Universiteit Leiden

Master Media Technology

AI Education Through Serious Games: Improving AI Literacy in Older Adults

Name:	Lisa Vrijdag
Student ID:	4467736
Date:	05/08/2026
1st supervisor:	G.Barbero
2nd supervisor:	M.F.T. Müller-Brockhausen

Master's Thesis in Media Technology

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands

AI Education Through Serious Games: Improving AI Literacy in Older Adults

Lisa Vrijdag
l.vrijdag@umail.leidenuniv.nl

MSc Creative Intelligence & Technology
Leiden Institute of Advanced Computer Science (LIACS)
Leiden University

Primary supervisor:
Dr. G. Barbero, Leiden University

Secondary supervisor:
Dr. M.F.T. Müller-Brockhausen, Leiden University

August 1, 2026

Abstract

Generative AI is growing rapidly, making it harder to distinguish AI-generated media from authentic media. This creates a need for AI literacy education, especially for older adults. This study investigated how a serious game could be designed to teach older adults about recognising AI-generated images and videos. The game introduced five recognition tips and included 20 classification questions containing both AI-generated and authentic media. The study was conducted with 30 Dutch-speaking participants aged 60-85 years. Data were collected through pre- and post-game surveys, participants' classification scores, and observations during 28 testing sessions. Participants achieved a mean score of 12.2 out of 20, corresponding to an accuracy of 61%. For 21 participants, animal media were the category in which they made the most mistakes. Only a small difference was found between employed and retired participants. The game was received positively: 80% considered it easy to understand, 93.3% found the recognition tips helpful, and 96.7% found the game fun. Participants' confidence in recognising AI-generated media also increased significantly after playing the game. These results show that a serious game can be an understandable, helpful, and enjoyable way to introduce older adults to AI-generated media. However, because no recognition test was completed before playing, this study cannot confirm that the game improved participants' recognition accuracy.

Keywords: AI literacy; serious games; older adults; artificial intelligence; game-based learning

Contents

1 Introduction

As Artificial Intelligence (AI) advances and becomes increasingly widespread across the internet, it is becoming more difficult to distinguish authentic content from AI-generated content. Recognising AI-generated media, including photographs, videos, audio, and text, is therefore becoming increasingly important. According to the Dutch National Cyber Security Centre (NCSC, n.d.), the ability to recognise AI-generated content can help individuals protect themselves against AI-related threats such as deepfakes, phishing, and misinformation. Developing a basic understanding of AI is consequently an important part of building AI literacy. AI literacy refers to having sufficient knowledge to understand AI and think critically about its benefits, limitations, and potential dangers (Digital Promise, 2024).

The need for AI literacy may be particularly relevant to older adults aged 60 years and above. Members of this group use smartphones for activities such as social communication and reading the news (Busch et al., 2021). Research suggests that some older adults experience greater difficulty distinguishing AI-generated content from authentic content and may overestimate their ability to do so (Mittal, 2025). Possible contributing factors include age-related changes in information processing and differences in digital and AI literacy.

Current AI-literacy education is primarily targeted at children and students, leaving older adults comparatively underserved (Tang et al., 2025). This does not necessarily indicate an unwillingness among older adults to learn about AI. Instead, they may be unsure where to begin because AI can appear inaccessible, difficult, and complex. Appropriate AI-literacy education could help older adults protect themselves against online threats, understand how AI works, and reduce their reliance on others for digital support. However, because most existing resources are designed for children or students, there remains a significant gap in educational tools tailored to older adults.

This study proposes an educational serious game as a possible response to this gap. Serious games are designed to achieve purposes beyond entertainment, such as education, training, or behaviour change. A review by Martínez Sanahuja (2019) found that serious games developed for adults aged 60 years and older can positively affect participants' cognitive and physical abilities. However, most of the games included in that review focused on maintaining existing cognitive or physical abilities and were situated within healthcare contexts. Comparatively few focused on teaching older adults new knowledge or skills.

A serious game may provide an accessible and engaging way for older adults to learn about AI. Research by Choudhury (2025) suggests that game elements can support learning by creating an interactive and engaging environment. Game-based approaches may also help users feel more comfortable by providing clear visual or auditory feedback. Furthermore, a game can be designed around the accessibility needs of older adults through understandable instructions, simple controls, readable visual elements, and immediate feedback.

The main objective of this study is to investigate how a serious game can be designed to introduce adults aged 60 years and older to the recognition of AI-generated media. For this purpose, an educational serious game was developed and evaluated with 30 members of the target population. Data was collected through a pre-game survey, participants' performance in the game, researcher observations, and a post-game survey.

The study was guided by the following subquestions:

1. Do older adults have more difficulty distinguishing AI-generated media involving humans or animals?
2. Are there differences in classification performance between employed and retired older adults?

2 Related Work

2.1 AI-Generated Content and AI-Supported Scams

Generative AI enables users to create content such as text, images, audio, and video. Although these technologies offer many opportunities, they also introduce several risks. According to Tredinnick and Laybats (2023), these risks include:

- **Fake content and misinformation.** AI-generated content can appear authentic despite being artificially produced. When this content is shared through social media, it may contribute to the spread of misinformation.
- **AI-supported criminal activity.** Criminals can use generative AI to produce phishing messages, create false documents, imitate individuals, and distribute deepfakes for purposes such as fraud or blackmail (NCSC, n.d.).

A related problem is that the origin of AI-generated content is not always transparent. Without appropriate disclosure, users may be unable to determine whether content was produced by a person or an AI system. Transparency allows users to evaluate generated content with greater awareness of how it was created (Medhat et al., 2025). For example, conversational AI systems can produce text that resembles human writing. If such output is not labelled, users may incorrectly assume that it represents human-authored information or opinions.

Given these risks, it is increasingly important that individuals can critically evaluate AI-generated media. The following section examines existing research on people's ability to distinguish AI-generated media from authentic content.

2.2 Recognition of AI-Generated Media

Recent improvements in generative AI have increased the realism, motion consistency, and visual quality of generated media. An informal illustration of this development is the "spaghetti benchmark," in which generative models have been prompted to create a video of actor Will Smith eating spaghetti (Shypunova & Poliakova, 2026). Changes in the quality of these videos illustrate how rapidly AI-generated media have improved.

Research indicates that people cannot always reliably distinguish AI-generated media from authentic media. Köbis et al. (2021) tested 210 participants using 16 videos and found an overall deepfake-detection accuracy of 57.6%. The researchers concluded that participants were not consistently able to identify deepfakes and frequently overestimated their detection abilities.

Similarly, Cooke et al. (2024) recruited 1,276 participants to assess their ability to distinguish between authentic and AI-generated media. Participants were presented with audio, images, video, and audiovisual material. Their recognition performance was close to 50%, indicating that their classifications were only slightly better than chance.

A study by Lu et al. (2023) focused specifically on image classification. Fifty participants were asked to distinguish authentic images from AI-generated images. The average correct-classification rate was 66.9% for authentic images and 44.2% for AI-generated images. This difference suggests that participants experienced particular difficulty identifying AI-generated images.

Although these studies support the general conclusion that AI-generated media are difficult to recognise, their results cannot be compared directly. Köbis et al. (2021) focused on deepfake videos, Cooke et al. (2024) examined several media formats, and Lu et al. (2023) focused exclusively on images. Differences in stimuli, procedures, sample sizes, and classification tasks may have influenced the reported accuracy rates.

Awareness of recognition methods also appears to be limited. In a German study reported by Krempl (2026), 47% of respondents believed that they could identify AI-generated images and videos. However, approximately one-third did not know common methods for evaluating such media. Only 28% reported looking for graphical inconsistencies, while 19% checked whether the source was reliable.

Taken together, these findings indicate that people cannot reliably distinguish AI-generated media from authentic content and may overestimate their ability to do so. As generative models continue to improve, intuition alone is unlikely to provide a reliable basis for classification. The following section therefore examines practical methods for evaluating potentially AI-generated media.

2.3 Methods for Recognising AI-Generated Media

Although AI-generated content is becoming more difficult to recognise, visual and contextual inconsistencies can still provide useful warning signs. A guide developed by Kamali et al. (2024) identifies five categories of common errors in AI-generated media. These categories formed the basis of the recognition tips presented in the serious game:

1. **Anatomical implausibilities.** Users can examine whether human and animal bodies appear anatomically plausible. Possible warning signs include extra or missing fingers, limbs, eyes, or other body parts, as well as unusual teeth, proportions, or body positions.

2. **Stylistic artefacts.** AI-generated media may appear unnaturally polished or visually perfect. Examples include glossy skin, an absence of natural textures, unusually smooth surfaces, and excessively blended backgrounds (Vasir & Huh-Yoo, 2025).

3. **Functional implausibilities.** Objects may be constructed or used in ways that would not function in reality. Examples include a person holding an object incorrectly, parts of an object merging, or objects clipping through one another.

4. **Violations of physics.** Generated images and videos may contain unrealistic shadows, reflections, lighting, gravity, perspective, depth, materials, or movement. For example, liquids or tears may move in physically implausible ways (Vasir & Huh-Yoo, 2025).

5. **Sociocultural implausibilities.** AI-generated media may depict events that are historically inaccurate, culturally inappropriate, or highly unlikely within a particular social context.

Related guides also recommend examining fingers, shadows, reflections, and other graphical details and taking sufficient time to evaluate images and videos (Yang, 2023) (van der Sluis, n.d.) . However, these indicators are becoming less reliable as generative models improve. Anatomical errors such as additional fingers, for example, are less common and less obvious in newer AI-generated images. Captitol Technology University (University, 2024) also mentions that looking for distortions and unnatural movements can help figure out whether a photo or video is AI-generated, but they also mention that we should trust our instincts.

