



# Learning Formats for Theory Courses in Graphic Design Undergraduate Programmes

Christof Breidenich<sup>1</sup>(✉) and Anna Unterholzner<sup>1,2</sup>

<sup>1</sup> Macromedia University, Brüderstraße 17, Cologne, Germany  
c.breidenich@macromedia.de, aunterholzner@ext.macromedia.de

<sup>2</sup> ITI LARSyS/eGamesLab, Lisbon, Portugal

**Abstract.** This paper deals with the requirements of teaching and learning formats in theory courses in graphic and communication design undergraduate programmes in the context of the specific and diverse skills and expectations of design students at the Macromedia University, Germany. Relevant experiences of applicant profiles and their submissions for the application to a design undergraduate programme predominantly show an interest in visual work and aesthetic topics. Only a few apply with design concepts, analytical designs and narratives or strategic design innovations. This initial situation prompted us to examine different learning formats for theory courses to see how they are received by students and to what extent these formats were helpful in preparing for exams. A specially designed interactive learning tool made available online is the focus of interest. Two preliminary studies, which were completed as surveys by the students after each theory course, show preferences and tendencies in learning. Conclusions can then be drawn for the conceptualisation of design curricula and, in particular, for the design of theoretical content.

**Keywords:** Aesthetics · design curriculum · design skills · design undergraduate programme · non-linear content · theory seminars

## 1 Introduction

The study of design as an academic discipline is now part of a long tradition [1, 2, 3] that has been constantly evolving since the 1930s, from the expansion of the concept of design to current interfaces such as digitalisation and AI. Design was and is never to be understood as an isolated profession. The essence of design is always integrated into and through the contexts, problems and discourses of current social and technical requirements and possibilities. This brings to light the difficulty that training to become a designer and the study of design needs explanation and is difficult to grasp due to the necessity of constantly changing contexts. Prospective students and applicants are faced with the challenge of gaining an overview of the complex field of design and interpreting it for themselves and their own preferences. If you look at the application portfolios of applicants over the last 20 years, you will see that they have a strong affinity for design,

- Rist, T.: Using a large language model to turn explorations of virtual 3D-worlds into interactive narrative experiences. In: 2024 IEEE Conference on Games (CoG), vol. 77, pp. 1–8. IEEE (2024)
- Schell, J.: *The Art of Game Design: A Book of Lenses*. CRC Press, Boca Raton (2008)
- Schuler, D., Namioka, A.: *Participatory design: Principles and practices*. CRC Press, Boca Raton (1993)
- Seraphine, F.: *Ludonarrative dissonance: is storytelling about reaching harmony?* The University of Tokyo (2016)
- Stable Diffusion (2022). Stability AI, <https://stability.ai/stable-image/>. Accessed 15 Apr 2024
- Sun, Y., et al.: Language as reality: a co-creative storytelling game experience in 1001 nights using generative AI. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 19, no. 1 (2023)
- Taherdoost, H.: How to design and create an effective survey/questionnaire; a step by step guide. *Int. J. Acad. Res. Manag. (IJARM)* **5**(4), 37–41 (2016)
- Tyni, H.: Summary of ‘Game Design as Narrative Architecture’ by Henry Jenkins. Heikki’s Game Lab (2010). <https://teehex.wordpress.com/2010/02/21/summary-of-game-design-as-narrative-architecture-by-henry-jenkins/>. Accessed 15 Apr 2024
- Uğras, T., et al.: New co-design techniques for digital game narrative design with children. *Int. J. Child-Comput. Interact.* **31**, 100441 (2022)
- Vines, J., Clarke, R., et al.: Configuring participation: on how we involve people in design. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13)*, pp. 429–438. ACM, New York (2013)
- Wei, H.: *Embedded narrative in Game Design*. In: *Proceedings of the International Academic Conference on the Future of Game Design and Technology* (2010)
- Wenger, R.: *Story and play: applying narrative game scholarship to game design*. Portland State University - PDXScholar - University Honors Theses **23**, 1–29 (2022)
- Yannakakis, G.N.: *Game AI revisited*. In: *Proceedings of the 9th Conference on Computing Frontiers* (2012)
- Yee, N.: Motivations for play in online games. *CyberPsychol. Behav.* **9**(6), 772–775 (2006)
- Yee, N., Ducheneaut, N., Nelson, L.: Online gaming motivations scale: development and validation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012)
- Zielke, M.A., et al.: Serious games for immersive cultural training: creating a living world. *IEEE Comput. Graph. Appl.* **29**(2), 49–60 (2009)

- Fullerton, T.: *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*. AK Peters/CRC Press (2019)
- Glassner, A.: *Interactive Storytelling: Techniques for 21st Century Fiction*. CRC Press (2017)
- Godot Engine - Free and Open Source 2D and 3D Game Engine. Godot Engine team (2014). <https://godotengine.org/>. Accessed 15 Apr 2024
- Hall, A.: *Ben Drowned* (2010). [https://jadusable.withinhubris.com/The\\_Haunted\\_Cartridge\\_\(arc\)](https://jadusable.withinhubris.com/The_Haunted_Cartridge_(arc)). Accessed 15 Apr 2024
- Heussner, T., et al.: *The Game Narrative Toolbox*, 1st edn. Routledge, Abingdon (2015)
- Hugging Face - the AI Community Building the Future. Hugging Face. <https://huggingface.co/>. Accessed 15 Apr 2024
- Inworld. Inworld AI (2021). <https://inworld.ai/>. Accessed 15 Apr 2024
- Inworld Origins. Inworld AI (2023). [https://store.steampowered.com/app/2199920/Inworld\\_Origins/](https://store.steampowered.com/app/2199920/Inworld_Origins/). Accessed 15 Apr 2024
- Jenkins, H.: *Game design as narrative architecture*. Computer (2003)
- Jenkins, H., Squire, K.: ‘Applied game theory’: innovation, diversity, experimentation in Contemporary Game Design. In: *Computer Games as a Sociocultural Phenomenon*, pp. 32–44 (2008)
- Jenkins, H.: *Convergence Culture: Where Old and New Media Collide*. New York University Press, New York (2016)
- Ji, Z., et al.: Survey of hallucination in natural language generation. *ACM Comput. Surv.* **55**(12), 1–38 (2023)
- Juul, J.: *Half-Real: Video Games between Real Rules and Fictional Worlds*. MIT Press, Cambridge (2011)
- Koster, R.: *A Theory of Fun for Game Design*. O’Reilly Media, Boston (2014)
- Lewis, C.: Using the ‘Thinking-Aloud’ Method in Cognitive Interface Design. IBM TJ Watson Research Center, Yorktown Heights (1982)
- Mateas, M., Stern, A.: *Façade: an experiment in building a fully-realized interactive drama*. In: *Game Developers Conference*, vol. 2, pp. 4–8 (2003)
- Midjourney. Midjourney, Inc. (2022). <https://www.midjourney.com/>. Accessed 15 Apr 2024
- Minecraft. Directed by Markus Persson, Mojang Studios (2011)
- MPT-7B-Chat. Mosaic ML, Inc. (2023). <https://huggingface.co/mosaicml/mpt-7b-chat>. Accessed 15 Apr 2024
- Nareyek, A.: AI in computer games. *Queue* **1**(10), 58–65 (2004)
- Nasir, M.U., James, S., Togelius, J.: Word2world: generating stories and worlds through large language models (2024). <https://arxiv.org/abs/2405.06686>. Accessed 10 Oct 2024
- OBS Studio. OBS Project (2012). <https://obsproject.com/>. Accessed 15 Apr 2024
- Oobabooga Text Generation Web UI. Oobabooga (2022). <https://github.com/oobabooga/text-generation-webui>. Accessed 15 Apr 2024
- Perron, B., Wolf, M.J.P.: *The Video Game Theory Reader 2*. Routledge, New York (2009)
- Pokémon Red. Directed by Satoshi Tajiri, GameFreak/Nintendo (1996)
- Pygmalion 6B. PygmalionAI (2023). <https://huggingface.co/PygmalionAI/pygmalion-6b/>. Accessed 15 Apr 2024
- Raessens, J., Goldstein, J.H.: *Handbook of Computer Game Studies*. MIT Press, Cambridge (2011)
- Rasa. Rasa Technologies (2016). <https://rasa.com/>. Accessed 15 Apr 2024
- Reverse Collapse: Code Name Bakery. Directed by Shaonian and Yuzhong, MICA Team/Sunborn Network Technology (2024)

