individual differences (such as background knowledge and personal experiences). This is because each participant serves as their own control group. This approach reduces measurement error and is more sensitive to whether changes caused by the intervention exist in small samples.

This study was approved by the Leiden University Faculty of Science ethics committee. All participants read the study description and gave informed consent before participating. Data will be collected anonymously (see in Appendix A).

The target audience is university students. They are generally familiar with digital games and interactive technologies, making them suitable for testing a learning activity designed as a game. Furthermore, recruiting participants on campus is manageable and helps ensure the study is carried out consistently. Another reason is that students are in a phase where they are developing views on technology and its impact on society.

We collected 50 responses. While this sample size is insufficient for large-scale deployment, it is feasible for exploratory experiments, and can reveal the potential impact of the intervention on knowledge and attitudes.

3.2. Game Design

The research tool used in this study is an interactive narrative game called Echoes of Algorithm, developed on the RPG Maker MV platform. The game characters were hand-drawn by the researcher, while the background scenes were drawn and modified on the Canvas platform to ensure consistency and originality of the visual style.

The goal of this game is to serve as an educational tool to help players understand the issue of "Algorithmic Bias". The game aims to explore what algorithmic bias is, how it arises, and why it matters.

3.2.1. Characters and Narrative Structure

The player assumes the role of a journalist investigating a hiring discrimination case at a fictional company. The story begins with a highly qualified female job applicant who is repeatedly rejected. A journalist's investigation guides the player from questioning individual responsibility to uncovering deeper, systemic issues. This process gradually reveals how bias is incorporated into algorithms, such as through historical data and proxy variables [4, 61]. Ultimately, the player understands the systemic bias behind algorithmic bias and the importance of human-centered accountability. This case is similar to the real-life incident in 2018 when Amazon suspended its AI recruitment program after the system learned patterns of gender inequality and was exposed as discriminating against female applicants [24].

The game's narrative structure is similar to a puzzle. Players engage in conversations with various non-player characters (such as an HR manager, an IT engineer, and the company's CEO). The HR representative argued that AI recruitment could avoid human bias and attempted to mask the issue with reasons of efficiency and fairness, reflecting the common view that algorithms are more objective than humans. This shows that algorithms are often viewed as objective and effective when performing rote tasks, but are perceived as unfair when it comes to personnel evaluations [5]. The IT engineer discussed issues with historical data and the selection of proxy variables in the data, demonstrating that simply removing sensitive indicators (such as gender and age) does not truly eliminate unfairness [26]. When questioned by a journalist, the CEO acknowledged the existence of systemic bias and said that AI systems act as a mirror to human society. Ultimately, the journalist's reflections suggested that humans should ultimately take responsibility for the choices the AI system made[62].

3.2.2. Interaction Mechanics

The game's interactive mechanics combine investigation, dialogue choices, and evidence presentation. Players can freely choose to visit HR, IT, or the CEO first in the company lobby. Different visit sequences yield different clue combinations. The Figure 1 below shows the company lobby.

Figure 1: Screenshot of Choice box

When players talk to the corresponding NPC, they can gather certain clues. The following Figure 2 shows a conversation scene with an IT engineer.

Figure 2: Screenshot of IT

At key moments, players need to present the right clues to unlock further dialogue and narrative development. This mechanic draws on the logic of detective mystery games, such as Ace Attorney [63], allowing players to take a more proactive role in analysis and argumentation. Such as shown in the Figure 3 below, when presenting evidence, you must select sufficient and correct evidence to unlock and proceed to the next storyline. In the conversation with the CEO, the CEO refuses to acknowledge the AI's shortcomings until the player presents the correct evidence they've previously collected.

Figure 3: Screenshot of CEO

3.3. Survey Evaluation

Survey evaluation is based on the Kirkpatrick model of training effectiveness, focusing on learning (knowledge and perception) and behavior (intention to act) [64]. The survey instrument was implemented in Qualtrics. The questionnaire used a 5-point Likert scale and was supplemented with open-ended background questions. The survey process is that players start playing the game immediately after completing the pre-test, and return to the questionnaire to complete the post-test questions after the game is over. Some questions in the questionnaire were reverse coded to prevent players from always choosing the same response options during testing. There are four main sections as follow.

1. Background Questions: Degree, study major, prior knowledge about AI fairness content, and interest in technology news.
2. Bias Perception and Fairness (Pre- and Post-game): Likert-scale items assessing perceptions of AI fairness and trust in algorithmic vs. human decisions.

3. Knowledge Assessment (Pre- and Post-game): Likert-scale items measuring understanding of algorithmic bias concepts (e.g., historical data bias, systemic bias). In addition, the post-test included a multiple choice item as a complementary knowledge check. This item presented a hiring scenario and asked participants to identify a potential case of algorithmic bias.
4. Behavioral Intentions (Post-game only): Likert-scale items assessing willingness to challenge unfair AI decisions, support ethical AI, and continue learning. An additional item asked about interest in future game chapters.

The learning aspect examines whether players have learned relevant knowledge points. The behavioral aspect primarily examines whether players' intentions toward algorithmic decisions change after playing the game. Furthermore, the fairness perception component focuses on changes in participants' perceptions of the fairness of AI decisions. This component measures whether the game helps players become aware of potential biases and societal risks associated with algorithms. Several items were adapted from the organizational justice scale originally developed by Colquitt [59]. For example, "Have those procedures been free of bias?" was reformulated as "AI decision processes are free from bias.", and "The decision-making process of AI systems is based on accurate and appropriate information." was adapted into "AI systems can create unfair outcomes through seemingly neutral criteria".

To compare pre- and post-test responses, we used a paired-samples t-test. This test is suitable for within-subjects designs where the same group of participants are measured at two different time points, as it effectively controls for individual differences [65]. The t-test can also be used in experiments with small sample sizes, but the results are more likely to reflect trends rather than strong statistical conclusions [66]. Its limited statistical power under small-sample conditions means that the results should be interpreted cautiously.

4. Result

4.1. Overview of Participants

A total of 61 responses were collected from the survey. Among them, 49 participants completed the full questions. Among these, 42 came from university students and were included in the analysis. Regarding their current degree, Figure 4 shows 40.5% ($n = 17$) were pursuing a bachelor's degree, 52.4% ($n = 22$) were master's students, and 7.1% ($n = 3$) were doctoral students. In terms of disciplinary background(see Figure 5), 47.6% ($n = 20$) reported studying in a field related to computer science, artificial intelligence, or data science, while 52.4% ($n = 22$) came from other fields. In Figure 6, most of the respondents (81.0%, $n = 34$) indicated that they regularly follow news about technology or social issues, while a smaller proportion (19.0%, $n = 8$) reported not. This disciplinary imbalance may have influenced the results. Students with backgrounds in related majors probably had more prior knowledge about algorithms than those from other fields, which may have shaped their responses.