Visual inconsistencies should therefore not be treated as definitive proof that media are AI-generated. Instead, they should be considered warning signs that encourage users to conduct further checks, such as investigating the original source, examining contextual information, or consulting a reliable verification service.

Applying these strategies requires both AI literacy and broader digital literacy. The following section discusses these forms of literacy among older adults and their relevance to the evaluation of AI-generated media.

2.4 AI and Digital Literacy Among Older Adults

AI literacy involves having sufficient knowledge to understand AI systems and critically evaluate their benefits, limitations, and potential dangers. Combined with digital literacy, this knowledge may help older adults protect themselves against online threats and reduce their reliance on others for digital support (Tang et al., 2025).

Adults aged 60 years and older use smartphones and other digital technologies for activities such as communication, social interaction, information gathering, and reading the news (Busch et al., 2021). However, evaluating the reliability of online content may become more demanding for some older adults because of differences in digital experience and age-related changes in information processing (Mittal, 2025).

Improving digital literacy may support greater independence and help older adults navigate AI-related threats (Tang et al., 2025). AI literacy is increasingly becoming part of this broader digital literacy because users regularly encounter automated systems and AI-generated content online. Some older adults may initially approach unfamiliar technologies with caution or scepticism, highlighting the need for learning materials tailored to their experience and requirements (Köbis et al., 2021).

Research nevertheless suggests that older adults can develop AI literacy when appropriate learning opportunities are available. Peng et al. (2025), for example, developed a ten-lesson AI-literacy programme that taught 12 participants between the ages of 56 and 75 to use AI creatively. The results indicated that participants' AI literacy increased during the programme.

These findings suggest that older adults can benefit from AI-literacy education when learning opportunities are relevant, accessible, and tailored to their needs. However, many existing educational resources are primarily designed for children, students, or working professionals. This leaves an important gap in AI-literacy education for older adults. Serious games may help address this gap by providing an interactive and accessible environment in which users can develop knowledge and practise recognition skills.

2.5 Serious Games for Older Adults

Serious games are games designed to achieve purposes beyond entertainment, such as education, training, assessment, or behaviour change (Bonnechère & Van Sint Jan, 2019). By combining educational content with interactive gameplay, serious games can make learning activities more engaging and accessible.

Motivation can influence whether older adults choose to engage with unfamiliar technologies. Users may be less willing to adopt a new system if its purpose or benefits are unclear (McLaughlin et al., 2012). Gamified learning may help address this barrier by presenting complex information in a more approachable and interactive form (Martínez Sanahuja, 2019).

Most serious games developed for older adults focus on maintaining or improving physical and cognitive abilities rather than teaching new knowledge. A review by Martinho et al. (2020) examined 42 serious-game studies involving older adults. Many of the reviewed games were situated in healthcare contexts and addressed activities such as medication management, cognitive assessment, medical instruction, and physical exercise.

Another review by Gutiérrez-Pérez et al. (2023) examined 108 studies involving a total of 15,902 participants. Of the included studies, 47.17% focused on physical health, 30.19% examined multimodal training, and 22.6% investigated cognitive training. These findings indicate that comparatively little research has examined serious games designed to teach older adults new forms of digital or AI-related knowledge.

2.5.1 Examples of Serious Games for Older Adults

Several types of serious games have previously been developed for older adults:

1. **Augmented-reality game.** Ghorbani et al. (2022) developed an augmented-reality serious game focused on decision-making and cognitive skills. The game was tested with 37 participants aged 65 years and older. Although the participants had limited previous computer experience, the game was generally received positively. Pop-up messages displayed on a tablet, smartphone, or computer provided assistance while participants completed the tasks.

2. **Exergame.** Adcock et al. (2020) investigated whether an exercise-based serious game affected physical functioning, cognitive abilities, and brain volume. The study compared an intervention group with a group that continued to receive its usual care. The intervention group consisted of 31 adults aged 65 years and older. No significant differences were found in physical functioning or brain volume, but the results suggested that the training improved participants' executive functioning.

3. **Farming game.** Eun et al. (2023) developed an exercise-based serious game with an AI-supported personalisation system. The system adjusted the difficulty level for individual players, while a scoring system encouraged continued participation. The game took place in an augmented-reality farming environment. One activity, called *Harvesting Crops*, required players to repeatedly squat and stand within a specified period to earn points.

4. **Tovertafel.** The Tovertafel is a commercial serious-game system designed for older adults with dementia. It allows users to play together with family members, caregivers, or other residents. Its games are designed to encourage movement, cognitive stimulation, and social interaction ("Tovertafel", n.d.).

2.5.2 Designing Serious Games for Older Adults

Salen and Zimmerman define a game as a system in which players engage in an artificial conflict defined by rules and resulting in a quantifiable outcome (Tekinbas & Zimmerman, 2003). Games are commonly digital and interactive, although game elements can also be applied in non-game contexts (Deterding et al., 2011). Serious games differ from entertainment games because they are designed to achieve an additional educational, therapeutic, or training-related purpose.

The ASGD methodology discussed by Djafarova et al. (2023) identifies four elements that should be considered when designing a serious game:

Learning: The knowledge or skills that the game is intended to develop.

Storytelling: The narrative, characters, setting, and context through which the educational material is presented.

Gameplay: The rules, mechanics, challenges, and interactions through which the player engages with the game.

User experience: The way players understand, navigate, and control the game, including the accessibility and clarity of its interface.

In addition to these elements, the visual and interactive design of a serious game should reflect the needs of its target population. Morris (1994) recommends clear text, sufficient colour contrast, appropriate brightness, and a legible display. Research by Guo et al. (2022) suggests that older adults value visible scores and clear visual feedback. Similarly, Salazar Cardona et al. (2024) recommends understandable interactions, accessible controls, and the use of touchscreen or button-based input where appropriate.

Taken together, these findings suggest that serious games for older adults should combine clear learning objectives with simple, motivating, accessible, and enjoyable gameplay. Large readable text, straightforward controls, visual feedback, and a limited amount of information on each screen may make a game easier to understand and use. These recommendations informed the design of the serious game developed in this study.

2.6 Relevance

Existing research indicates that serious games for older adults predominantly focus on maintaining physical or cognitive functioning (Gutiérrez-Pérez et al., 2023; Martinho et al., 2020). Comparatively little attention has been given to games that help older adults acquire new knowledge about emerging technologies.

This represents a relevant research gap because AI-generated media are becoming increasingly common and difficult to recognise. Older adults may benefit from accessible opportunities to develop the knowledge and critical-evaluation skills needed to navigate this changing media environment. A serious game focused on recognising AI-generated media could provide an interactive way to introduce these skills while accounting for the accessibility requirements of the target group.

The present study therefore investigates the design and evaluation of a serious game intended to introduce adults aged 60 years and older to practical methods for evaluating AI-generated images and videos.

3 Methodology

3.1 Research Design

This study used an exploratory mixed-methods user-study design to investigate how a serious game for older adults could be designed to support the recognition of AI-generated visual media. Quantitative and qualitative data were collected to examine participants' classification performance and their experiences with the game.

The study was guided by the following research questions:

Main RQ: How can a serious game for adults aged 60 years and older be designed to support the recognition of AI-generated visual media?

Sub-RQ1: How does classification performance differ between visual media involving humans and visual media involving animals?

Sub-RQ2: How does classification performance differ between employed and fully retired older adults?

The quantitative data consisted of participants' classification scores, classification errors, and survey responses. These data were used to compare performance across human and animal media, examine differences between employed and retired participants, and evaluate participants' perceptions of the game.

The qualitative data consisted of the researcher's observations during the testing sessions. These observations provided additional information about participants' behaviour, difficulties, comments, classification decisions, and experiences while playing the game. Combining these data sources allowed the study to evaluate both classification performance and the accessibility and acceptability of the game design.

3.2 Participants

A total of 30 older adults participated in the main study. Participants were required to be at least 60 years old, speak Dutch, and provide informed consent. Previous experience with video games or AI tools was not required. The minimum age of 60 years was selected to include both recently retired adults and older adults who were still employed.

Participants ranged in age from 60 to 85 years. The mean age was 69.10 years, with a standard deviation of 6.45 years. The descriptive age statistics are presented in Table 1.

Table 1: Descriptive statistics for participants' ages

Statistic	**Age in years**
Mean	69.10
Minimum	60
Maximum	85
Standard deviation	6.45

Participants were recruited through convenience and volunteer sampling. Recruitment took place through the researcher's personal network and through a poster displayed at a community centre in the researcher's city. The community centre also served as an in-person testing location, making participation more accessible to people living in the surrounding area.

Of the 30 participants, 27 completed the study in person and three participated remotely. One remote participant was observed through screen sharing in Microsoft Teams after providing consent. The other two remote participants completed the study without direct observation. Quantitative data were therefore available for all 30 participants, while observational data were available for 28 participants. All participants completed the game and both parts of the survey.

The sample consisted of 18 men and 12 women. Eight participants reported being employed, 19 were fully retired, two were retired but still employed, and one was unemployed. For the exploratory employment-status comparison, the two participants who were retired but still working were included in the employed group. This resulted in 10 employed participants and 19 fully retired participants.

The unemployed participant was excluded from this comparison. The participant characteristics are summarised in Table 2.

Table 2: Participant characteristics

Characteristic	Category	n	%
Gender	Men	18	60.0
	Women	12	40.0
Employment status	Employed	8	26.7
	Retired	19	63.3
	Other	3	10.0
Participation format	In person	27	90.0
	Online	3	10.0

11 (36.7%) of the participants used AI before, 5 (16.6%) did not know for sure whether they use AI before and 14(46.7%) did not use it. Of the 30 participants, 28 reported using social media. 8(26.6%) of the participants tried to learn about AI while 22(73.3%) did not. This represents 93.3% of the sample. Among these 28 participants, 18 (64.3%) primarily used social media for communication and social interaction, while six (21.4%) primarily used it for information gathering or following the news. See Table 3 for participants AI experience and social media use.