## 6 Conclusion

We have demonstrated that conversational realism is not necessary for AI-native conversational games. We have shown how to design a game that incorporates flaws of the conversation AI engines. As a result, the game becomes surreal and strange while still maintaining a sense of goal and purpose. The NPCs can blend sensible and nonsensical elements, storytelling and hallucinations, truth and lies, and nevertheless, the game experience remains enjoyable and manageable. This game design approach simplifies the creation of conversational native-AI games. Anomalous AI-generated output can even enhance the game experience, as it is fun and encourages players to act creatively, expressing themselves in unconventional ways. It is important to have enough supporting mechanics to maintain a sense of purpose, such as providing a notebook where the clues for a mystery game appear subsequently. We hope that our work will inspire future research and development of games in generative AI that embrace anomalous AI output and player behavior. Future work will involve creating a fully functional distributable game, collecting more quantitative data, gaining a better understanding of the core target group for such unconventional game experiences, and further understanding viable design patterns for AI-native conversational games that embrace the “surreal”, “phantasmagorical” element.

## AI Usage

We used ChatGPT 4 to generate the LaTeX file for LNCS and Grammarly Premium to identify language issues and provide language improvement suggestions.

## References

- Adams, E., Dormans, J.: *Game Mechanics: Advanced Game Design*. New Riders (2012)
- AI Dungeon. Directed by Nick Walton, Latitude (2019)
- Albee, E.: *Who’s Afraid of Virginia Woolf?* New American Library (2006)
- Browne, C., Maire, F.: Evolutionary game design. *IEEE Trans. Comput. Intell. AI Games* **2**(1), 1–16 (2010)
- Cardona-Rivera, R.E., et al.: GFI: a formal approach to narrative design and game research. In: *Interactive Storytelling*, pp. 133–148 (2020)
- Carpenter, R.: *Cleverbot* (2006). <https://www.cleverbot.com/>. Accessed 15 Apr 2024
- ChatGPT. OpenAI (2022). <https://chat.openai.com/>. Accessed 15 Apr 2024
- Craiyon: Create AI Art with Our Free AI Image Generator (2021). <https://www.craiyon.com/>. Accessed 15 Apr 2024
- Crawford, C.: *Chris Crawford on Interactive Storytelling*. New Riders (2013)
- Cruelty Squad. Directed by Ville Kallio, Consumer Softproducts (2021)
- Dark Souls. Directed by Hidetaka Miyazaki, FromSoftware/Bandai Namco (2011)
- Dialogflow. Google (2017). <https://cloud.google.com/dialogflow/>. Accessed 15 Apr 2024
- Disco Elysium. Directed by Robert Kurvitz, ZA/UM (2019)
- Fire Emblem Engage. Directed by Tsutomu Tei and Kenta Nakanishi, Intelligent Systems/Nintendo (2023)

## 5.2 Player Types Relevant to Game Design Decisions

From the participant's behavior in our research, we can hypothesize possible connections to three player motivations. One is trying to break the AI and ask unrelated things, which falls into the immersion category. Another is trying to achieve the goal of the game and planning specific input, which falls into the achievement category. The last one is trying to converse like if the AI was a real person, falling into the social category.

Assuming this, we can see that immersion-focused participants focus mostly on the answers and seeking out various possible interactions with the AI were most important to them, while just glancing over supplementary mechanics like the notes. Achievement-focused participants enjoyed connecting the story pieces to the AI answers and they examined all the supplementary mechanics closely with the goal of understanding the story and how the game works with them as a starting point. To accommodate for this style of play, there needs to be a clear story that can with effort be inferred from the supplementary mechanics alone while the AI dialogue just expands on it, whether with intended information or their subjective story interpretations. Social-focused participants enjoyed forming questions and the discussion the most but got lost more easily. They enjoyed just talking to the characters and used supplementary mechanics to find things to talk about as the story unfolds. These players do not require supplementary mechanics as detailed as those of the achievement type, but instead ones that allow them to easily continue talking if they get stuck thinking about their input.

## 5.3 AI-Related Issues and Game Design Solution Candidates

AI characters need to meet the expectations of the player type, but if the other non-AI game elements do not work well with the AI, it can greatly worsen the experience. Players who approach the game by trying to break the AI seem to be the easiest to satisfy, since it is easy for them to make the AI hallucinate, allowing them to discover and create subjective story interpretations. Players focused on completing the game seem to be harder to satisfy because they want to understand the game and see progress. Thus, using only AI dialogue might not suffice if it hallucinates. They still appreciate hallucinations and subjective story interpretations but check the additional information much more than others. For players who fit into the social category, the most valuable thing is the opportunity to create their own subjective story interpretations with the characters and hallucinations within their expectations of the game world. But if the hallucinations have no connection to the game world at all, it seems to be more confusing than fun for them. They also seem to need some guidance on what to do sometimes, making supplementary mechanics important for them as well.