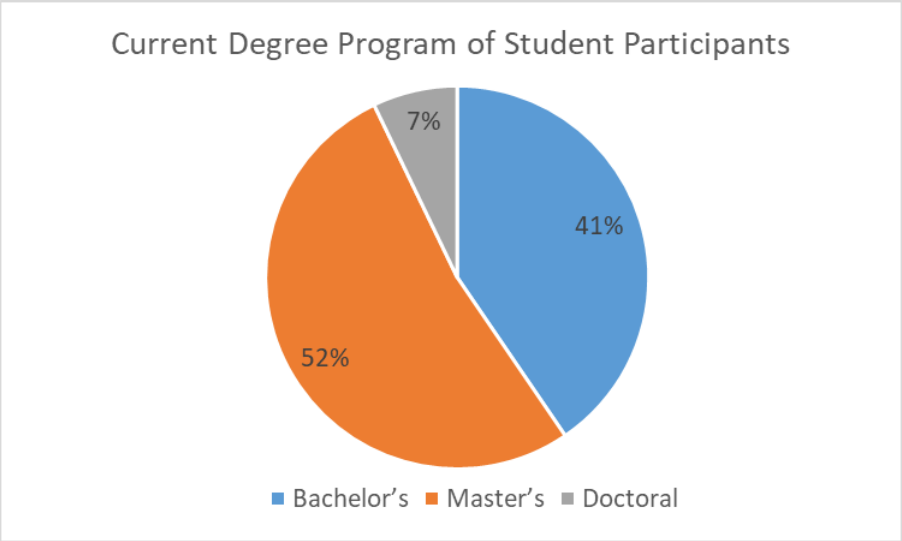

Figure 4: Distribution of student participants by current degree program

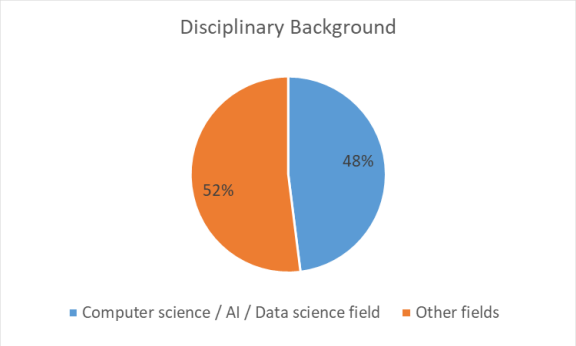

Figure 5: Distribution of student participants by disciplinary background

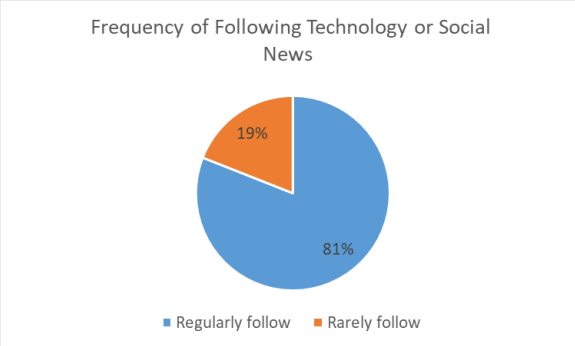

Figure 6: Distribution of student participants by frequency of following technology or social news

4.2. Knowledge Assessment

In order to evaluate the effect of the game on the perceived knowledge level, this study set up seven related questions in the questionnaire. These included six questions related to knowledge comprehension related to algorithmic bias (Q15-Q20, including the pre-test and post-test), and a situational choice question in the post-test. From the overall trend, the post-test scores of the six knowledge questions were all higher than the pre-test, suggesting that the participants' understanding of the relevant concepts improved after playing the game (see Figure 8).

Table 1
Overall knowledge assessment (Q15–Q20 average): pre–post comparison

Group (n=42)	Mean	SD	$t(41)$	p	Cohen's d
Pre-test	3.72	0.71	3.64	0.0008	0.56
Post-test	4.01	0.66			

In the comparison of the mean scores, the results showed that the general perceived knowledge of the participants after playing the game was higher than that of the pre-test (see Table 1). Questions were scored on a 1–5 Likert scale, with higher scores indicating greater understanding. The pre-test mean score was M=3.72, SD=0.71, and the post-test was M=4.01, SD=0.66, an increase of approximately

0.29 points. The results of the paired sample test showed that the difference reached a certain level of statistical significance ($t(41)=3.64$, $p=0.0008$), and the effect size Cohen's $d=0.56$, which was a medium effect. Although the sample size is limited, this result still provide preliminary evidence that playing the game helps participants feel like they know more about algorithmic bias.

Beyond the mean comparison, Figure 7 uses a Sankey diagram to visualize how participants' perceived knowledge scores shifted from pre- to post-test across the six knowledge questions, as well as the overall flow of scores. The nodes in this diagram are arranged in ascending order of total flow (including all people who started from and moved into that score).

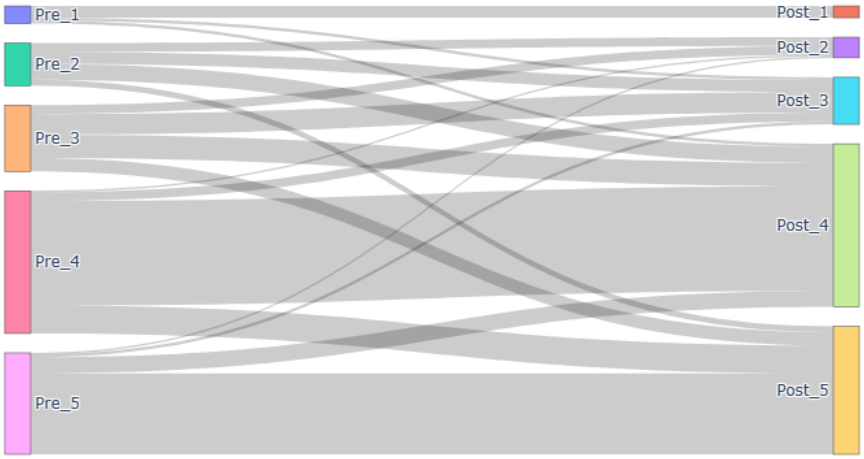

Figure 7: Sankey diagram of transitions on six knowledge questions (pre-test to post-test)

As shown in Figure 8, the post-test mean scores for all questions were higher than the pre-test. Questions 17 and 18 showed the largest pre- and post-test differences, with lower pre-test scores (mean score about 3.4) and highest post-test improvements. Q16 "I know that algorithms can replicate existing inequalities when trained on biased historical data." and Q20 "Responsible AI should focus on supporting human capabilities and protecting ethical and societal values." scored the highest (mean score higher than 4.2), indicating that most participants had a high degree of recognition of the clues *Biased Past Data* and *Human Accountability* in this game.