Table 3: Participants' AI experience and social-media use

Characteristic	Category	n	%
Ever used AI	Yes	11	36.7
	No	14	46.7
	Do not know	5	16.7
Tried to learn about AI	Yes	8	26.7
	No	22	73.3
Uses social media	Yes	28	93.3
	No	2	6.7
Main purpose of social-media use	Entertainment	4	14.3
	Communication	18	64.3
	Information	6	21.4

Note. Percentages for AI experience and social-media use are based on all participants ($N = 30$). Percentages for the main purpose of social-media use are based only on participants who use social media ($n = 28$).

Because this was an exploratory user study, the target sample size of 30 participants was based on the feasibility of recruitment within the available study period and resources.

3.3 Game Design

3.3.1 Learning Objectives

The primary learning objective of the game was to introduce participants to practical methods for evaluating AI-generated visual media. The game presented five recognition tips based on common inconsistencies found in AI-generated images and videos. These tips were intended to encourage participants to examine online media critically.

Participants were also informed that generative AI is developing rapidly and that visual inconsistencies cannot always be detected. The recognition tips were therefore presented as warning signs rather than definitive proof that an image or video was AI-generated.

3.3.2 Gameplay

Before playing the game, participants completed the pre-game survey. They then received an explanation of the five recognition tips and completed a short practice round to familiarise themselves with the controls and classification task.

The scored part of the game consisted of 20 classification questions presented one at a time. Each question displayed either an image or a video. Participants classified each media item by selecting either *AI* or *Not AI*. They could respond using a computer mouse or the left and right arrow keys.

The classification task contained 10 AI-generated items and 10 authentic items. It also contained 10 items involving humans and 10 involving animals. Participants were not informed of this distribution before completing the task.

After answering all 20 questions, participants were shown their total score and classification errors. They could then review their mistakes and see which media and recognition categories they had found most difficult. The scoring and feedback system is described in Section 3.3.6.

3.3.3 Recognition Tips

The five recognition tips were based on the guidelines proposed by Kamali et al. (2024). The original categories were reformulated as short questions that participants could apply directly while evaluating an image or video:

1. **Is it anatomically correct?**

   This question represented *anatomical implausibilities*. Participants were encouraged to examine whether human and animal bodies appeared anatomically plausible. Possible warning signs included extra or missing eyes, tails, limbs, or fingers, as well as unnatural body proportions or positions. Figures 1 and 2 were presented to participants as tutorial examples

Figure 1: Tutorial example of an anatomical implausibility: a human hand with an extra thumb.

Figure 2: Tutorial example of an anatomical implausibility: a cat with a horn.

2. **Is the scenario plausible?**

   This question represented *sociocultural implausibilities*. Participants considered whether the situation shown was realistic within its social and cultural context. Historically inaccurate events or unusual combinations of people, animals, and objects could indicate AI-generated media. Figures 3 and 4 were presented to participants as tutorial examples.

Figure 3: Tutorial example of an sociocultural implausibilities: Mark Rutte in a pink suit

Figure 4: Tutorial video example of an sociocultural implausibilities: Land turtle swimming in the sea

3. **Is it functionally correct?**

   This question represented *functional implausibilities*. Participants examined whether objects were constructed and used correctly. Objects being used for the wrong purpose, incorrectly constructed objects, or objects clipping through one another could indicate AI-generated content. Figures 5 and 6 were presented to participants as tutorial examples.

Figure 5: Tutorial example of a functional implausibilities: Dog shaped bread

Figure 6: Tutorial video example of a functional implausibilities: Dog warping

Does it follow physical laws?

This question represented *violations of physics*. Participants examined shadows, reflections, lighting, gravity, perspective, materials, and natural movement. Media that appeared to violate physical rules could contain indications of AI generation. Figures 7 and 8 were presented to participants as tutorial examples.

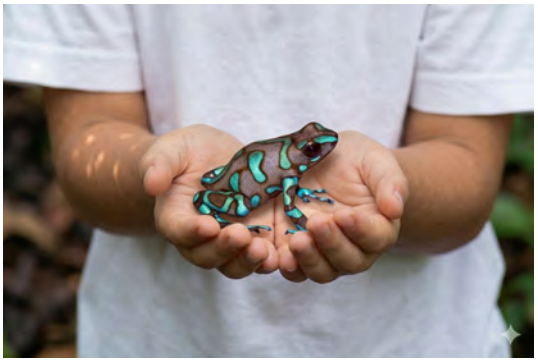

Figure 7: Tutorial example of an functional implausibilities: Frog with accurate shadows

Figure 8: Tutorial video example of an functional implausibilities: Ballet dancer with incorrect mirror

Does it look too good to be true?

This question represented *stylistic artefacts*. Participants looked for media that appeared unnaturally smooth, flawless, symmetrical, or perfectly staged. Examples included skin without visible pores or blemishes, missing surface textures, and an excessively polished appearance. Figures 9 and 10 were presented to participants as tutorial examples.

Figure 9: Tutorial example of a stylistic artefact: Mink posed with flowers

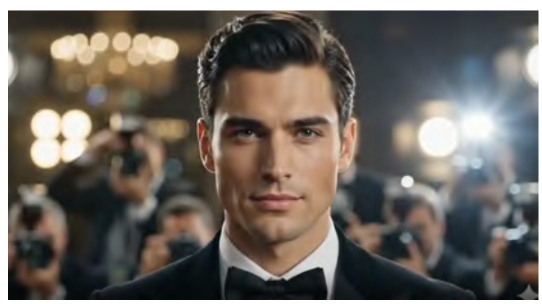

Figure 10: Tutorial video example of a stylistic artefact: Man on red carped

3.3.4 Gameplay and Controls

Because the game was designed for older adults, its interface was kept simple and visually clear. During the in-person sessions, the game was displayed on a laptop with a screen resolution of 1920×1080 pixels. Text and buttons were displayed at a large size to improve readability. Each classification screen contained only the media item, the classification question, and the *AI* and *Not AI* buttons.

When a participant selected an answer, the card containing the media item moved towards the corresponding side of the screen. This animation provided immediate visual confirmation that the response had been registered. It did not indicate whether the answer was correct, preventing participants from receiving performance feedback during the scored classification task. The game and mascot design screenshots are provided in Appendix 10.3 and Appendix 10.4.

Figure 11 presents the dialogue and instruction interface, while Figure 12 presents the interface used for the classification questions.

Figure 11: Screenshot of game: Dialogue screen

Figure 12: Screenshot of game: Classification game screen

3.3.5 Educational and Interface Design

The educational content was based on the media-recognition guidelines proposed by Kamali et al. (2024). These guidelines were reformulated as short questions so that participants could apply them directly while classifying media. Examples demonstrated how each type of implausibility could appear in an image or video. Simplifying the guidelines was intended to make them easier to understand, remember, and apply.

The interface was designed to minimise unnecessary interactions and make the game accessible to older adults. This approach was informed by the recommendations of Guo et al. (2022) and Salazar Cardona et al. (2024). Large interface elements and a limited amount of information on each screen were used to direct participants' attention towards the classification task.

The game also included a mascot named Robbie, who introduced the instructions and guided participants through the different stages. Robbie provided a consistent source of guidance and communicated the information participants required to complete each stage.

3.3.6 Scoring and Feedback

The game recorded participants' responses to all 20 classification questions. Participants received one point for each correct answer, resulting in a maximum score of 20 points.

The game recorded two types of classification error:

- AI-generated media incorrectly classified as authentic; and
- authentic media incorrectly classified as AI-generated.

Feedback was presented only after participants had answered all 20 questions. This prevented feedback from influencing subsequent classification decisions. The results screen displayed the participant's total score, the number of each classification-error type, and whether the participant had achieved lower accuracy for human or animal media. It also identified the recognition-tip categories associated with the participant's mistakes.

An example of the feedback was:

> You answered 10 of the 20 questions correctly.
>
> Of the 20 items, 10 were AI-generated and 10 were authentic.
>
> You classified an AI-generated item as authentic five times and an authentic item as AI-generated five times.
>
> You found the animal category the most difficult, particularly items involving violations of physics and scenario plausibility.

Participants were subsequently given the opportunity to review their mistakes. The purpose of this feedback was to connect participants' classification errors to the five recognition tips introduced during the tutorial.

3.3.7 Survey Design

The survey was divided into pre-game and post-game components. The complete survey is provided in Appendix 10.1.

The pre-game survey collected demographic information, including age, gender, and employment status. Employment status was included to support the exploratory comparison between employed and fully retired participants in Sub-RQ2.

The survey also examined participants' previous experience with AI. Participants were asked whether they had previously used AI tools, attempted to learn about AI, and how they assessed their ability to recognise AI-generated media. Participants were additionally asked about their social-media use and their primary reasons for using social media.

The post-game survey collected participants' classification scores and their evaluations of the game. Participants answered questions about the clarity of the instructions, their enjoyment of the game, and the perceived helpfulness of the recognition tips.

A confidence question concerning participants' perceived ability to identify AI-generated images, videos, and audio was included in both surveys. This allowed participants' self-reported confidence before and after playing the game to be compared.

3.4 Materials and Media

The materials used in the study consisted of the educational game, the AI-generated and authentic media presented in the game, and the pre-game and post-game surveys. The following sections describe the equipment, software, and media stimuli.

3.4.1 Equipment and Software

The 27 in-person sessions were conducted using a laptop with a screen resolution of 1920×1080 pixels. Participants controlled the game using a computer mouse or the left and right arrow keys. The game was developed in Unity and distributed as a Windows application.

The pre-game and post-game surveys were administered through Qualtrics. At the end of the game, participants copied their result data using an in-game button and pasted the data into the post-game survey.