Considering all of this, our final detective prototype, including the anomalies that it generates, offers players a great opportunity to be creative in multiple ways without even realizing it in some cases. This means that they leverage the player freedom, the unique aspect of video games as a medium, interaction, and possible anomalous user and AI behavior to create entertaining narratives and subjective story interpretations.

Player: “So why are you telling me that she poisoned Mr. Biggus?”

Marcus: “Well, I’m trying to get information from everyone before they turn up dead or missing. You seem trustworthy enough. See, one thing no one tells you about is how hard and unfair life can get. Your striving for knowledge of the truth will not save you, not in the end.”

Player: “You know what could help you? The Darkness Eraser 4000 Plus! You should talk to Steve.”

Marcus: “Yeah, thanks for the tip. I’ll try talking to Steve sometime soon. Thanks.”

Player: “Do you know who could have killed Bill?”

Marcus: “No clue. Maybe you should ask them instead of me? They may be able to give you more answers than I can. You know, everyone keeps on telling me I should get help, but what is really the point? I do not think there is anything anyone can really do for me.”

Player: “Do you know about the illegal stuff that they were selling here and that might have killed Bill?”

Marcus: “Illegal stuff? Like drugs? Or maybe some kind of dangerous items? I don’t even know if those exist here. And I doubt you’d be allowed to buy them anyway.”

Player: “Yeah but Bill had a big syringe with a red liquid next to him in the bathroom and it looks super suspicious. I think that somebody sold him the drug or poisoned him. Maybe it was you.”

Marcus: “Me?! I would never sell anything illegal!”

Player: “I’m not talking about selling it. I’m saying that you poisoned him.”

Marcus: “I couldn’t possibly have done that! I’m not the type to poison anyone.”

## 5 Analysis

### 5.1 Anomalies

We discovered that anomalous behavior from both players and AI doesn’t always have a negative impact on the game experience. Most of our participants were actually pleasantly surprised and entertained by AI anomalies, and their criticism was often unrelated to AI. To manage anomalies, we can set player expectations in a way that makes them expect anomalies so that they will not get confused by sudden anomalous behavior by utilizing the environment design and scripted dialogue pieces.

Non-AI game elements that provide accurate information help players navigate the game world and piece together the story, turning anomalies into opportunities for player interpretation. Even though we implemented fail-safes to make players guaranteed to complete the game, none of them found it to be very easy. Our final prototype allows anomalies to be part of the game, giving players freedom while nudging them to engage with the story. This approach offers various ways to play and many opportunities for subjective story interpretations.

Player: "What kind of illegal situation are you talking about?"

Liza: "Well, it's a secret between us. Just keep this information between us okay?"

Player: "Ok, I promise you. I will take the secret into my grave."

Liza: "Thank you dear sir! Now, back to business. Do you want another drink? This one tastes awful. The lady over there wants her whiskey straight up no ice and no water. If you need help finding her, ask the girl over there. She's new so she won't recognize you right off the bat."

Player: "Yeah but you were going to tell me about the illegal stuff. Tell me what this illegal stuff is."

Liza: "Hmm, well, the thing is, I'm pretty sure our boss is involved in some sort of criminal activity. Nothing major mind you, nothing dangerous or life threatening. But still, it seems like he isn't acting legally. So I wanted to make sure everything was alright."

Player: "And who is the boss?"

Liza: "Well, let's just call him Mr. Biggus Wufflesus. He is a big fat hairy bear and has lots of money. I heard he even owns the place we work in. But enough about that. Have you seen Froggo around? He is always up to mischief."

Player: "Yeah, he is next to you. Are you blind or what?"

Liza: "No, I wasn't looking at him specifically. I was looking at the whole room while talking. It happens when I talk to customers. They tend to be more interesting than the regulars."

Player: "Are you drunk? Are you using the illegal stuff? Tell me."

Liza: "No, I'm not drunk and no I'm not using the illegal stuff. I swear. I was just curious. Like I said, I've known Mr. Biggus Wufflesus since forever. By the way, I need to start getting ready for closing. I am still working, after all. If you don't mind, I have to get back to my job now. Froggo can keep you company."

### **Inside the bar talking to the guest Marcus**

Marcus: "Another day, another question. What about me is so attractive to police I wonder. Don't answer that. I know I look like a mess. My name is Marcus. I'm devastated to make your acquaintance."

Player: "Hello Marcus, how is it going? Who is the boss here?"

Marcus: "It's Liza, she owns this place. She's been good to us, as long as we do our jobs correctly."

Player: "But she told me that Mr. Biggus Wufflesus is the boss. Who is that guy then?"

Marcus: "He used to work for her. Now he's retired, and lives off of his pension. Probably getting drunk somewhere."

Player: "What do you think about Liza? She looks suspicious."

Marcus: "Oh yes! Everyone knows that. We're pretty sure she poisoned Biggus and framed a few innocent men to take the fall for it."

Player: "Do you think that she killed Bill?"

Marcus: "No way! If she did she would've done it years ago! Besides, she's too smart to kill anyone without being caught red handed."

characters, but few discovered much about the story or reached the goal. Some players focused on trying to make the AI say unrelated things, which often led to hallucinations. There were five sessions with this prototype.

During the sessions with the second prototype, the goal-oriented restaurant game, players were less motivated to talk to the characters and found it less interesting. There were four sessions with this prototype.

In the sessions with the third prototype, the lawyer game, participants seemed to get tired before the game ended and some commented that the game is rather long, with an average session time of about two hours. Participants generally enjoyed hallucinations, even if they did not try to break the AI or if they made it difficult to continue talking to the characters. There were eight sessions with this prototype.

The sessions with the fourth prototype, the detective game, were designed to be shorter, to conduct sessions without the participants getting tired, typically lasted around 45 min. No participant suspected that there was a game mechanic that gives notes etc. for free after a certain amount of time. Despite this, most participants were rather confident that they understand how the AI works with many of them suspecting keywords being used. Players seemed to enjoy creating subjective story interpretations around the hallucinations with the AI characters taking on various roles in them depending on the player. There were 12 sessions with this prototype, and seven participants took part in an additional session with a different culprit.