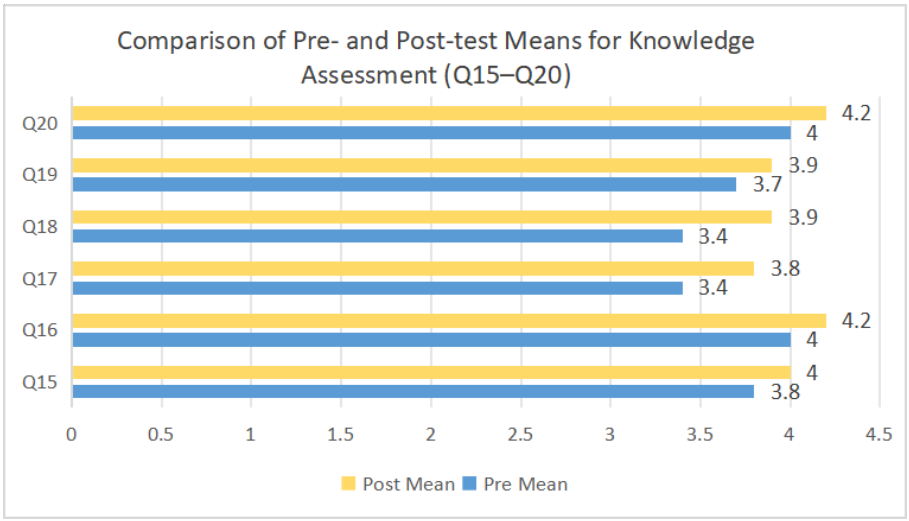

Figure 8: Means for Knowledge Assessment

Figure 9 and Figure 10 shows the distribution of pre- and post-test responses for Q17 and Q18. On Q18, "I know what 'systemic bias' means in the context of algorithmic decision-making," the proportion

of participants who chose "Strongly Agree" increased from 12% to 24% in the post-test. This question assesses knowledge of *systemic bias* in games, and the results suggest that games may be effective in explaining basic concepts. On Q17, "I am aware that AI systems may unintentionally discriminate through factors like commute distance or overtime work," the proportion of participants who chose "Strongly Agree" increased to 26% in the post-test, while the proportion of participants who chose "Neutral" or "Disagree" decreased, indicating that more participants were able to identify potential bias caused by proxy variables.

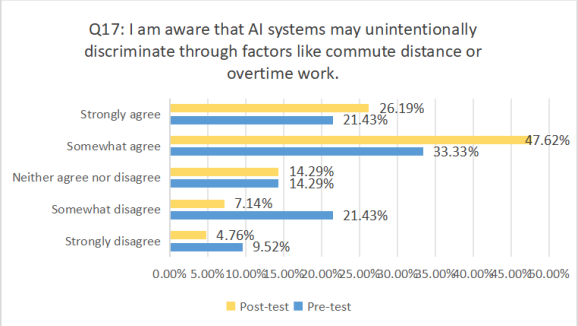

Figure 9: Pre- and post-test distribution of responses for Q17

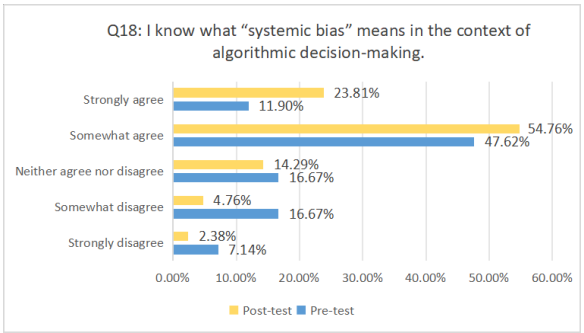

Figure 10: Pre- and post-test distribution of responses for Q18

In addition, the post-test included a multiple-choice question: "Which of the following is a type of algorithmic bias that can potentially happen in the hiring process?" This question served as an additional knowledge check, designed to assess participants' understanding of the concept through a contextualized multiple-choice question. The results showed that 83.7% (n = 36) of the participants selected the correct answer ("Preferring candidates who live closer to the office, which unintentionally favors certain groups"), demonstrating that most of the players were able to not only express their understanding of the concept in a Likert-style question but also identify instances of bias in specific contexts.

4.3. Bias Perception and Fairness

Four Likert scale questions (Q5–Q8) were designed to explore whether the game affected participants' attitudes toward algorithmic bias and fairness. The questions include the perception of algorithmic bias, the judgment of fairness, and the degree of human trust in algorithmic decisions.

Table 2
Overall bias perception and fairness assessment (Q5–Q8 average): pre–post comparison

Group (n=42)	Mean	SD	$t(41)$	p	Cohen's d
Pre-test	3.40	0.84	3.17	0.0029	0.49
Post-test	3.83	0.88			

Table 2 shows that the Mean score rose from M = 3.40 (SD = 0.84) in the pre-test to M = 3.83 (SD = 0.88) in the post-test—an increase of about 0.43 points. The comparison results of the pre-test and post-test showed that the difference reached a certain statistically significant level (t(41)=3.17, p=0.0029), and the effect size Cohen's d=0.49 was a medium effect. Although the sample is limited, the overall trend shows that the game increased the participants' sensitivity to algorithmic unfairness. This suggests that it may play a positive role in the attitude change level. The game may have played a positive role in the attitude change of participants.

This Sankey diagram (Figure 11) shows the pre-test and post-test score changes and flow trends for four questions on Bias Perception & Fairness. Each line in the diagram represents the number of people who moved from one score in the pre-test to another in the post-test. Thicker lines represent more people who moved, and therefore greater flow. The nodes in this diagram are arranged in ascending order of total flow (including all people who started from and moved into that score).

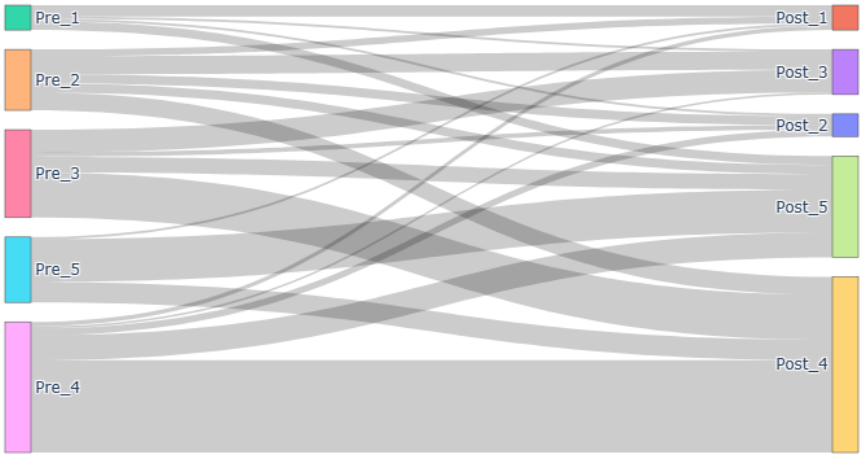

Figure 11: Sankey diagram of transitions on six knowledge questions (pre-test to post-test)

In Figure 12 , the average scores for the four fairness-related questions in the post-test were all higher than those in the pre-test, a trend consistent with the results in Table 2 . Among them, Q8, “I trust algorithmic decisions more than human decisions because humans are more biased.” (reverse-scored), saw the largest change, increasing from M = 3.1 to M = 3.6. Q6, “Algorithmic decisions are not always fair,” was already at a high level in the pre-test and increased further in the post-test (M = 3.7 to 4.1), indicating that participants generally acknowledge the potential for algorithmic bias and unfairness. Q5 and Q7 also showed varying degrees of growth.