3.4.2 Media Selection and Distribution

The three remote participants used their own computers to access the game and surveys. One remote session was observed through screen sharing in Microsoft Teams after the participant provided consent. The other two participants completed the study without direct observation. They received a PDF containing instructions and could contact the researcher if they experienced technical difficulties.

An internet connection was required to complete the Qualtrics surveys. The game itself could be played offline.

The scored classification task contained 20 media items, consisting of images and videos. Ten items were AI-generated and ten were authentic. The media were additionally divided into 10 human items and 10 animal items. Media containing both humans and animals were excluded so that each item belonged to only one category. Authentic media had only one category: human or animal. AI-generated media had two categories attached to it: human or animal, and one of the five classification categories. The media used in this study can be found in Appendix 10.2.

Within each authenticity condition, five items were images and five were videos. The resulting media distribution is summarised in Table 4.

Table 4: Distribution of media stimuli by authenticity and subject category

Media type	**Human**	**Animal**	**Total**
AI-generated	5	5	10
Authentic	5	5	10
Total	**10**	**10**	**20**

The AI-generated media were created by the researcher using Google Gemini 3.6 Flash (Google, n.d.) on 1 May 2026. The authentic images and videos were obtained from online sources that permitted reuse.

3.5 Data-Collection Measures

The survey was administered in Dutch through Qualtrics and was divided into pre-game and post-game components. It included multiple-choice, yes–no, Likert-scale, and open-ended questions. Closed questions provided quantitative data, while open-ended responses provided additional information about participants' experiences.

The pre-game survey collected participants' age, gender, employment status, social-media use, and previous experience with AI. It also examined whether participants had attempted to learn about AI, whether they considered AI a threat or an opportunity, and whether they believed a video game could be an effective method of AI education.

Participants rated their confidence in their ability to identify AI-generated images, videos, and audio using a Likert-scale question. This question was repeated in the post-game survey.

The post-game survey collected participants' scores and their opinions about the game and its educational content. Participants evaluated the clarity of the instructions, their enjoyment, the helpfulness of the recognition tips, and the perceived effectiveness of the serious-game format.

3.6 Procedure

Each session was scheduled to last no more than 60 minutes. The game, surveys, and study materials were presented in Dutch. The procedure consisted of the following stages:

1. **Briefing and informed consent**

   Participants received a briefing sheet explaining the purpose of the study, the game, their right to withdraw, and how their data would be used. They then provided informed consent.

2. **Pre-game survey**

   Participants completed the first part of the Qualtrics survey. This collected demographic information, employment status, social-media use, previous AI experience, and self-assessed confidence in recognising AI-generated media.

3. **Tutorial and practice questions**

   Participants received an introduction to generative AI and an explanation of the five recognition tips. They then completed a practice round to familiarise themselves with the gameplay and controls.

4. **Classification task**

   Participants completed the 20-item classification task. During observed sessions, the researcher recorded notes about participants' behaviour, difficulties, comments, and requests for assistance. Technical assistance was provided when required, but the researcher did not help participants classify the media.

5. **Scores and feedback**

   After answering all questions, participants received their score and copied their result data into the post-game survey. They could then review their mistakes and read explanations of the correct answers. Participants attending an observed session could also discuss their results with the researcher.

6. **Post-game survey and debriefing**

   Participants completed the post-game survey and received a debriefing sheet. They were given the opportunity to ask questions about the game, generative AI, and the study.

3.7 Pilot Study

A pilot study was conducted with one participant before the main study. The pilot participant's data were not included in the final analysis because the game was modified following the pilot session.

The complete pilot procedure took approximately 45 minutes. An additional 15 minutes were included in the planned duration of the main sessions to account for differences in participants' reading, decision-making, and interaction speeds.

3.8 Data Analysis

Sub-RQ1 was addressed by dividing the 20 scored media items into 10 human and 10 animal items, as shown in Table 4. Items containing both humans and animals were excluded. The task also contained equal numbers of AI-generated and authentic items.

For each participant, the number of classification errors in the human and animal categories was calculated. The category containing the greater number of errors was recorded as the participant's most difficult category.

Sub-RQ2 was addressed by comparing the mean classification scores of employed and fully retired participants. This resulted in 8 employed participants and 19 fully retired participants. The results of the three other work status participants were calculated separately. Because the analysis was exploratory, the group means were compared descriptively.

Pre-game and post-game confidence scores were compared using a paired-samples t-test.

The researcher's observational notes were reviewed descriptively to identify recurring behaviours, difficulties, comments, and item-specific classification patterns. These observations were used to contextualise the quantitative findings.

3.9 Ethical Considerations

Participation was voluntary, and participants received no financial or material compensation. Participants were required to be capable of understanding the study information and providing informed consent.

Before participating, each individual received an information sheet, a briefing sheet, and an informed-consent form. After completing the study, participants received a debriefing sheet. Participants could withdraw their data for up to one month after participation.

4 Results

This chapter presents the results of the classification task and the pre-game and post-game surveys. It also summarises the researcher's observations during the testing sessions. The complete questionnaire is provided in Appendix 10.1.

4.1 Participant Characteristics

All 30 participants completed the experiment, and their quantitative data were included in the analysis. Twenty-seven participants completed the study in person. Of the three remote participants, one was observed through screen sharing and two completed the study without direct observation. Observational data were therefore available for 28 participants.

Participants ranged in age from 60 to 85 years. The mean age was 69.10 years ($SD = 6.45$). The sample consisted of 12 women (40.0%) and 18 men (60.0%).

Eight participants (26.7%) reported being employed, 19 (63.3%) were fully retired, two (6.7%) were retired but still employed, and one (3.3%) was unemployed. The recorded participant characteristics are presented in Table 2.

4.2 Pre-Game Survey

4.2.1 Social-Media Use

Twenty-eight participants (93.3%) reported using social media, while two (6.7%) did not. Among the 28 social-media users, 18 (64.3%) reported that their primary reason for using social media was communication and social interaction. Six participants (21.4%) primarily used it to gather information or follow the news.

4.2.2 Previous AI Experience

Eight participants (26.7%) had previously attempted to learn about AI, whereas 22 (73.3%) had not.

11 participants (36.7%) reported having used an AI tool before the experiment, 14 (46.7%) reported that they had not used one, and five (16.7%) were unsure whether they had previously used AI.

4.2.3 Perception of AI

Participants were asked whether they considered AI a threat or an opportunity for society. Results are shown in Table 5. One participant (3.3%) considered AI a major threat, while 13 (43.3%) considered it a threat. Eight participants (26.7%) considered AI neither a threat nor an opportunity. Seven (23.3%) considered it an opportunity, and one (3.3%) considered it a major opportunity.

Table 5: Participants' perceptions of AI as a threat or opportunity

Response	n	**Percentage**
Major threat	1	3.3%
Threat	13	43.3%
Neither threat nor opportunity	8	26.7%
Opportunity	7	23.3%
Major opportunity	1	3.3%
Total	**30**	**100.0%**

4.2.4 Expected Effectiveness of a Video Game

Before playing the game, participants were asked whether a video game could be an effective way to teach people about AI. Fifteen participants (50.0%) answered yes, one (3.3%) answered no, and 14 (46.7%) indicated that they did not know. See Table 6.

Table 6: Responses before testing

Response	n	**Percentage**
Yes	15	50.0%
No	1	3.3%
I don't know	14	46.7%
Total	**30**	**100.0%**

4.3 Classification Results

The mean classification score was 12.2 out of 20, corresponding to a mean accuracy of 61.0%.

For 21 participants (70.0%), animal media were the category in which they made the greatest number of classification errors. This suggests that most participants experienced more difficulty classifying animal media than human media.

The results screen could identify up to two recognition-tip categories in which each participant had made the most mistakes. Violations of physics were identified as a frequent error category for 27 participants. Incorrect or implausible scenarios were identified for eight participants. Because a maximum of two categories could be recorded for each participant, the category frequencies were not mutually exclusive.

The 19 fully retired participants obtained a mean score of 11.9 out of 20. The eight participants who reported being employed obtained a mean score of 11.3. The three participants in the other employment-status category also obtained a mean score of 11.3.

4.4 Post-Game Survey

4.4.1 Scores

The participants' average score was 12.2 out of 20 correct (Table 7). The most mistakes were made in the Animal category, with 21 participants having this category. The top two classification mistakes were violations of physics with 27 participants and sociocultural implausibilities with eight participants. Taking into account that participants could have up to 2 main mistakes.

Table 7: Summary of test performance and classification mistakes

Measure	**Result**
Average test score	12.2 out of 20
Category with the most mistakes	Animal category ($n = 21$)
Most common classification mistake	Violations of physics ($n = 27$)
Second most common classification mistake	Sociocultural implausibilities ($n = 8$)

Note. Participants could make up to two main classification mistakes. Therefore, the mistake categories were not mutually exclusive, and their counts do not sum to the total number of participants.

4.4.2 Perceived Effectiveness

After playing the game, the results show that 10 participants (33.3%) were neutral about whether a video game could be an effective way to teach people about AI. Thirteen participants (43.3%) rated the approach as effective, while seven (23.3%) rated it as very effective. Overall, 20 participants (66.7%) considered a video game an effective or very effective educational method.

4.4.3 Ease of Understanding

One participant (3.3%) rated the game as very difficult to understand. Five participants (16.7%) were neutral, 13 (43.3%) rated the game as easy to understand, and 11 (36.7%) rated it as very easy to understand. Overall, 24 participants (80.0%) considered the game easy or very easy to understand.

4.4.4 Helpfulness of the Recognition Tips

Two participants (6.7%) were neutral about the helpfulness of the recognition tips. Fourteen participants (46.7%) rated the tips as somewhat helpful, and 14 (46.7%) rated them as very helpful. Overall, 28 participants (93.3%) considered the tips somewhat or very helpful.