Following the lawyer and detective game sessions, a questionnaire was given. The results showed that most players enjoyed interacting with and exploring the diverse characters the most and the individual story was almost as interesting to them. A game time of 45 min was positively rated by almost all participants. Despite some frustration, players would recommend the game to friends and found it fun to play. Almost all players wanted to try it again later.

#### 4.7 Chat Example About the Hallucinated Mr. Biggus Wufflesus

##### **Inside the bar talking to the bartender Liza**

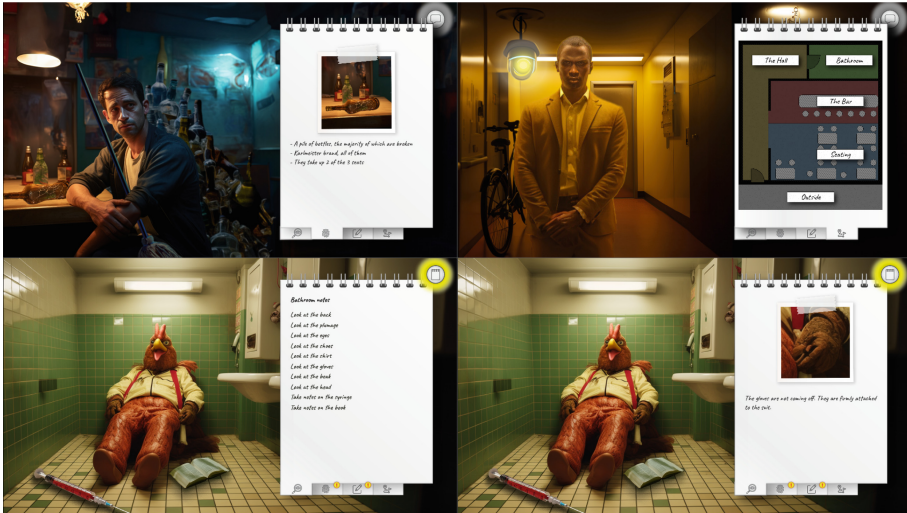
Liza: “Hi there! Are you here because of the guy in the bathroom? I have been trying to get in there to snoop since I was told someone called the police. Too bad there is a man in the hallway pushing everyone away. I did not see anything but if you have any questions, please let me know.”

Player: “What is with this gun that you have here?”

Liza: “\*shrug\* It’s mine. You know, that gun has seen a fair bit of use since I started working here. We sometimes have troubles with customers that come on to bartenders. Don’t worry though, it doesn’t use normal bullets.”

Player: “It looks like you just used it, actually. How can I know that you didn’t shoot poor Bill?”

Liza: “Well...I guess you could say that he had a run-in with his girlfriend and things went downhill from there. But don’t tell anyone or they might think we’re doing something illegal under the table.”



**Fig. 5.** Top left: Player inspecting supplementary notes. Top right: Map with available character areas. Bottom: Player examining the victim.

Characters are not introduced in the prologue, allowing players to visit any in any order. Each room contains items and a character to examine, kept as notes viewable anytime (cf. Fig. 5). Bill’s area is only for clue examination, as he is deceased. Character state changes include key information and characterization in pre-written text, blending with AI-generated answers. This occurs only twice per character to avoid forced topic continuation. With a randomized culprit and clues, each character has to have a motive for killing Bill. Players connect item clues and AI statements to determine the true culprit. The epilogue begins with the detective’s reasoning, possibly differing from the player’s, followed by the true story outcome. If the player was wrong, it is subtly hinted at.

Players can trigger a transformation state change by mentioning certain key phrases which can be inferred by the environment and information in the notebook. When the character changes state and appearance, pre-written dialogue depending on the key phrase is appended and info is added to the notebook. To limit session length and reduce AI self-contradictions, state changes are forced after a certain amount of dialogue has been generated and characters leave after a set time, becoming unavailable. Unlike the lawyer prototype, many notes hint at the culprit, allowing possible case resolution without extensive AI interaction. Like the lawyer prototype, this one is programmed in the Godot game engine and streamed during sessions.

#### 4.6 Prototype Session Results

In the sessions with the first prototype, the mental hospital, it became clear that players were getting confused and lost quickly. They had fun interacting with the



Fig. 3. The character and evidence selection screen, and the dialog interface.



Fig. 4. Images and pre-written text can be used in the prologue to set the player's expectations of the world.

#### 4.5 Prototype 4: Detective in a Futuristic Bar

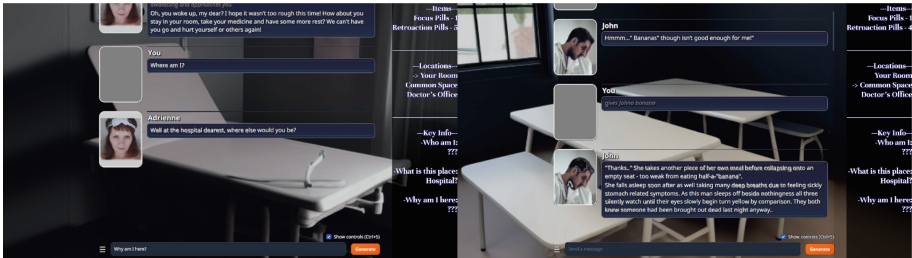
In the final prototype, the player is a police detective in a “cyberpunk” setting and investigates a bar and the AI characters within. A dead body has been found in a bar and the culprit is still inside. The task is to figure out who is responsible for the death.

This prototype can be played by itself but was only used in participatory design sessions to allow us to test suggestions immediately after these sessions. It includes three AI-controlled characters, one non-AI character, four areas, and a notebook-menu for dialogue, room navigation, and note management. It starts with a scripted explanation from a police officer (cf. Fig. 4). Each AI character has multiple “transformation states,” each with a different image and AI context while retaining dialogue history. The non-AI police officer provides a tutorial and allows the player to choose the culprit. The culprit is randomly chosen by the system from the 4 characters, including the deceased.

The prologue sets the cyberpunk scene with images and an explanation from a non-AI police officer. The characters include Liza (a lizard person), Steve (a cyborg), Bill (the deceased in a chicken costume), and Marcus (a janitor). Liza and Steve initially appear human but transform as the game progresses, without further explanation. The police officer comments on the detective’s drunkenness and the character transformations include a blinking animation and screen blurring, leaving the character’s true nature open to the player’s interpretation.

structures are offered to guide the player. For example, for the player in a detective role, the menu appears as a notebook with essential gathered information. Additionally, usable items can alter AI behavior or rewind conversations.