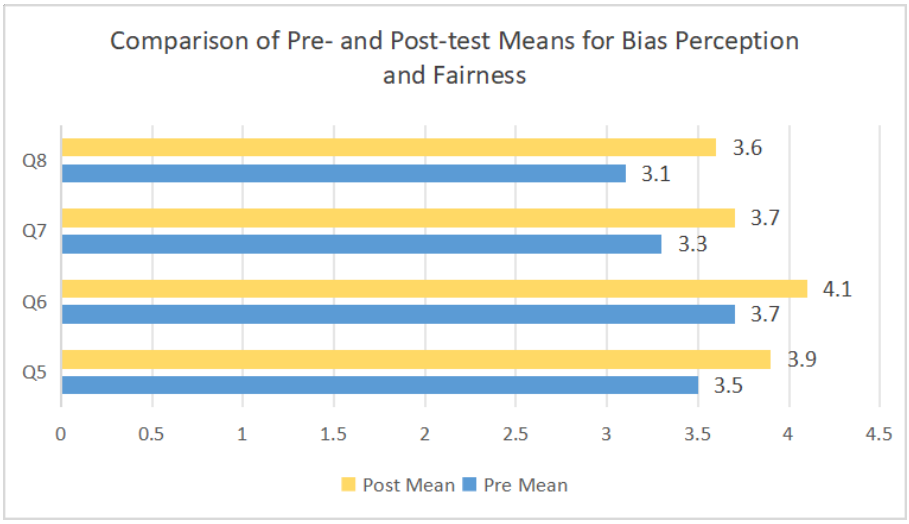

Figure 12: Means for Bias Perception and Fairness

Specifically, in Q5, the general trend showed that participants were more aware of the possibility that algorithms reflect social biases such as gender and race after playing the game. Specifically, the proportion of participants who chose ”strongly agree” increased from 19.05% to 33.33%, while the proportion who chose ”somewhat agree” remained high (38.10% to 42.86%) (see Figure 14). At the same time, the proportions of participants who expressed neutrality and disagreement decreased, showing that participants displayed a higher sensitivity to social dimensions related to fairness.

In Q6, the post-test showed that the proportion of participants who chose ”strongly agree” increased to 35.71%, and the proportion of those who expressed ”somewhat agree” also increased to 50.00% (see Figure 14), while the number of participants who expressed neutrality or disagreement decreased. This suggests that after playing the game, more and more participants have become more aware that algorithms are not completely objective and unbiased.

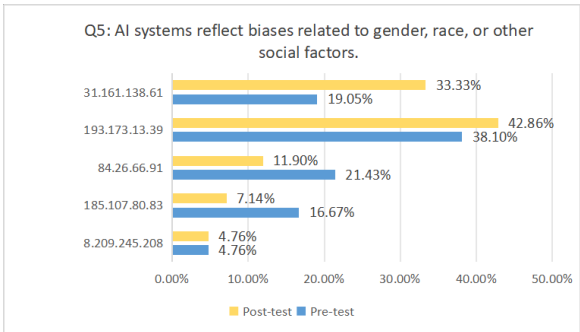

Figure 13: Pre- and post-test distribution of responses for Q5

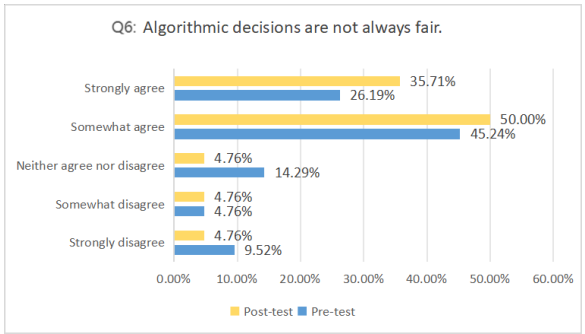

Figure 14: Pre- and post-test distribution of responses for Q6

In Question 8 (see Figure15), "I trust algorithmic decisions more than human decisions because humans are more biased," more participants chose the disagreement option "Somewhat disagree," rising from 23.81% to 42.86% in the post-test, and "Strongly disagree," rising from 11.90% to 23.81%. This suggests that after playing the game, participants' belief that "algorithms are more trustworthy than humans" has weakened, leading to a more skeptical and cautious attitude.

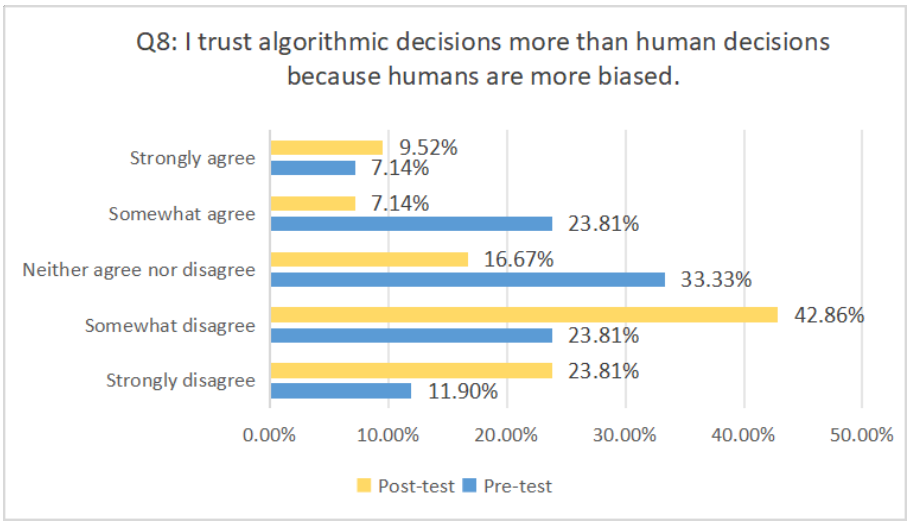

Figure 15: Pre- and post-test distribution of responses for Q8

4.4. Behavioral Intention

To examine the impact of the game on the behavioral level, six behavioral intentions related to the algorithmic system were measured after the game.

Table 3 below summarizes the means and standard deviations for each item. In general, the mean scores of the participants in most items were above the scale midpoint (3). These results suggest that the game may have increased players' awareness and engagement with real-life issues concerning algorithmic bias. Players may be more willing to question the fairness of AI systems and more interested in participating in similar games in the future.

Specifically, participants agreed most strongly with Q9, "I believe that the training data sources for AI hiring systems should be reviewed" (M=4.2, SD=1.3), indicating a strong intention to support data source review. The average scores for Q10, "If I encountered an unfair AI decision, I would raise a concern or give feedback," and Q13, "I would advocate for fairer and more ethical AI systems when given the chance," were above 4, indicating that most people are willing to question and promote fairer AI systems in real-world situations. In Q14, "I would be interested in playing a second chapter or a follow-up story in this game," received the lowest score (M=3.5, SD=1.3), indicating that while participants have strong intentions to act on ethical and social issues, their interest in the game itself is

Table 3
Descriptive statistics for Behavioral Intentions (Post-game only)

Question	Mean (M)	SD
Q9. I believe that the training data sources for AI hiring systems should be reviewed.	4.2	1.3
Q10. If I encountered an unfair AI decision, I would raise a concern or give feedback. AI	4	1.1
Q11. I am willing to actively learn more about algorithmic systems and bias.	3.9	1.2
Q12. I would question the fairness of decisions made solely by AI systems.	4.1	1
Q13. I would advocate for fairer and more ethical AI systems when given the chance. AI	4.1	1.1
Q14. I would be interested in playing a second chapter or a follow-up story in this game.	3.5	1.3

relatively limited. From the perspectives of entertainment and sustained engagement, the appeal of the game still needs improvement.