4.4.5 Enjoyment

One participant (3.3%) was neutral about whether the game was enjoyable. Eighteen participants (60.0%) rated the game as fun, while 11 (36.7%) rated it as very fun. Overall, 29 participants (96.7%) considered the game fun or very fun.

4.4.6 Confidence in Recognising AI-Generated Media

Before playing the game, seven participants (23.3%) reported low confidence in their ability to identify AI-generated images, videos, and audio. Twelve participants (40.0%) were neutral, while 11 (36.7%) reported some confidence in their ability.

After playing the game, two participants (6.7%) reported no or low confidence in their recognition ability, seven (23.3%) were neutral, 16 (53.3%) reported some confidence, and five (16.7%) reported high confidence. Overall, 21 participants (70.0%) reported some or high confidence after playing the game.

The mean confidence score increased from $M = 3.13$ before the game to $M = 3.83$ after the game. A paired-samples t-test indicated that this increase was statistically significant, $t(29) = 4.03$, $p < .001$. Table 8 shows the results of the paired-samples t-test for participants' confidence scores. With Figure 13 showing the distribution of participants in their self-identified confidence.

Table 8: Paired-samples t-test results for confidence scores

Measure	**Before M**	**After M**	**Difference**	**t(df)**	**p**
Confidence score	3.13	3.83	0.70	4.03 (29)	$< .001$

How much do you trust your ability to identify AI-generated images, videos, and audio?

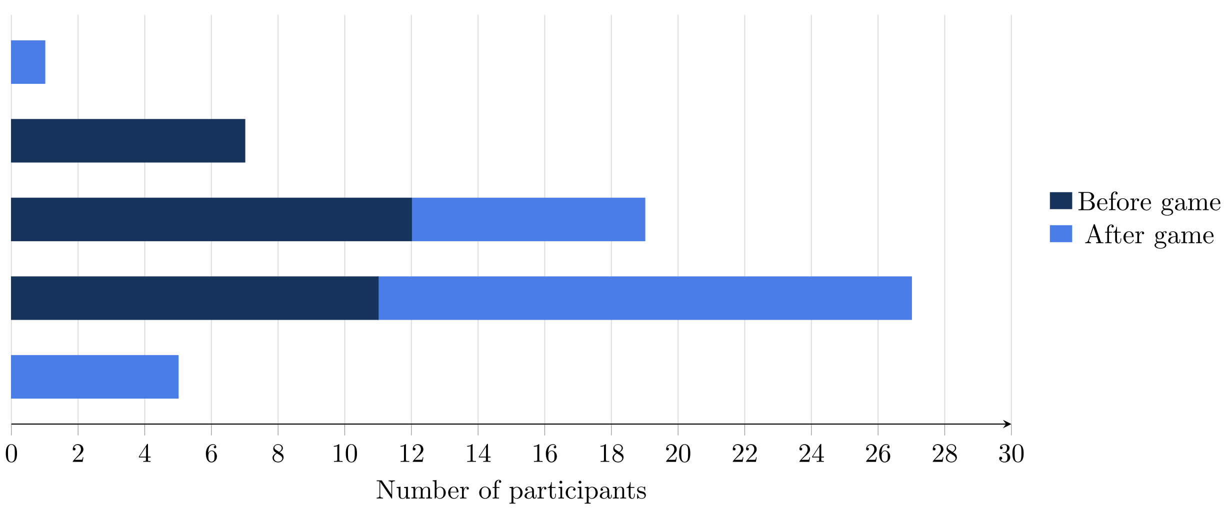

Figure 13: Trust in identifying AI-generated media before and after the game.

4.5 Employed and retired older adults' scores

As for the score difference between the employed and retired participants. 27 out of the 30 participants reported that they were either employed or retired. With 19 participants being retired, and eight still being employed. The other three gave an other working status. These three were excluded from the other 2 main groups, but their average score was still calculated. The retired participants had a mean score of 11.9 out of 20, the employed participants had a mean score of 11.3, and the other category scored a 11.3 out of 20 as well. Table 9 shows the scores as well.

Table 9: Employment status and mean classification scores

Employment status	Participants (n)	Percentage	Mean score (out of 20)	Main comparison
Fully retired	19	63.3%	11.9	Included
Employed	8	26.7%	11.3	Included
Other status	3	10.0%	11.3	Excluded
Total	30	100.0%	–	–

4.6 Researcher Observations

Observational data were available for 28 of the 30 participants. These data included 27 participants who completed the study in person and one remote participant who was observed through screen sharing. The researcher recorded recurring behaviours, comments, and item-specific observations during the sessions.

1. **Increased scepticism**

   Participants were not informed that the task contained equal numbers of AI-generated and authentic items. During the 20 classification questions, several participants appeared to become increasingly sceptical of the media presented to them.

   Participants misclassified authentic media as AI-generated more frequently than they misclassified AI-generated media as authentic. On average, participants made 2.1 errors in which AI-generated media were classified as authentic. In comparison, they made an average of 5.6 errors in which authentic media were classified as AI-generated.

2. **AI-generated hamster video**

   Figure 14 shows an AI-generated video of a hamster moving on a bedsheet. Of the 28 observed participants, 27 classified the video as authentic. The hamster video was therefore one of the AI-generated items that participants found most difficult to recognise.

Figure 14: Screenshot of AI-generated hamster video

3. **AI generation and photo editing**

   Several participants commented that it was difficult to distinguish AI-generated media from media altered using conventional photo-editing software. Participants sometimes recognised that an image appeared unusual but were uncertain whether this resulted from AI generation or manual editing.

4. **Authentic swallow image**

   More than half of the observed participants classified the authentic swallow image shown in Figure 15 as AI-generated. Participants frequently referred to the unusual position and appearance of the bird's feathers when explaining their decisions.

5. **Plausible but unlikely scenarios**

   Figure 16 shows an AI-generated image of two Dutch politicians drinking together in a cafe. The image was selected as an example of a socioculturally implausible scenario.

   Some participants laughed when they saw the image, while others argued that the depicted event remained possible. One participant explained their classification by stating:

   > "It could happen so that it could be true."

   This response illustrates the difficulty of using an unlikely scenario as an indication of AI generation when the depicted event remains physically possible.

Figure 15: Photograph of a swallow.

Figure 16: AI-generated image of two Dutch politicians sitting at a café.

5 Discussion

This chapter interprets the quantitative and qualitative findings in relation to the research questions and previous literature. Where the available data do not provide a direct explanation, possible interpretations are presented cautiously.

5.1 Overall Findings

Participants achieved a mean classification score of 12.2 out of 20, corresponding to an accuracy of 61.0%. Although the tasks and media used in previous studies differed from those used in the present study, this result broadly aligns with research showing that people experience difficulty distinguishing AI-generated media from authentic media. Previous studies have reported recognition accuracy ranging from approximately 50% to 57.6% (Cooke et al., 2024; Köbis et al., 2021).

Participants classified authentic media as AI-generated more frequently than they classified AI-generated media as authentic. On average, participants made 5.6 errors in which authentic media were classified as AI-generated, compared with 2.1 errors in which AI-generated media were classified as authentic. This suggests that participants were more successful at detecting AI-generated media than at correctly accepting authentic media.

The researcher's observations suggested that some participants became increasingly sceptical as the classification task progressed. One possible explanation is that the recognition tips encouraged participants to search for visual defects, causing them to interpret unusual features in authentic media as evidence of AI generation. However, because the study did not include a comparison group that completed the task without receiving the recognition tips, it is not possible to determine whether the tips caused this pattern.

The selection of authentic stimuli may also have influenced the results. Some authentic media depicted unusual subjects or situations and may therefore have appeared visually similar to AI-generated content. This pattern was particularly visible in the authentic swallow image, which more than half of the observed participants classified as AI-generated. Participants frequently referred to the unusual position and appearance of the bird's wings when explaining their decisions.

In contrast, 27 of the 28 observed participants classified the AI-generated hamster video as authentic. Together, these examples demonstrate that authentic media can appear unusual while AI-generated media can appear convincing. Visual recognition tips should therefore be presented as warning signs rather than definitive evidence that media are AI-generated. Players should also be encouraged to investigate contextual information and verify the source of questionable media.

5.2 Confidence and Game Evaluation

Participants were asked before and after playing the game to rate their confidence in identifying AI-generated images, videos, and audio. As shown in Table 8 and Figure 13, the mean confidence score increased from $M = 3.13$ before the game to $M = 3.83$ after the game. A paired-samples t-test indicated that this increase was statistically significant, $t(29) = 4.03$, $p < .001$, with Cohen's $d = 0.74$.

This result indicates that participants felt more capable of recognising AI-generated media after playing the game. However, the confidence measure represents perceived ability rather than objectively measured improvement. Because participants did not complete a classification test before receiving the recognition tips, the study cannot determine whether their recognition accuracy improved.

The difference between participants' confidence and classification errors is also important. Although participants missed relatively few AI-generated items, they frequently classified authentic items as AI-generated. Increased confidence may therefore not necessarily indicate better-calibrated judgement. Future versions of the game should help players understand both types of error: failing to recognise AI-generated media and incorrectly treating authentic media as AI-generated.

Participants nevertheless evaluated the game positively. Eighty per cent considered it easy or very easy to understand, 93.3% found the recognition tips somewhat or very helpful, and 96.7% considered the game fun or very fun. These findings suggest that the serious-game format was accessible and engaging for this sample. Although positive evaluations do not demonstrate learning effectiveness, they indicate that participants were receptive to the format and educational content.

5.3 Sub-RQ1: Human Versus Animal Media

Sub-RQ1 examined whether participants experienced greater difficulty classifying media involving humans or animals. For 21 participants (70.0%), animal media were the category in which they made the greatest number of classification errors. As shown in Table 7.