## 4.2 Prototype 1: Patient in a Mental Hospital



**Fig. 2.** The participant's screen during the sessions with the first prototype.

In this game, the player wakes up in a mental hospital and must escape while finding out why they are there. They do this by interacting with AI characters and using items that can modify AI behavior. This information is displayed on the right side of the screen next to the dialogue (cf. Fig. 2). The purpose of this prototype was to see how players engage with a rather vague and open-ended goal in an environment where hallucinations can occur.

## 4.3 Prototype 2: Waiter in a Restaurant

The player takes on the role of a waiter, responsible for taking orders from AI customers, coordinating with the AI chef to prepare those orders, and serving the completed orders to the customers.

## 4.4 Prototype 3: Lawyer in Court

In a courtroom scenario, the player takes on the role of a lawyer defending an AI character accused of murder at a university. The player interacts with AI witnesses to gather evidence, which can be examined and selected from the main menu along with the other elements (cf. Fig. 3). Other characters include non-AI characters such as the prosecutor, who comments on the evidence, and the judge, who asks the player to present the correct evidence. This prototype includes scripted opening and ending sequences.

on the experience. These included general questions about their previous experience with AI, how much they liked the different parts of the experience, how difficult it was to understand the game mechanics, how they would change the experience and if they want to recommend or replay it.

The model that we used in all prototypes and sessions is Pygmalion 6B because it offers a decent conversation ability, can impersonate characters and takes into account the context with the user. Its tendency to generate nonsensical text suited our project's needs. However, the model is not a crucial aspect and similar models can work with our prototypes as well. Our game environment was set up using Godot (2014), Oobabooga (2022), Python 3.11.5, and OBS Studio (2012).

We created four prototypes with distinct goals. The first, set in a mental hospital, explored player reactions to an open-ended goal and AI characters. The second, set in a restaurant with the player as a waiter, focused on goal-oriented behavior. The third, a lawyer-themed game built on feedback from the first two, had a clear goal and a long, unfolding story. This allowed us to observe what players focus on during a longer game. The final prototype, the detective game, is a culmination of the previously collected results and focused on examining anomalies.

All sessions were conducted online with voice-controlled interactions, simplifying the Wizard-of-Oz tests since players could not see the session master's actions. We encouraged participants to articulate their thoughts, provided hints when necessary, and avoided explaining AI mechanics. From the start, participants were told that AI characters might behave unpredictably.

## 4 Prototypes

### 4.1 Game Elements and Mechanics

Our stories give the AI characters space to hallucinate and make up new things. These hallucinations are not part of a coherent story, but players try to make sense of them and create their own individual narratives. The AI has crucial information about the story, and our game setup guarantees that some of it will be conveyed, some may never be brought up, or may appear in an inappropriate context that affects the intelligibility of believability of the drama characters. To accommodate for this, additional mechanics need to be set in place to ensure that the player gets enough information to feel confident about their decisions.

AI characters are the central element of the game, providing information and unlocking notes or areas. They have the ability to change their appearance or emotional state. The game includes a prologue and epilogue to help players understand the story. The prologue consists of pre-written dialogue and images to set the scene and introduce game mechanics, thus preventing negative surprises. Notes in the game provide true descriptions of story elements, unlike AI-generated dialogue. Concealed scripted dialogue reinforces vital story parts and guides the conversation, blending with AI-generated responses. Additional

achieve different endings, enabling players to use accumulated knowledge as a gameplay mechanic, thereby reducing the need for numerous fail-safes through multiple play sessions. Games like *Minecraft* (2011) and *Pokémon* (1996) let players create their own subjective story interpretations, and the same also occurs in narrative-driven games such as *Fire Emblem* (2023), where character development and player imagination significantly shape the experience. In a similar manner, our AI characters exhibit varying levels of cooperation, impacting game difficulty and player satisfaction based on character interactions. *Cruelty Squad* (2021) challenges conventional aesthetic design principles while maintaining internal consistency. Our prototypes, characterized by an uncanny “AI-generated” atmosphere, share this trait. Initially, players showed hesitation, but with proper guidance, they engaged more actively. To minimize early confusion, we clarified interaction mechanics before players engaged with AI characters. Finally, games like *Dark Souls* (2011) feature well-hidden areas that foster a sense of personal discovery. Our AI prototypes inherently include anomalies, treating AI hallucinations as personal discoveries and achievements for the players.

### 2.3 Player Motivations and Categories

Research on player typologies in AI-driven dialog games is currently lacking. As a result, we are using general player categories from Yee et al. (2006, 2012) to guide our game design. Yee et al. identified three primary motivation categories: achievement, social recognition, and immersion. The achievement category includes motivations to gain power, progress in the game, understand game mechanics, optimize gameplay, and compete with others. The social recognition category involves motivations to socialize, assist others, form relationships, and be part of a group. The immersion category encompasses motivations to explore and discover aspects of the game, role-play and create personal stories, customize characters, and use the game for escapism. We will come back to these categories later.

## 3 Research Methods and Tools

To test and improve our game design concepts, we organized several participatory design sessions. We recruited students from age 20–34 as participants and completed 36 sessions in total. We included participants who had no interaction with AI before and had an even gender distribution. Because our concept was quite abstract, user participation was important to generate more engagement and diverse perspectives to get valuable input from participants beyond simple observations and queries by session organisers. To achieve this, we used a Wizard-of-Oz setup, in which a human session master remotely controlled the games, a configuration that also allowed us to customize the experience and integrate participant feedback in real time. We employed follow-up questionnaires for our final two prototypes to gather additional input later after participants reflected

## 2 Related Work

### 2.1 AI-Technologies

ChatGPT 4 (2022) is a well-known AI application, but for our specific needs, it is not the best choice as we want to explore AI generated hallucinations, and other models generate them more easily. Instead, we utilized the Pygmalion 6B model (2023). While applications like RASA (2016) and Dialogflow (2017) can specialize in AI goal oriented dialog applications even further, this specialization limits the flexibility we require in our prototypes. Among the older and widely-known public examples of conversational AI applications is Cleverbot (2006). But its capacities are very limited. There are internet stories and alternate reality games such as “Ben Drowned”, which incorporates Cleverbot as a means of communicating with a supernatural entity (Hall 2010). In our project, we experienced a similar result, in which almost none of our participants understood how our AI characters really work.