4.5. Subgroup Analysis

This study compared changes in knowledge comprehension (average scores for Q15–Q20) and fairness perceptions (average scores for Q5–Q8) in different subgroups (see Tables 4 and 5). The results showed that both students in related and non-related majors, as well as participants who regularly followed science, technology, and social news, showed some improvement after playing the game. The table shows that groups with lower pre-test scores tended to see greater improvements after playing the game.

In knowledge understanding, the mean score for the related major group was higher at pre-test (M = 3.93, SD = 1.05), which increased to M = 4.12 (SD = 0.97) at post-test. The non-related major group started at a lower mean (M=3.53, SD=1.09), which increased to M = 3.92 (SD = 1.00) after the test. Similarly, among the news follower group, participants who frequently followed the news saw their self-perceived knowledge level increase from M=3.80 (SD=1.09) to M=4.02 (SD=0.98), while those who rarely followed the news saw their self-perceived knowledge level increase from M=3.40 (SD=1.11) to M=3.98 (SD=0.97). These results suggest that participants with relevant professional backgrounds or those who frequently followed the news had higher levels of self-perceived knowledge at the beginning of the study. The changes brought about by the game were more pronounced among those with non-related professional backgrounds and those who did not follow news.

Table 4: Perceived knowledge (Q15–Q20) by subgroups

Subgroup	N	Pre M (SD)	Post M (SD)
Related major	20	3.93 (1.05)	4.12 (0.97)
Non-related major	22	3.53 (1.09)	3.92 (1.00)
Follow news	34	3.80 (1.09)	4.02 (0.98)
Not follow news	8	3.40 (1.11)	3.98 (0.97)

A similar trend was observed in fairness perceptions. The related major group saw an improvement from M=3.53 (SD = 1.23) to M=3.79 (SD = 1.22), while the non-related major group saw an increase from M=3.30 (SD = 1.03) to M=3.88 (SD = 0.98). Within the news attention group, the group that did not pay attention to the news saw a higher increase, from M = 3.28 (SD = 1.13) to M=4.03 (SD = 1.00). However, due to the limited size of some subsamples (e.g., the group that does not regularly follow the news, n = 8), these results need further verification in a larger sample size study.

Table 5: Fairness perception (Q5–Q8) by subgroups

Subgroup	N	Pre M (SD)	Post M (SD)
Related major	20	3.53 (1.23)	3.79 (1.22)
Non-related major	22	3.30 (1.03)	3.88 (0.98)
Follow news	34	3.43 (1.14)	3.78 (1.11)
Not follow news	8	3.28 (1.13)	4.03 (1.00)

5. Discussion

This section discusses how the game influenced participants' learning outcomes, perceptions of fairness, and intentions to act in relation to algorithmic systems.

5.1. How does this game influence players' understanding of algorithmic bias?

This study found that the interactive game had a positive impact on improving the understanding of concepts related to algorithmic bias by participants and their ability to identify specific biases. Generally, the average scores on the post-test for all six self-reported knowledge-related questions were higher than those on the pre-test. In the scenario test, over 80% of participants correctly identified bias examples in the recruitment scenario. This is consistent with the findings of numerous educational game studies, which show that interactive and embodied learning experiences (e.g., educational games) can effectively promote perceived knowledge absorption and conceptual understanding [67]. Interactive storytelling can serve as a stimulus within the stimulus-organism-response (SOR) [68] framework: story scenarios serve as stimuli, and students process them cognitively and emotionally (organismically), ultimately acquiring richer conceptual knowledge and becoming better able to identify bias (response).

5.2. How does this game change players' attitudes and perceptions of fairness?

In the bias perception and fairness section, the game also appears to have had some impact on the attitudes of the participants. The average scores on these questions generally increased after the game, indicating a heightened sensitivity to the potential unfairness of algorithms. the game appears to have encouraged participants to adopt a more cautious stance towards algorithmic decisions and encouraged them to view technical decisions more critically. In particular, the proportion of participants who chose "disagree" in the reverse-scored question 8 ("I trust algorithmic decisions more than human decisions because humans are more biased") increased most in the post-test. This indirectly suggests a deeper understanding of how algorithms reflect societal biases.

This trend is consistent with organizational justice theory, which states that distributive justice (fairness of outcomes) and procedural justice (fairness of the decision-making process) jointly influence people's judgments [69, 70]. Through narrative of the game, participants learned about the potential for algorithmic bias (fairness of outcomes) and experienced this unfairness through storytelling (procedural fairness). In this sense, narratives may help make abstract ethical issues more immersive and relatable experiences.

5.3. How does this game influence players' behavioral intentions?

Participants showed positive responses to questions related to behavioral intentions, such as expressing a willingness to "question unfair AI decisions." This suggests that the game, after influencing participants' cognition and attitudes, also stimulated their willingness to take action. While the survey captures only self-reported intentions and not actual behavior, such expressions of willingness are often considered an important precursor to behavioral change. It also shows that the game may have enhanced participants' motivation to engage more critically and proactively with algorithmic systems.

At the same time, participants showed relatively low interest in continuing to play subsequent chapters of the game, reflecting a common challenge in serious game design. While narrative-driven games facilitate understanding and learning and may change players' attitudes, further enhancements in terms of fun or narrative engagement are needed to maintain engagement.

5.4. How does this game impact the abovementioned effects on different subgroups?

Subgroup analysis revealed that while all groups improved after playing the game, those with lower initial knowledge levels showed more improvement. For example, students in non-related majors showed greater improvements in knowledge and perceived fairness. Similarly, participants who rarely followed science news also showed greater improvement. These results suggest that interactive games, as a science communication tool, can also be useful for developing new approaches to bridging digital gaps between people of different backgrounds.

6. Conclusion and Future Work

This study developed and evaluated an interactive narrative game to explore its potential to improve students' understanding and attitudes toward algorithmic bias. The results suggest that the game has the potential to positively influence learning, fairness perceptions, and behavioral intentions. It may have helped participants better understand complex concepts, such as historical data bias and systemic bias. Furthermore, the game appeared to make participants more cautious and critical about algorithmic fairness, and increased their awareness of the potential for algorithms to replicate societal biases. The study also highlighted positive behavioral intentions, such as raise concerns about unfair outcomes, and advocating for fairer AI systems. In summary, interactive narrative games have the potential to be an effective science communication tool.

Although the study presented some preliminary results, it still has some limitations. Due to the use of convenience sampling, participants were only included students, which limited the representativeness of the sample. In addition, the sample was overrepresented with students with backgrounds in computer science, artificial intelligence, or data science, and their background knowledge may have influenced the results. The overall sample size was small, and in certain subgroups it was so limited that the reported effects should only be interpreted as tendencies without statistical support. The sample size should be increased in future researches, and the respondents of other educational backgrounds and cultures should be considered.