One possible explanation is that participants were more familiar with human anatomy and with common AI-generation errors involving human faces, hands, and bodies. Animal anatomy and movement may have been less familiar and therefore more difficult to evaluate. This explanation remains speculative because participants' familiarity with human and animal anatomy was not measured.

The animal-media result may also have been strongly influenced by individual stimuli. The AI-generated hamster video was classified as authentic by 27 of the 28 observed participants, while more than half classified the authentic swallow image as AI-generated. Because the game contained only 10 human and 10 animal items, these particularly difficult items could have had a substantial effect on the category-level result.

The findings therefore indicate that animal media were more difficult within the stimulus set used in this study. They do not establish that older adults generally experience greater difficulty recognising AI generation in all forms of animal media. Future research should use a larger and more varied stimulus set to separate subject-category effects from the difficulty of individual media items.

5.4 Sub-RQ2: Employed Versus Retired Participants

Sub-RQ2 examined whether classification performance differed between employed and fully retired participants. Fully retired participants achieved a mean score of 11.9 out of 20, compared with a reported mean of 11.3 among employed participants. The descriptive difference was therefore 0.6 points. The other 3 participants with other employment statuses reported the mean of 11.3. For the full data collected look at Table 9 Because this comparison involved a small convenience sample and groups of unequal size, the results should be interpreted cautiously. Only descriptive statistics were reported, and the uncertainty surrounding the difference was not estimated. Consequently, the study cannot determine whether the observed difference represents a systematic relationship between employment status and classification performance.

The findings therefore provide no clear evidence of a difference between employed and fully retired participants. A larger and more balanced sample would be required to examine this question more reliably. Future analyses should report the variability within each group, an effect size, and a confidence interval in addition to the group means.

5.5 Main Research Question

The main research question asked how a serious game for adults aged 60 years and older could be designed to support the recognition of AI-generated visual media.

The findings suggest that such a game should use clear instructions, simple controls, large and accessible interface elements, and short recognition tips that can be applied directly during classification. These design features are consistent with recommendations for accessible serious games for older adults (Guo et al., 2022; Salazar Cardona et al., 2024).

The media set should include both convincing AI-generated examples and unusual authentic examples. This variety can demonstrate that visual inconsistencies are not definitive evidence of AI generation and that some AI-generated media contain few visible errors. Feedback should address both types of classification error and explain why an item is authentic or AI-generated.

The game should also encourage players to use recognition tips as the beginning of a verification process rather than as a final decision rule. In addition to examining visual details, players should be encouraged to check the original source, consider the context in which the media appeared, and consult reliable verification resources where appropriate.

Participants generally considered the prototype understandable, helpful, and enjoyable, and playing the game was associated with increased confidence. These findings suggest that a serious game can be an acceptable and engaging way to introduce older adults to AI-generated media.

However, the study cannot establish that the game improved participants' objective recognition ability because it did not include a pre-game classification test or a control condition. The results therefore support conclusions about the prototype's accessibility, acceptability, and perceived educational value, but not its learning effectiveness.

6 Limitations

This study has several limitations that should be considered when interpreting its findings.

First, participant recruitment was affected by practical and environmental circumstances, including a period of unusually hot weather in the Netherlands. Recruitment was primarily conducted through the researcher's personal network and a local community centre. The resulting convenience sample was geographically limited and may not represent the wider population of older adults in the Netherlands.

The sample was also unbalanced in terms of employment status. The number of fully retired participants was considerably larger than the number of employed participants. Consequently, the comparison between these groups was exploratory and does not support a clear conclusion about the relationship between employment status and classification performance.

Second, the game was originally intended to run in web browsers and on tablets. Technical problems with the Unity implementation meant that most participants played the game on the researcher's laptop. The unavailable browser version also prevented the development of an English version within the study period. Participation was therefore limited to Dutch-speaking individuals who could attend a local session or run the Windows application remotely.

The use of different participation settings may also have influenced the results. Most participants completed the study in person on the same laptop, whereas three remote participants used their own computers. Differences in screen size, display quality, playback performance, and testing environment may have affected how the media were presented.

Third, the classification task contained only 20 scored media items. This number was selected to keep each session under one hour and reduce participant fatigue. Consequently, the human, animal, image, video, AI-generated, and authentic categories contained relatively few examples. The results were therefore sensitive to the difficulty and visual characteristics of individual items. In particular, the hamster video and swallow image may have had a substantial influence on the apparent difference between human and animal media.

One video also experienced a playback problem during testing. This technical issue may have appeared to be a defect in the media itself and could have influenced participants to classify the video as AI-generated.

Finally, the study did not include an objective classification test before participants received the recognition tips. The pre-game and post-game surveys measured perceived recognition ability, whereas objective performance was measured only after the participants had completed the tutorial. The study also did not include a control group. It is therefore not possible to determine whether the game improved participants' objective recognition accuracy or whether their performance would have differed without the recognition tips.

The increase in confidence should consequently be interpreted as a change in perceived ability rather than evidence of improved recognition performance.

7 Future Work

Several directions for future research emerged from the findings and limitations of this study.

First, future research could use a longitudinal design. Participants could complete an objective AI-recognition test before using the game, play the game regularly over several weeks, and complete a comparable post-test afterwards. A delayed follow-up test could determine whether participants retained the knowledge and skills introduced by the game.

The pre-test, post-test, and delayed test should use different but similarly difficult media to reduce memory and practice effects. A control group that does not receive the recognition tips could also help determine whether changes in performance are attributable to the game.

Future studies should use a larger and more varied stimulus set. The media should include multiple examples of humans and animals, different image and video styles, and content generated by several AI models. Item order could be randomised or counterbalanced to reduce order effects and determine whether participants become more sceptical as the task progresses.

The game could also be developed into a mobile application containing several learning modules with progressively increasing difficulty. Each module could introduce new information, provide practice opportunities, and conclude with an assessment of the user's knowledge. This structure could be similar to the progressive learning systems used by applications such as *Duolingo* (Duolingo, n.d.) and *Babbel* (Babbel GmbH, n.d.). A browser-based version could be developed using web technologies instead of requiring participants to download a large Unity application. This would make the game easier to access on laptops, tablets, and smartphones and would remove the need to distribute large application files by email.

Future versions should also place greater emphasis on source verification. In addition to teaching users to recognise visual inconsistencies, the game could introduce strategies such as checking the original source, comparing reports from reliable organisations, examining contextual information, and using verification tools.

Finally, future studies should recruit a larger, more balanced, and more diverse participant sample. Recruitment could take place across several regions of the Netherlands and include participants with different educational backgrounds, levels of digital experience, and employment situations. A more balanced sample of employed and fully retired adults would support a more reliable comparison between these groups.

8 Conclusion

This study investigated how a serious game for adults aged 60 years and older could be designed to support the recognition of AI-generated visual media. An educational game was developed and evaluated with 30 Dutch-speaking participants between the ages of 60 and 85.

The findings suggest that a serious game for this target group should use clear instructions, simple controls, large and accessible visual elements, short recognition tips, and understandable feedback. The game should include both convincing AI-generated examples and unusual authentic examples to demonstrate that visual inconsistencies are warning signs rather than definitive evidence of AI generation. Feedback should address both AI-generated media classified as authentic and authentic media incorrectly classified as AI-generated.

Participants achieved a mean classification score of 12.2 out of 20, corresponding to an accuracy of 61.0%. For 21 participants (70.0%), animal media were the category in which they made the greatest number of mistakes. However, this result may have been influenced by particularly difficult animal stimuli and cannot be generalised beyond the media set used in this study.

Fully retired participants achieved a slightly higher mean score than employed participants. However, the difference was small, the groups were unbalanced, and only descriptive statistics were reported. The study therefore provides no clear evidence of a relationship between employment status and classification performance.

Participants generally evaluated the game positively. Most considered it understandable, helpful, and enjoyable, and their self-reported confidence in recognising AI-generated media increased after playing. These findings indicate that the serious-game format was accessible and engaging for the participants in this study.

However, increased confidence does not demonstrate improved recognition ability. Because the study did not include an objective pre-game classification test or a control condition, it cannot determine whether the game improved participants' recognition accuracy. The study therefore supports conclusions about the prototype's accessibility, acceptability, and perceived educational value, but not its learning effectiveness.

Overall, the project demonstrates that a serious game can provide an engaging way to introduce older adults to the challenges associated with AI-generated media. It also identifies several design considerations that can inform the development of more extensive and rigorously evaluated AI-literacy interventions.

Acknowledgements

I would like to thank my thesis advisor for their guidance, patience, and understanding throughout the research and writing process. I would also like to thank my colleagues and the staff and visitors of the community centre, whose support made participant recruitment and the testing sessions possible. Finally, I am grateful to all participants who volunteered their time and contributed to this study.

References

Adcock, M., Fankhauser, M., Post, J., Lutz, K., Zizlsperger, L., Luft, A. R., Guimarães, V., Schättin, A., & de Bruin, E. D. (2020). Effects of an in-home multicomponent exergame training on physical functions, cognition, and brain volume of older adults: A randomized controlled trial. *Frontiers in medicine*, *6*, 321.

Babbel GmbH. (n.d.). Learn a language online – fast & effective [Accessed: 2026-07-22]. https://www.babbel.com/

Bonnechère, B., & Van Sint Jan, S. (2019). Serious games. In B. Pilote & G. Chiniara (Eds.), *Clinical simulation* (2nd ed.). Elsevier. https://doi.org/10.1016/B978-0-12-815657-5.00002-4

Busch, P. A., Hausvik, G. I., Ropstad, O. K., & Pettersen, D. (2021). Smartphone usage among older adults. *Computers in Human Behavior*, *121*, 106783.