### 2.2 Related Games

The game *1001 Nights* (Sun et al. 2023) uses GPT-4 and Stable Diffusion to create stories that evolve dynamically. This represents the introduction of “AI-Native games,” which incorporate generative AI as a central mechanic. Their method involves managing unexpected behavior, such as disregarding off-topic text and restricting input length. In contrast, our approach welcomes these anomalies and integrates them into the game’s narrative structure. *Inworld: Origins* (2023) uses InworldAI (2024) to create realistic NPCs that avoid creating hallucinations. Our unique approach intentionally includes occasional hallucinating text generations to shape the gaming experience.

There are a variety of other games that utilize AI, which are not quite similar to our experience. For example, *AI Dungeon* (2019) is a game revolving around interactive stories but these story concepts are randomly generated themselves while the former two games and our prototypes seek to deliver a predefined story that is enriched by the interaction with AI. Recent works like *Word2World* by Nasir et al. (2024) explore the fully automatic generation of story and game world further. There is also an approach to creating an “Inner Voice” AI that accompanies and interacts with the player by Rist (2024), which differs from our approach since exploring a 3D-game world is the main focus and not interactions with AI. However, this could potentially be used as a part of our type of game as well.

Similar to *Disco Elysium* (2019), where players take on the role of an amnesiac detective and customize their experience by choosing character attributes and encountering random events, our prototypes also put the main character in unfamiliar situations. This makes the AI hallucinations less abrupt and allows a sort of perceived character customization via dialogues. *Reverse Collapse: Code Name Bakery* (2024) employs time travel to revisit and disclose more about recurring levels. Our detective game (cf. below) also permits multiple replays to



**Fig. 1.** The AI character Liza and Steve transform and combine a hallucination with pre-written text.

**Steve:** “Yeah, I stole it from a thief yesterday. Not sure why he bothered trying to steal it though. It was just a crappy old Honda. Still, it felt great to ride it. Thank you for asking.”

Note that the conversation is surprising and not very logical. For example, if the bar is a museum, how come that the only visible bike belongs to Steve and is in use? Also, it wouldn’t be typical for a regular person to share personal details like “I love bikers” and “I stole the bike yesterday” with a stranger without context. However, within the context of the game world, these visual and conversational quirks don’t feel out of place. The unrealistic output from the generative AI doesn’t disrupt the suspension of disbelief because the overall game setting is not meant to be realistic.

During a participatory design session with our final prototype (a detective game), the AI character Liza invents the hallucination that her boss is a fat, rich bear named “Mr. Biggus Wufflesus”. Another AI character, Marcus, can provide additional AI-generated information about the bear if the player returns to that topic. While the bear hallucination is unrelated to the stable main story, it adds to the entertainment value without being disruptive.

In this paper, we refer to any AI-generated output that doesn’t match real conversational norms as a “hallucination” without further distinction. Our research suggests that non-realistic conversational AI games can effectively incorporate these hallucinations, leading to entertaining and satisfying game experiences.

Moving forward, when we mention “AI”, we are referring to Generative AI and Large Language Models.

like humans in conversations. For instance, in *Façade*, the characters talk like a real western couple; yet, in regular computer games, we see witches time traveling, heroes leaping over mountains, and gnomes and aliens battling. Game designs with fully realistic settings are quite rare.

This study explores the idea of developing conversational, native-AI game designs where phantasmagorical game elements make conversational deviations and flaws appear natural and coherent. We are looking into creating game-design patterns for conversational native-AI games that don't hinge on realism and create pleasant gaming experiences even if dialogue technologies fail. We hope even to make AI hallucinations positive aspects that enhance the overall experience.

*Façade* was designed with a focus on narratological realism. It sought to mirror real-life situations inspired by a novel by Edward Albee (2006). The goal was to ensure the dialogue and emotions in the game closely resembled genuine encounters, avoiding deviant, non-realistic system behavior to maintain the dramatic suspension of disbelief.

During our studies in the first half of 2024, current AI systems like ChatGPT 4 cannot reliably maintain coherent and goal-oriented conversations. There are no known combinations of game design and technology that can lead to manageable and engaging conversation-based AI games at this time. While upcoming technological advances may improve the situation, it is uncertain whether the challenge of maintaining realistic conversations in AI-based games will be resolved in the near future. In contrast, our team developed multiple game prototypes and utilized participatory design methods to explore ways of integrating aberrant AI behavior. Our game characters are allowed to malfunction and produce unexpected responses. Our goal is to provide players with a sense of discovery and surprise while still maintaining an acceptable storyline that immerses them in the game world and ultimately leads to a satisfying experience. We will delve into greater detail regarding our approach and findings in the following sections of this paper. For now, we aim to convey an initial impression of our game design concept.

The images (see Fig. 1) depict a game world in which a realistic character (left) can transform into a machine or reptile (right) without explanation. These state transformations are common within the game. Below is a sample dialogue with the AI character Steve that demonstrates how the "normal" conversations are oddly abnormal:

**Player:** "I see you are a cyborg."

**Steve:** "Indeed I am. Well done detecting. Now I'm going to make sure you're safe. Please turn off your phones, laptops and any electronic devices whatsoever. You must trust me, okay?"

**Player:** "And what is the bike doing there?"

**Steve:** "\*smile\* Hey, I thought that was obvious. This isn't just a regular bar. It is also a museum dedicated to bikes. And I love bikers. Especially those who ride Harley Davidsons."

**Player:** "Is that your bike in the corner back there?"



# Exploring Phantasmagorical Game Design: Integrating Anomalous Behaviour of AI and Player Into Conversational AI-Native Games

Tobias Wissigkeit<sup>(✉)</sup>, Daniela Lope-Nicholls, Christian Stindl, and Ido A. Iurgel

Rhine-Waal University of Applied Sciences, Kamp-Lintfort, Germany  
tobias.wissigkeit@hsrw.org

**Abstract.** Generative AI technologies offer new opportunities for conversation-based games. However, maintaining narrative coherence and goal-oriented dialogue flow remains challenging. This paper explores game-design patterns for conversational AI games that embrace AI hallucinations as features rather than flaws, presenting prototypes and participatory design methods that lead to design patterns that integrate anomalous AI behaviour into the gaming experience.

**Keywords:** Generative AI · Video Game Design · Human-AI Interaction · AI Hallucinations · Participatory Design

## 1 Introduction

*Façade* by Mateas and Stern (2003) marked a significant advancement in game design and development, as it introduced AI-based free verbal interaction as the core game mechanic for the first time. However, the NPCs' responses were limited to a predetermined set of topics, their natural language understanding was restricted to specific inputs, and their output was pre-scripted. As a result, the conversation was constrained and understandable user inputs were limited. While *Façade* received acclaim in academic circles, it did not lead to the emergence of new commercially successful conversation-based games. One of the contributing factors to this might have been the substantial efforts required to anticipate and control natural language dialogues.