Second, although the game enhanced knowledge and perceived fairness, the willingness of the participants to keep playing the game was low. The effectiveness of this tool in maintaining public interest in science communication should be enhanced. Further studies should focus their design process on fostering user engagement.

Third, the evaluation was based on a short-term pre-test and post-test comparison. Although the results show positive trends in learning outcomes and attitude changes, it is unclear whether these effects persist. Since the survey only measured behavioral intentions, it is still unclear whether these positive intentions will lead to real-world actions. Furthermore, this study used a single-group design without a control group. In the future, it may be possible to further incorporate a control group.

Finally, this survey only tested behavioral intentions rather than actual behavior, and the research context focused only in hiring case. Interactive narrative methods can be extended to other areas, such as credit scoring, predictive policy, or medical decision-making. Expanding to richer scenarios may attract diverse audiences and help evaluate the effectiveness of this tool in scientific communication.

References

[1] D. Acemoglu, The simple macroeconomics of AI, Economic Policy 40 (2024). doi:10.1093/epolic/eiae042.

[2] S. Afroogh, A. Akbari, E. Malone, M. Kargar, H. Alambeigi, Trust in AI: progress, challenges, and future directions, Humanities and Social Sciences Communications 11 (2024). URL: https://doi.org/10.1057/s41599-024-04044-8. doi:10.1057/s41599-024-04044-8.

[3] K. Ukanwa, Algorithmic bias: Social science research integration through the 3-D dependable AI framework, Current Opinion in Psychology 58 (2024) 101836. URL: https://www.sciencedirect.com/science/article/pii/S2352250X24000496. doi:https://doi.org/10.1016/j.copsyc.2024.101836.

[4] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, ACM Comput. Surv. 54(6) (2021). URL: https://doi.org/10.1145/3457607. doi:10.1145/3457607.

[5] M. K. Lee, Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management, Big Data & Society 5 (2018). doi:10.1177/2053951718756684.

[6] A. Fejerskov, Algorithmic bias and the (false) promise of numbers, Global Policy 12 (2021) 101–103. doi:10.1111/1758-5899.12915.

[7] M. O. Riedl, V. Bulitko, Interactive Narrative: an Intelligent Systems approach, AI Magazine 34 (2013) 67–77. URL: https://doi.org/10.1609/aimag.v34i1.2449. doi:10.1609/aimag.v34i1.2449.

[8] N. Kordzadeh, M. Ghasemaghaei, Algorithmic bias: review, synthesis, and future research directions, European Journal of Information Systems 31 (2022) 388–409. URL: https://doi.org/10.1080/0960085X.2021.1927212. doi:10.1080/0960085X.2021.1927212. arXiv:https://doi.org/10.1080/0960085X.2021.1927212.

[9] D. Danks, A. J. London, Algorithmic bias in autonomous systems, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 4691–4697. URL: https://doi.org/10.24963/ijcai.2017/654. doi:10.24963/ijcai.2017/654.

[10] P. Lopez, Bias does not equal bias: A socio-technical typology of bias in data-based algorithmic systems, Internet Policy Review: Journal on Internet Regulation 10 (2021) 1–29. URL: https://ideas.repec.org/a/zbw/iprjir/250397.html. doi:10.14763/2021.4.1598.

[11] M. Dolata, S. Feuerriegel, G. Schwabe, A sociotechnical view of algorithmic fairness, Information Systems Journal 32 (2021) 754–818. URL: https://doi.org/10.1111/isj.12370. doi:10.1111/isj.12370.

[12] E. Ferrara, Fairness and Bias in Artificial intelligence: A brief survey of sources, impacts, and mitigation strategies, Sci 6 (2023) 3. URL: https://doi.org/10.3390/sci6010003. doi:10.3390/sci6010003.

[13] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: S. A. Friedler, C. Wilson (Eds.), Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81 of *Proceedings of Machine Learning Research*, PMLR, 2018, pp. 77–91. URL: https://proceedings.mlr.press/v81/buolamwini18a.html.

[14] R. S. Baker, A. Hawn, Algorithmic bias in education, International Journal of Artificial Intelligence in Education 32 (2021) 1052–1092. URL: https://doi.org/10.1007/s40593-021-00285-9. doi:10.1007/s40593-021-00285-9.

[15] A. C. B. Garcia, M. G. P. Garcia, R. Rigobon, Algorithmic discrimination in the credit domain: what do we know about it?, AI Society 39 (2023) 2059–2098. URL: https://doi.org/10.1007/s00146-023-01676-3. doi:10.1007/s00146-023-01676-3.

[16] M. Glickman, T. Sharot, How human–AI feedback loops alter human perceptual, emotional and social judgements, Nature Human Behaviour (2024). URL: https://doi.org/10.1038/s41562-024-02077-2. doi:10.1038/s41562-024-02077-2.

[17] R. Agarwal, M. Bjarnadottir, L. Rhue, M. Dugas, K. Crowley, J. Clark, G. Gao, Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware framework, Health Policy and Technology 12 (2022) 100702. URL: https://doi.org/10.1016/j.hlpt.2022.100702. doi:10.1016/j.hlpt.2022.100702.

[18] Z. Chen, Ethics and discrimination in artificial intelligence-enabled recruitment practices, Humanities and Social Sciences Communications 10 (2023). URL: https://doi.org/10.1057/s41599-023-02079-x. doi:10.1057/s41599-023-02079-x.

[19] Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, Dissecting racial bias in an algorithm used to manage the health of populations, Science 366 (2019) 447–453. URL: https://doi.org/10.1126/science.aax2342. doi:10.1126/science.aax2342.

[20] J. Angwin, J. Larson, S. Mattu, L. Kirchner, Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks, ProPublica (2016). URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, accessed: 2025-08-25.

[21] D. F. Mujtaba, N. R. Mahapatra, Fairness in AI-Driven Recruitment: Challenges, metrics, methods, and future directions, arXiv (Cornell University) (2024). URL: https://arxiv.org/abs/2405.19699. doi:10.48550/arxiv.2405.19699.

[22] K. Shrishak, AI-Complex Algorithms and Effective Data Protection Supervision – Bias Evaluation, Support Pool of Experts Programme report, European Data Protection Board (EDPB), 2024. Document submitted in March 2024.

[23] A. Fabris, N. Baranowska, M. J. Dennis, D. Graus, P. Hacker, J. Saldivar, F. Z. Borgesius, A. J. Biega, Fairness and Bias in Algorithmic Hiring: A multidisciplinary survey, ACM Transactions on Intelligent Systems and Technology (2024). URL: https://doi.org/10.1145/3696457. doi:10.1145/3696457.

[24] J. Dastin, Insight – amazon scraps secret AI recruiting tool that showed bias against women, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, 2018. Accessed: 2025-09-14.

[25] C. L. Agbasiere, G. R. Nze-Igwe, Algorithmic Fairness in Recruitment: Designing AI-Powered hiring tools to identify and reduce biases in candidate selection, Path of Science 11 (2025) 5001. URL: https://doi.org/10.22178/pos.116-10. doi:10.22178/pos.116-10.