Choudhury, A. (2025, July 16). *Gamification as a tool for lifelong learning*. Eurospeak Ireland. Retrieved February 20, 2026, from https://www.eurospeak-ireland.com/erasmus-projects/gamification-as-a-tool-for-lifelong-learning-how-seniors-are-using-games-to-learn-new-skills/

Cooke, D., Edwards, A., Barkoff, S., & Kelly, K. (2024). As good as a coin toss: Human detection of ai-generated images, videos, audio, and audiovisual stimuli. *arXiv preprint arXiv:2403.16760*.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining" gamification". *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, 9–15.

Digital Promise. (2024). Ai literacy [Accessed: 2026-07-27]. https://digitalpromise.org/initiative/artificial-intelligence-in-education/ai-literacy/

Djafarova, N., Dimitriadou, A., Zefi, L., & Turetken, O. (2023). The art of serious game design: A framework and methodology. *AIS Transactions on Human-Computer Interaction*, *15*(3), 322–349.

Duolingo. (n.d.). Duolingo – leer gratis een taal [Accessed: 2026-07-22]. https://nl-nl.duolingo.com/

Es, K., & Nguyen, D. (2024). “your friendly ai assistant”: The anthropomorphic self-representations of chatgpt and its implications for imagining ai. *AI SOCIETY*, *40*, 3591–3603. https://doi.org/10.1007/s00146-024-02108-6

Eun, S.-J., Kim, E. J., & Kim, J. (2023). Artificial intelligence-based personalized serious game for enhancing the physical and cognitive abilities of the elderly. *Future Generation Computer Systems*, *141*, 713–722.

Ghorbani, F., Taghavi, M. F., & Delrobaei, M. (2022). Towards an intelligent assistive system based on augmented reality and serious games. *Entertainment Computing*, *40*, 100458.

Google. (n.d.). *Gemini: Je ai-assistent van google* [Geraadpleegd op 19 juli 2026]. Retrieved July 19, 2026, from https://gemini.google/nl/about/?hl=nl

Guo, Y., Yuan, T., & Yue, S. (2022). Designing personalized persuasive game elements for older adults in health apps. *Applied Sciences*, *12*(12), 6271. https://doi.org/10.3390/app12126271

Gutiérrez-Pérez, B.-M., Martín-García, A.-V., Murciano-Hueso, A., & de Oliveira Cardoso, A.-P. (2023). Use of serious games with older adults: Systematic literature review. *Humanities and Social Sciences Communications*, *10*(1), 1–17.

Kamali, N., Nakamura, K., Chatzimparmpas, A., Hullman, J., & Groh, M. (2024). How to distinguish ai-generated images from authentic photographs. *arXiv preprint arXiv:2406.08651*.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *Iscience*, *24*(11).

Krempl, S. (2026, April). AI fraud: Germans overestimate their ability to detect deepfakes [Accessed: 2026-5-11].

Lu, Z., Huang, D., Bai, L., Qu, J., Wu, C., Liu, X., & Ouyang, W. (2023). Seeing is not always believing: Benchmarking human and model perception of ai-generated images. *Advances in neural information processing systems*, *36*, 25435–25447.

Martínez Sanahuja, L. (2019, June). *Gaming and healthcare: An intelligent way to help our elders* [Master's thesis]. Universitat Politècnica de Catalunya (UPC) – BarcelonaTech.

Martinho, D., Carneiro, J., Corchado, J. M., et al. (2020). A systematic review of gamification techniques applied to elderly care. *Artificial Intelligence Review*, *53*, 4863–4901. https://doi.org/10.1007/s10462-020-09809-6

McLaughlin, A., Gandy, M., Allaire, J., & Whitlock, L. (2012). Putting fun into video games for older adults. *Ergonomics in Design*, *20*(2), 13–22.

Medhat, M., Ayoub, L. W., Daher, M., & Mohamed, K. M. (2025). Ethical considerations in ai-generated content on social media. In R. K. Hamdan (Ed.), *Sustainable data management: Navigating big data, communication technology, and business digital leadership. volume 1* (pp. 611–620). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-83911-5_52

Mittal, A. (2025). Navigating the digital mirage: The impact of deepfakes and ai on older adults' ability to discern authentic content. *International Journal For Multidisciplinary Research*, *7*(5). https://doi.org/10.36948/ijfmr.2025.v07i05.59035

Morris, J. M. (1994). User interface design for older adults. *Interacting with computers*, *6*(4), 373–393.

NCSC. (n.d.). *Ai: Kans of bedreiging?* Retrieved March 5, 2026, from https://www.ncsc.nl/artificial-intelligence/ai-kans-bedreiging

Peng, P., Xu, L., Ng, D. T. K., Lee, C. S. Y., & Chu, S. K. W. (2025). Exploring the impact of ai-generated image, story and song creations on ai literacy and well-being among older adults: A mixed-methods study. *Proceedings of the Association for Information Science and Technology*, *62*(1), 1045–1050.

Salazar Cardona, J., Arango Lopez, J., Gutiérrez Vela, F. L., & Moreira, F. (2024). Meaningful learning: Motivations of older adults in serious games. *Universal Access in the Information Society*, *23*, 1689–1704. https://doi.org/10.1007/s10209-023-00987-y

Shypunova, A., & Poliakova, T. (2026). Evolution of deepfakes: From an internet joke to a perfect illusion.

Tang, E., KangJie, Song, T., Zhu, Z., Li, J., & Lee, Y.-C. (2025). Ai literacy education for older adults: Motivations, challenges and preferences. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–15.

Tekinbas, K. S., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. MIT press.

Tovertafel [Accessed: 2026-5-11]. (n.d.).

Tredinnick, L., & Laybats, C. (2023). The dangers of generative artificial intelligence.

University, C. T. (2024). How to Spot AI-generated Content: Is It Fact or Fiction? [[Accessed 12-05-2026]].

van der Sluis, B. (n.d.). Zien is geloven? Dit moet je weten over deepfake video's — mediawijsheid.nl [[Accessed 13-05-2026]].

Vasir, G., & Huh-Yoo, J. (2025). Characterizing the flaws of image-based ai-generated content. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3706599.3720004

Yang, A. (2023). Welke van deze foto's is gemaakt met behulp van AI? — nationalgeographic.nl [[Accessed 13-05-2026]].

9 Appendix

9.1 Appendix A: AI Disclosure

For this research, AI was used; in this appendix, the researcher explains what they used and how they used it.

AI model: ChatGPT Plus
Version number: GPT-5.6 Sol
Purpose: ChatGPT was used to help structure the thesis, identify inconsistencies in terminology, correct errors in the LaTeX code in Overleaf, and perform statistical analyses based on the data collected in this study. Examples include calculating the mean score of 30 participants, the standard deviation of the participants' ages, and the result of paired-samples t-test. All output generated by ChatGPT was subsequently checked by the researcher.

AI model: Gemini
Version number: 3.6 Flash
Purpose: Gemini was used to create the AI-generated content included in the game, as shown in Appendix 10.2.

AI model: Grammarly Pro
Purpose: Grammarly Pro was used to correct spelling and grammatical errors in the thesis and to provide suggestions for revising sentence structure and wording.

10 Appendix

10.1 Appendix B: Survey questions

Survey questions part 1, before playing the game

1. Are you participating in this test in person (with the student) or online?
2. What is your age?
3. What is your gender?
4. What is your current working status?
5. Do you use social media?
6. Why do you use social media?
7. Have you used AI before?
8. How confident are you about your ability to recognise AI-generated images/videos/audio?
9. I think AI is an ... for society. (Threat or opportunity Likert scale)
10. Learning about AI is important. (Likert scale)
11. Have you ever tried using AI or learning how it works?
12. If not, what has prevented you from learning more about this topic?
13. Do you think that a video game could be an effective way to learn about AI?
14. Why, or why not?

Survey questions part 2, after playing the game

15. What was your score (max. 20)?
16. Was the game easy to understand?
17. Did you have problems while playing the game?
18. If yes, what was the problem?
19. Do you think this is an effective way of teaching?
20. Why, or why not?
21. Did the game teach you something?
22. If yes, what did it teach you?
23. Did you enjoy playing the game?
24. What do you think overall of the game?
25. How confident are you about your ability to recognise AI-generated images/videos/audio?
26. Any comments or remarks?

10.2 Appendix C: Media used in game

The media was generated with the Gemini 3 series. The game included 10 AI-generated and 10 real images/videos. The AI-generated and real media were tagged with one of the two main categories: human or animal. And only the AI-generated media was tagged with one of the 5 subcategories: Socio-cultural implausibilities, anatomical implausibilities, Functional implausibilities, violations of physics and Stylistic artefacts. In this appendix, the media used is shown, and the AI-generated media includes an extra caption explaining why it's AI-generated, highlighting the mistakes in the media.

10.2.1 AI-generated media

Figure 17: Category; AI, Human, Anatomical incorrect, Extra thumb.

Figure 18: Category; AI, Animal, Anatomical incorrect, Dog with 2 tails video

Anatomical incorrect The person in Figure 17 has an extra thumb and the dog in Figure 18 has an extra tail.

Figure 19: Category; AI, Human, Physically incorrect, Walking on water.

Figure 20: Category; AI, Animal, Physically incorrect , Hamster walking video

Physically incorrect The person in Figure 19 walks on the water and not the floor of the pool, as well as the rings of water around the footstep don't add up. The sheet the hamster in Figure 20 is walking on shifts.

Figure 21: Category; AI, Human, Functionally incorrect, Woman with coffeecup.