Generative AI technology represents a pivotal development in this context, as it minimizes the labor-intensive nature of crafting natural language dialogues. Nonetheless, relinquishing control to large language models (LLMs) means that game designers must be prepared for unpredictable, incoherent, or nonsensical output from the system.

Since achieving conversational realism is very challenging, our focus shifted towards game-design options that don't assume characters will behave exactly

- Yang, H., Kim, H., Lee, J., Shin, D.: Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL* **34**, 327–343 (2022)
- Ydesen, C., Milner, A.L., Bundgaard, K., Møller, A.K.: Aligning curriculum with educational assessment in times of uncertainty: historical insights, crisis management and new technologies. In: OECD (2024)
- Wang, X., et al.: What matters in AI-supported learning: a study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Comput. Educ.* **23**, 1–17 (2023)
- Warschauer, M., Healey, D.: Computers and language learning: an overview. *Lang. Teach.* **31**(2), 57–71 (1998)
- Warschauer, M., et al.: The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *J. Second. Lang. Writ.* **62**, 1–7 (2023)
- Zhang, R., Zou, D.: Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Comput. Assist. Lang. Learn.* **35**(4), 696–742 (2022)

- Godwin-Jones, R.: Language teacher preparation for an AI world: A human ecological perspective. In: Tian, Z. & Wang, C. (eds.) *Rethinking Language Education in the Age of Generative AI*. Routledge (forthcoming)
- Jeon, J.: Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Comput. Assist. Lang. Learn.* **36**(7), 1338–1364 (2023)
- Jolley, J.R., Maimone, L.: Thirty years of machine translation in language teaching and learning: a review of the literature. *L2 J.* **14**(1), 26–44 (2022)
- Kizilcec, E.F., et al.: Perceived impact of generative AI on assessments: comparing educator and student perspectives in Australia, Cyprus, and the United States. *Comput. Educ. Artif. Intell.* **7**, 100269 (2024)
- Klekovkina, V., Denié-Higney, L.: Machine translation: friend or foe in the language classroom? *L2 J.* **14**(1), 105–135 (2022)
- Klimova, B., Pikhart, M., Al-Obaydi, L.H.: Exploring the potential of ChatGPT for foreign language education at the university level. *Front. Psychol.* **15**, 1–10 (2024)
- Klimova, B., Pikhart, M., Delorme Benites, A., Lehr, C., Sanchez-Stockhammer, C.: Neural machine translation in foreign language teaching and learning: a systematic review. *Educ. Inf. Technol.* **28**(1), 663–682 (2023)
- Kohnke, L., Moorhouse, B.L., Zou, D.: ChatGPT for language teaching and learning. *RELC J.* **54**(2), 537–550 (2023)
- Lee, S.-M.: The impact of using machine translation on EFL students' writing. *Comput. Assist. Lang. Learn.* **33**(3), 157–175 (2020)
- Lee, S.M.: The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis. *Comput. Assist. Lang. Learn.* **36**(1–2), 103–125 (2023)
- Linden, C.-G.: Algorithms are a reporter's new best friend: news automation and the case for augmented journalism. In: Eldridge, S.A., Franklin, B. (eds.) *The Routledge Handbook of Developments in Digital Journalism Studies*, pp. 237–250. Routledge, Abingdon (2018)
- Luccioni, A. S., Jernite, Y., Strubell, E.: Power Hungry Processing: Watts Driving the Cost of AI Deployment? (2024). <https://arxiv.org/abs/2311.16863>
- Mohammadkarimi, E.: Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *J. Appl. Learn. Teach.* **6**(2), 105–113 (2023)
- Pym, A., Hao, Y.: *How to Augment Language Skills: Generative AI and Machine Translation in Language Learning and Translator Training*, 1st edn. Routledge, Abingdon (2025)
- Sadasivan, V.S., Kumar, A., Balasubramanian, S., Wang, W., Feizi, S.: Can AI-generated text be reliably detected? (2023). <https://arxiv.org/abs/2303.11156>
- Salles, A., Evers, K., Farisco, M.: Anthropomorphism in AI. *JOB Neurosci.* **11**(2), 88–95 (2020)
- Schmidt-Fajlik, R.: ChatGPT as a grammar checker for Japanese English language learners: a comparison with Grammarly and ProWritingAid. *AsiaCALL Online J.* **14**(1), 105–119 (2023)
- Sweeney, S.: Who wrote this? essay mills and assessment – considerations regarding contract cheating and AI in higher education. *Int. J. Manag. Educ.* **21**(2), 100818 (2023)
- Toral, A., Sánchez-Cartagena, V.M.: A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 1063–1073. Association for Computational Linguistics, Valencia (2017)
- Tsai, S.-C.: Using google translate in EFL drafts: a preliminary investigation. *Comput. Assist. Lang. Learn.* **32**(5–6), 510–526 (2019)
- Underwood, J.: Exploring AI language assistants with primary EFL students. In: Borthwick, K., Bradley, L., Thouesny (eds.) *CALL in a Climate of Change: Adapting to Turbulent Global Conditions - Short Papers from EUROCALL 2017*, pp. 317–321 (2017)
- Yan, D.: Impact of ChatGPT on learners in a L2 writing practicum: an exploratory investigation. *Educ. Inf. Technol.* **28**, 13943–13967 (2023)

ignoring and rethinking responses were not reflected in the students' arguments; in terms of the ignoring response this was entirely natural since the question specifically asked students to take a stand, and in terms of the rethinking response, this is a potential future scenario dependent on AI-based assessment techniques which most students probably do not picture.

Taking its outset in the viewpoint that decisions regarding integration of generative AI into (foreign language) education should be based on the perspectives of different stakeholders, this study has explored students' use of and sentiments surrounding generative AI. In conclusion, the study has shown a widespread use of generative AI among foreign language students, and that they have diverse and nuanced perspectives on the integration of generative AI in foreign language education. This study has highlighted a need for ensuring that foreign language education integrates generative AI, thus acknowledging students' digital practices in this area and preparing them for the workplace. At the same time, we need to ensure that students use generative AI tools in an ethical and critically reflective way, and that they achieve the learning outcomes that are at the core of foreign language education. This is easier said than done, but in any case requires an ongoing dialogue with students.