[26] E. Drage, K. Mackereth, Does AI debias recruitment? Race, gender, and AI's "Eradication of Difference", Philosophy & Technology 35 (2022). URL: https://doi.org/10.1007/s13347-022-00543-1. doi:10.1007/s13347-022-00543-1.

[27] European Union, Regulation (eu) 2016/679 (general data protection regulation), https://gdpr-info.eu/art-22-gdpr/, 2016. Accessed: 2025-09-14; see Articles 13-15 Article 22 regarding automated decision-making, transparency, right to explanation, right to refuse fully automated decisions.

[28] C. Starke, J. Baleis, B. Keller, F. Marcinkowski, Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature, Big Data Society 9 (2022). URL: https://doi.org/10.1177/20539517221115189. doi:10.1177/20539517221115189.

[29] M. K. Lee, S. Baykal, Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division, in: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1035–1048. URL: https://doi.org/10.1145/2998181.2998230. doi:10.1145/2998181.2998230.

[30] M. Mourali, D. Novakowski, R. Pogacar, N. Brigden, Public perception of accuracy-fairness trade-offs in algorithmic decisions in the United States, PLoS ONE 20 (2025) e0319861. URL: https://doi.org/10.1371/journal.pone.0319861. doi:10.1371/journal.pone.0319861.

[31] R. Wang, F. M. Harper, H. Zhu, Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–14. URL: https://doi.org/10.1145/3313831.3376813. doi:10.1145/3313831.3376813.

[32] M. K. Lee, K. Rich, Who is included in human perceptions of ai?: Trust and perceived fairness around healthcare ai and cultural mistrust, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021. URL: https://doi.org/10.1145/3411764.3445570. doi:10.1145/3411764.3445570.

[33] M. Yurrita, T. Draws, A. Balayn, D. Murray-Rust, N. Tintarev, A. Bozzon, Disentangling fairness perceptions in algorithmic decision-making: the effects of explanations, human oversight, and contestability, 2023, pp. 1–21. doi:10.1145/3544548.3581161.

[34] A. Haim, D. Yogev, What Do People Want from Algorithms? Public Perceptions of Algorithms in Government, SSRN Electronic Journal (2025). URL: https://doi.org/10.2139/ssrn.5332425. doi:10.2139/ssrn.5332425.

[35] P. Stachyra, M. Roughley, Playing with Science: Games as a Medium of Science Communication, Springer International Publishing AG, 2023. URL: https://doi.org/10.1007/978-3-031-41652-1_8. doi:10.1007/978-3-031-41652-1\{_8.

[36] S. N. A. Noh, H. Mohamed, N. A. M. Zin, Serious Games model for Higher-Order Thinking Skills in Science Education, International Journal of Advanced Computer Science and Applications 15 (2024). URL: https://doi.org/10.14569/ijacsa.2024.0151022. doi:10.14569/ijacsa.2024.0151022.

[37] G. M. Robinson, M. Hardman, R. J. Matley, Using games in geographical and planning-related teaching: Serious games, edutainment, board games and role-play, Social Sciences Humanities Open 4 (2021) 100208. URL: https://doi.org/10.1016/j.ssaho.2021.100208. doi:10.1016/j.ssaho.2021.100208.

[38] T. Tene, D. F. V. López, P. E. V. Aguirre, N. I. C. Oviedo, C. V. Gomez, S. Bellucci, A systematic review of serious games as tools for STEM education, Frontiers in Education 10 (2025). URL: https://doi.org/10.3389/feduc.2025.1432982. doi:10.3389/feduc.2025.1432982.

[39] S. Champlin, J. James, Breaking health insurance knowledge barriers through games: Pilot Test of Health Care America, JMIR Serious Games 5 (2017) e22. URL: https://doi.org/10.2196/games.7818. doi:10.2196/games.7818.

[40] E. N. Speelman, E. Escano, D. Marcos, N. Becu, Serious games and citizen science; from parallel pathways to greater synergies, Current Opinion in Environmental Sustainability 64 (2023) 101320. URL: https://doi.org/10.1016/j.cosust.2023.101320. doi:10.1016/j.cosust.2023.101320.

[41] M. H. Al-Khresheh, The cognitive and motivational benefits of gamification in English Language learning: a systematic review, The Open Psychology Journal 18 (2025). URL: https://doi.org/10.2174/0118743501359379250305083002. doi:10.2174/0118743501359379250305083002.

[42] M. Taub, R. Sawyer, A. Smith, J. Rowe, R. Azevedo, J. Lester, The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment, Computers Education 147 (2019) 103781. URL: https://doi.org/10.1016/j.compedu.2019.103781. doi:10.1016/j.compedu.2019.103781.

[43] J. Vahlo, K. Tuuri, T. Välisalo, Exploring gameful motivation of autonomous learners, Frontiers in Psychology 13 (2022). URL: https://doi.org/10.3389/fpsyg.2022.825840. doi:10.3389/fpsyg.2022.825840.

[44] M. S. Alotaibi, Game-based learning in early childhood education: a systematic review and meta-analysis, Frontiers in Psychology 15 (2024). URL: https://doi.org/10.3389/fpsyg.2024.1307881. doi:10.3389/fpsyg.2024.1307881.

[45] M. Sailer, J. U. Hense, S. K. Mayr, H. Mandl, How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction, Computers in Human Behavior 69 (2016) 371–380. URL: https://doi.org/10.1016/j.chb.2016.12.033. doi:10.1016/j.chb.2016.12.033.

[46] K. Loderer, R. Pekrun, J. L. Plass, Emotional foundations of game-based learning, The MIT Press, 2020. URL: https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/68937.

[47] S. K. Yeo, M. McKasy, Emotion and humor as misinformation antidotes, Proceedings of the National Academy of Sciences 118 (2021). URL: https://doi.org/10.1073/pnas.2002484118. doi:10.1073/pnas.2002484118.

[48] I. Zimmerman, T. Tal, A. Baram-Tsabari, Cognitive, affective, and behavioral engagement with science news predicted by the use of accessibility strategies in science-minded and general audiences, Public Understanding of Science (2024). URL: https://doi.org/10.1177/09636625241252561. doi:10.1177/09636625241252561.

[49] L. Wei, M. Schmierbach, B. Liu, J. Kang, C. Chen, F. E. Dardis, R. Tan, O. Cohen, Amplifying