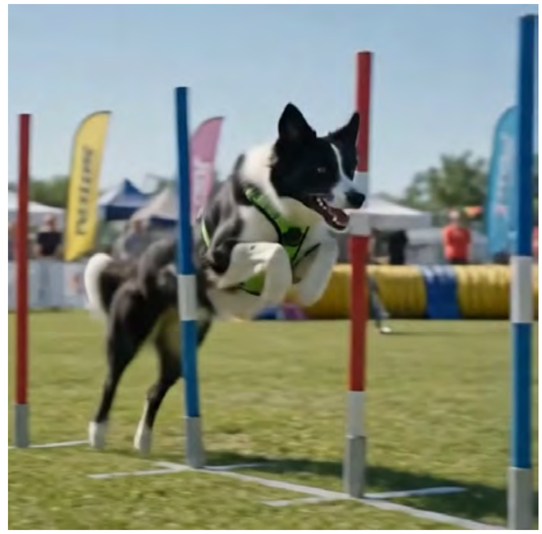

Figure 22: Category; AI, Animal, Functionally incorrect , Dog jumping parkour video

Functionally incorrect The woman in Figure 21 uses the coffee cup upside down, meaning the coffee would fall out. The dog in Figure 22 jumps through the objects of the parkour.

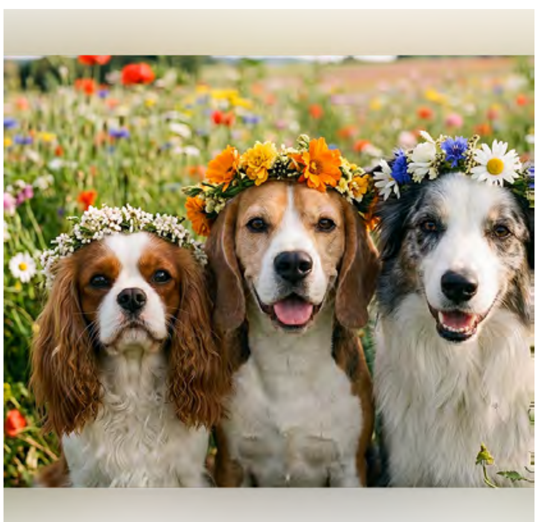

Figure 23: Category; AI, Human, Too good to be true, Dogs in flower field.

Figure 24: Category; AI, Animal, Too good to be true , Man in the rain video

Too good to be true In Figure 23 you see three dogs perfectly posed in a flower field, this is a highly unlikely photo, due to the dogs standing perfectly still for long. Figure 24 shows a man with an umbrella, when opened the rain start immediately pouring.

Figure 25: Category; AI, Human, Incorrect scenario, Squirrel and bunny.

Figure 26: Category; AI, Animal, Incorrect scenario, 2 men sitting at a cafe

Incorrect scenario The Figure 25 shows a bunny and a squirrel fighting for a nut. And the Figure 26 shows 2 dutch politicians that are highly unlikely to go to a cafe together.

10.2.2 Authentic media

The authentic media was found on the internet or taken by the researcher.

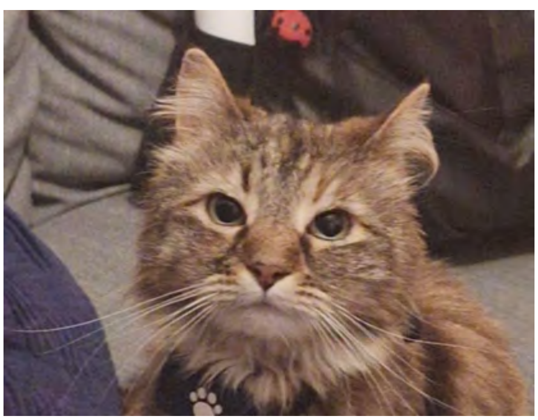

Figure 27: Category; Animal, Authentic, Cat with thick nose, photo taken by researcher

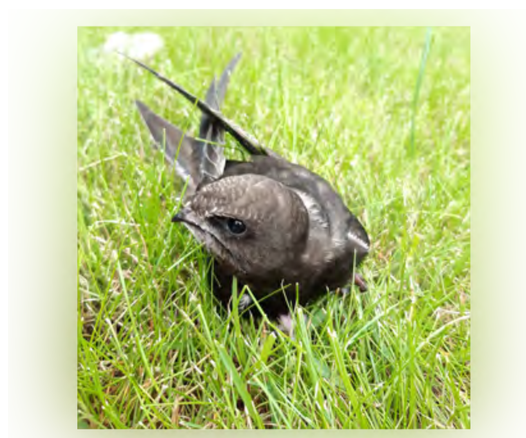

Figure 28: Category; Animal,Authentic, Swallow in the grass, photo taken by researcher

Figure 29: Category; Animal, Authentic, Ducks on hydrangeas, Video taken from Getmotivated519 on Instagram

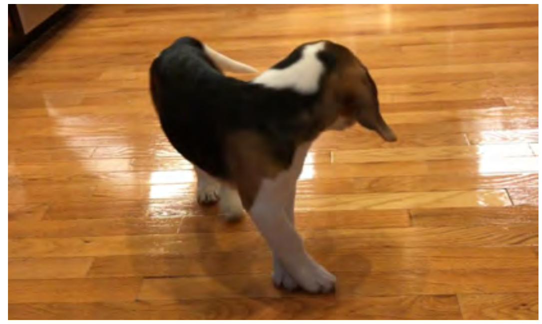

Figure 30: Category; Animal,Authentic, Dog chasing its tail, video taken by Oliver the beagle on Youtube

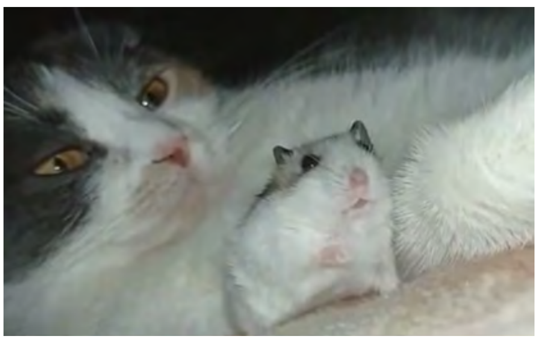

Figure 31: Category; Animal, Hamster and cat, video taken by Hanjurroki Yuko on Youtube

Figure 32: Category; Human,Authentic, elf cosplay in woods, photo taken by Emackphoto on Instagram

Figure 33: Category; Human, Authentic, Angled photo of girl, Photo taken from Xiaohongshu, ID Sr10997

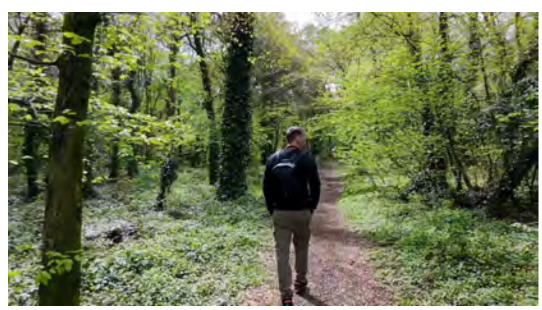

Figure 34: Category; Human ,Authentic, Man walking in the woods, video taken of Pexels

Figure 35: Category; Human, Authentic, Makeup , Photo taken by Mei Pang on Instagram

Figure 36: Category; Human ,Authentic, Man in room with mirrors, video taken of Pexels

10.3 Appendix D: Game screenshots

For this research, a game was developed. This game was developed in the game engine Unity. In this appendix, screenshots Figures 37–50 of what the game looked like. The design choices are clarified in Section 3.3. All the items made in the game, aside from the pictures and videos, were created by the researcher in Photoshop, Procreate, and Canva. As for the game's programming, the Gemini's 3.6 Flash model was used.

Figure 37: First screen

Figure 38: Start screen

Figure 39: Dialogue box

Figure 40: Folder

Figure 41: Anatomical-implausibility explanation

Figure 42: Implausible-scenario explanation

Figure 43: Functional-implausibility explanation

Figure 44: Physics-violation explanation

Figure 45: Too-good-to-be-true explanation

Figure 46: Game explanation

Figure 47: Question format

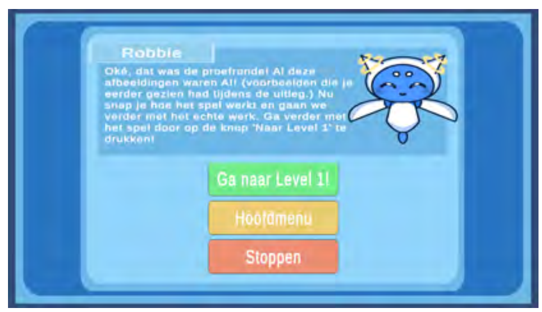

Figure 48: Explanation after the tutorial

Figure 49: Results screen

Figure 50: Mistake-review screen

10.4 Appendix E: Mascot design

Mascott design The game included a mascot who guided participants through it. The sprites were created in Procreate. As for Robbie's design, it was based on a research article by (Es & Nguyen, 2024) about the results she found when googling 'Artificial Intelligence'. The design is based on the findings in that article and the results from today's search engines, such as Google. In Figure 51, you see the first colour concepts, and in Figure 52, the final design is shown. The name was chosen to sound more like a robot name than an AI agent's name.

Figure 51: Colour concepts for Robbie

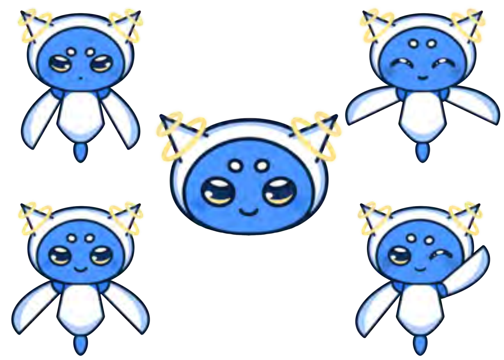

Figure 52: All sprites for Robbie