player experience to facilitate prosocial outcomes in a Narrative-Based serious game, Media and Communication 13 (2024). URL: https://doi.org/10.17645/mac.8637. doi:10.17645/mac.8637.
[50] J. S. Wu, J. J. Lee, Climate change games as tools for education and engagement, Nature Climate Change 5 (2015) 413–418. URL: https://doi.org/10.1038/nclimate2566. doi:10.1038/nclimate2566.
[51] P. Schimit, L. Monteiro, A vaccination game based on public health actions and personal decisions, Ecological Modelling 222 (2011) 1651–1655. URL: https://doi.org/10.1016/j.ecolmodel.2011.02.019. doi:10.1016/j.ecolmodel.2011.02.019.
[52] R. Rodela, A. Ligtenberg, R. Bosma, Conceptualizing Serious Games as a Learning-Based Intervention in the context of natural Resources and environmental governance, Water 11 (2019) 245. URL: https://doi.org/10.3390/w11020245. doi:10.3390/w11020245.
[53] T. Kermavnar, V. T. Visch, P. M. A. Desmet, Games in Times of a Pandemic: Structured overview of COVID-19 Serious Games, JMIR Serious Games 11 (2023) e41766. URL: https://doi.org/10.2196/41766. doi:10.2196/41766.
[54] R. Jiang, B. Shao, S. Si, R. Sato, J. Tsuneo, Health communication in games at the early Stage of COVID-19 Epidemic: A Grounded Theory Study based on Plague, Inc., Games for Health Journal 10 (2021) 408–419. URL: https://doi.org/10.1089/g4h.2020.0135. doi:10.1089/g4h.2020.0135.
[55] A. Dimitriadou, N. Djafarova, O. Turetken, M. Verkuyl, A. Ferworn, Challenges in serious game design and Development: Educators' experiences, Simulation Gaming 52 (2020) 132–152. URL: https://doi.org/10.1177/1046878120944197. doi:10.1177/1046878120944197.
[56] M. Stankova, D. Tuparova, G. Tuparov, P. Mihova, Barriers to the Use of Serious Computer Games in the Practical Work with Children with Educational Difficulties, TEM Journal (2021) 1175–1183. URL: https://doi.org/10.18421/tem103-22. doi:10.18421/tem103-22.
[57] W. Westera, Why and How Serious Games can Become Far More Effective: Accommodating Productive Learning Experiences, Learner Motivation and the Monitoring of Learning Gains, DOAJ (DOAJ: Directory of Open Access Journals) (2019). URL: https://doaj.org/article/8c8296a61e5b4f268aae10578b639e6d.
[58] A. Alexiou, M. C. Schippers, I. Oshri, S. Angelopoulos, Narrative and aesthetics as antecedents of perceived learning in serious games, Information Technology and People 35 (2020) 142–161. URL: https://doi.org/10.1108/itp-08-2019-0435. doi:10.1108/itp-08-2019-0435.
[59] J. A. Colquitt, On the dimensionality of organizational justice: A construct validation of a measure., Journal of Applied Psychology 86 (2001) 386–400. URL: https://doi.org/10.1037/0021-9010.86.3.386. doi:10.1037/0021-9010.86.3.386.
[60] C. S. Reichardt, Experimental and Quasi-Experimental Designs for Generalized Causal Inference. By William R. Shadish, Thomas D. Cook, and Donald T. Campbell. Boston: Houghton Mifflin, 2002. Pp. 623. (paper)., Social Service Review 76 (2002) 510–514. URL: https://doi.org/10.1086/345281. doi:10.1086/345281.
[61] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, J. Vertesi, Fairness and abstraction in sociotechnical systems, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, New York, NY, USA, 2019, p. 59–68. URL: https://doi.org/10.1145/3287560.3287598. doi:10.1145/3287560.3287598.
[62] E. Commission, E. P. S. Centre, The age of artificial intelligence – Towards a European strategy for human-centric machines, Publications Office, 2018. doi:doi/10.2872/23955.
[63] Capcom, Phoenix wright: Ace attorney, Video game, 2001. Originally released for Game Boy Advance in Japan; later localized for Nintendo DS and other platforms.
[64] D. Kirkpatrick, J. Kirkpatrick, Evaluating training programs: The four levels, Berrett-Koehler Publishers, 2006.
[65] A. Ross, V. L. Willson, Paired Samples T-Test, 2017. URL: https://doi.org/10.1007/978-94-6351-086-8_4. doi:10.1007/978-94-6351-086-8\{_4.
[66] J. de Winter, Using the student's t-test with extremely small sample sizes, Practical Assessment, Research Evaluation 18 (2013).
[67] C.-H. Hsiao, K.-Y. Tang, Who captures whom – Pokémon or tourists? A perspective of the Stimulus-Organism-Response model, International Journal of Information Management 61 (2021) 102312.

URL: https://doi.org/10.1016/j.ijinfomgt.2021.102312. doi:10.1016/j.ijinfomgt.2021.102312.
[68] A. Mehrabian, J. A. Russell, The basic emotional impact of environments, Perceptual and Motor Skills 38 (1974) 283–301. doi:10.2466/pms.1974.38.1.283.
[69] O. Iyigun, Game theory in organizational justice: An experimental study on teams, Journal of Business Financial Affairs 06 (2017). doi:10.4172/2167-0234.1000289.
[70] Y. Cohen-Charash, P. E. Spector, The role of justice in organizations: A meta-analysis, Organizational Behavior and Human Decision Processes 86 (2001) 278–321. URL: https://www.sciencedirect.com/science/article/pii/S0749597801929589. doi:https://doi.org/10.1006/obhd.2001.2958.

A. Survey Questions

A.1. Background Questions

1. What is your current degree? [Multiple Choice]
(Bechalor, Master, PHD, Somethin else)
2. Not a student or something else(text in)
3. Is your major related to computer science, artificial intelligence, or data science? [Yes/No]
4. Do you regularly follow news about technology or social issues? [Yes/No]

A.2. Bias Perception & Fairness (Pre- and Post-game)

Scale: 1 = Strongly Disagree, 5 = Strongly Agree (Likert 1–5)
5. AI systems reflect biases related to gender, race, or other social factors.
6. Algorithmic decisions are not always fair.
7. AI systems can create unfair outcomes through seemingly neutral criteria.
8. I trust algorithmic decisions more than human decisions because humans are more biased. (reverse coded)

A.3. Behavioral Intentions (Post-game only)

Scale: 1 = Strongly Disagree, 5 = Strongly Agree (Likert 1–5)
9. I believe that the training data sources for AI hiring systems should be reviewed.
10. If I encountered an unfair AI decision, I would raise a concern or give feedback.
11. I am willing to actively learn more about algorithmic systems and bias.
12. I would question the fairness of decisions made solely by AI systems.
13. I would advocate for fairer and more ethical AI systems when given the chance.
14. I would be interested in playing a second chapter or a follow-up story in this game.

A.4. Knowledge Assessment (Pre- and Post-game)

Scale: 1 = Strongly Disagree, 5 = Strongly Agree (Likert 1–5)
15. I understand what “algorithmic bias” means.
16. I know that algorithms replicate existing inequalities when trained on biased historical data.
17. I am aware that AI systems may unintentionally discriminate through factors like commute distance or overtime work.
18. I know what “systemic bias” means in the context of algorithmic decision-making.
19. I understand the ethical tension between making fair decisions and maximizing company profit.
20. Responsible AI should focus on supporting human capabilities and protecting ethical and societal values.

A.5. Adapted Items Reference

Table 6 shows the adapted survey items along with the original items from Colquitt (2001).

Table 6: Adapted survey items compared with original Colquitt (2001) items

Adapted Survey Item	**Original Colquitt (2001) Item**
AI systems can create unfair outcomes through seemingly neutral criteria.	The decision-making process of AI systems is based on accurate and appropriate information.
AI decision processes are free from bias. (reverse coded)	Have those procedures been free of bias?
Decisions made by AI systems are consistent and fair. (reverse coded)	Have those procedures been applied consistently?

Declaration on Generative AI

Grammarly was used to check grammar and spelling in this thesis.