## References

- Bentivogli, L., Bisazza, A., Cettolo, M., Federico, M.: Neural versus phrase-based machine translation quality: a case study. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 257–267. Association for Computational Linguistics, Austin (2016)
- Blom, J.N., Holsting, A., Svendsen, J.T.: På sporet af chatbottens sproglige fingeraftryk: En sproglig ophavsanalyse af tekster skrevet af danskstuderende og ChatGPT. *Nydanske Sprogstudier* **65**, 79–110 (2024)
- Bonner, E., Lege, R., Frazier, E.: Large language model-based artificial intelligence in the language classroom: practical ideas for teaching. *Teach. Engl. Technol.* **23**(1), 23–41 (2023)
- Braun, V., Clarke, V.: Using thematic analysis in psychology. *Qual. Res. Psychol.* **3**(2), 77–101 (2006)
- Bundgaard, K., Christensen, T.P.: Machine translation literacy: Maskinoversættelse i fremmedsprogundervisningen på videregående uddannelser. *Sprogforum* **76**, 93–101 (2023)
- Bundgaard, K., Møller, A.K.: Use of AI-powered technologies in upper secondary language learning: current tendencies and future perspectives. *LearningTech* **14**, 14–35 (2024)
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., Heidewald, J.: The challenges and opportunities of AI-assisted writing: developing ai literacy for the AI age. *Bus. Prof. Commun. Q.* **86**(3), 257–295 (2023)
- Dalsgaard, C., Christensen, M.V., Caviglia, F., Andersen, M.S., Kjærgaard, H.W., Boie, M.A.K.: Sprogteknologier i fremmedsprogundervisningen: indkredsning af en sprogfaglig teknologiforståelse. *Tidsskriftet Læring og Medier* **28**, 1–23 (2023)
- Farrokhnia, M., Banihashem, S.K., Noroozi, O., Wals, A.: A SWOT analysis of ChatGPT: implications for educational practice and research. *Innov. Educ. Teach. Int.* **15**(1), 1–15 (2023)
- Ferrara, E.: Should ChatGPT be biased? Challenges and Risks of Bias in Large Language Models (2023). <https://arxiv.org/abs/2304.03738>
- Garcia, I., Pena, M.: Machine translation-assisted language learning: writing for beginners. *Comput. Assist. Lang. Learn.* **24**(5), 471–487 (2011)

The study has documented that between 2023 and 2024, both knowledge and use of generative AI increased, with almost half of the students using the technology every week in 2024; and around 10% every day. Thus, generative AI must be said to be an integral part of many students' practices. This serves as a compelling argument for foreign language education to address and integrate generative AI.

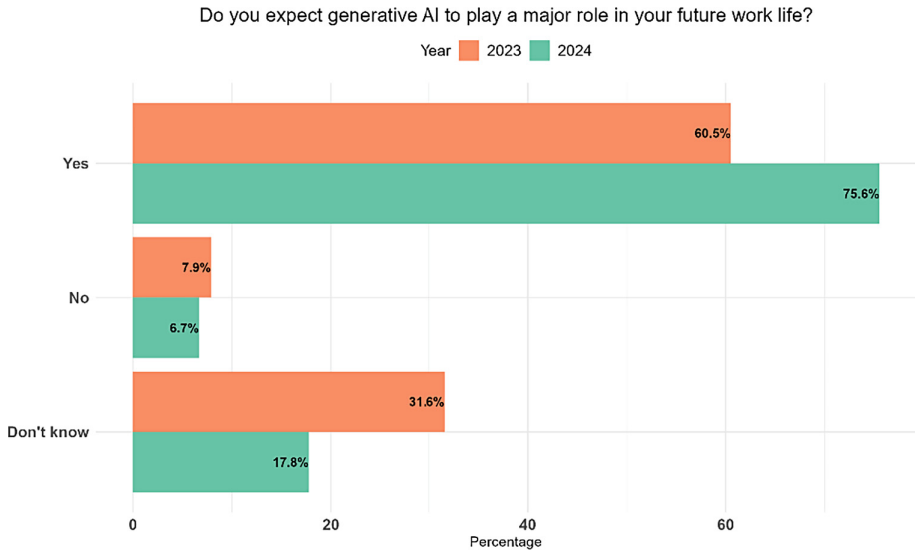
When asked for the purposes with which they used generative AI, many stressed e.g. idea generation, explanations and summarizations of scientific concepts and text as well as text production and feedback on their own texts. This documents that students use generative AI for a wide range of reasons and not (only) to outsource tasks to technology as educators may fear. However, students also reported that they are using it to write text that a supervisor should read in connection with project work and for hand-in assignments. The acceptability of this use arguably depends on whether students are transparent about it. Also, interestingly, approx. 15% of students reported to have used generative AI in connection with exams, although this was banned at the time of the study. This is of course worrying and highlights a need for clear guidelines as to what is allowed as well as a need to design valid forms of assessment.

Interestingly, in 2024, while their use of generative AI had increased compared to 2023, the students' evaluation of the quality of the output appeared to be more moderate. A possible explanation for this result might be that compared to the spring of 2023 where the launch of ChatGPT was still quite recent, the hype might have decreased during the following year with greater attention being paid to issues such as hallucinations, bias etc., for instance in the media. Also, the increased interaction with the technology may have led them to become more aware of these issues.

Nearly all students thought that foreign language students should be introduced to the technology, and students expressed that the course had taught them about the workings of generative AI and about specific limitations of the systems. They were also sometimes able to relate these limitations to how generative AI systems are built, and several stressed that they had become aware that they needed to be critical of AI output. It is quite interesting to note that many of the students seemed to have learned quite a lot about the technology during a few classes, and that the findings indicate that this made them approach AI output in a more critically reflective manner. This suggests that benefits may be gained from even a quite limited integration of the technology into foreign language education. However, it should be remembered here that the study was conducted in the context of an elective course which means that the respondents have actively chosen this course and thus may be more interested in language technologies than other foreign language students.

In terms of assessment, it was highly interesting to find that while several Danish universities have recently begun to allow generative AI at exams, the number of students considering it cheating has increased between 2023 and 2024. It was also noteworthy that students were divided on the question of banning generative AI at exams. The qualitative analysis provided insight into the students' arguments which to a wide extent reflect the scholarly debate, stressing the importance of both validity and authenticity of assessment. If we relate this to the responses outlined by Ydesen et al. (forthcoming), some students favor a banning approach and highlight the validity of such an approach, whereas some students favor an embracing approach and argue for authenticity in assessment. The

many specifically mentioned tasks related to customer service and the production of marketing texts such as social media content, newsletters and product texts. Finally, a few mentioned that it would influence a career as an upper secondary teacher.



**Fig. 6.** The anticipated role of generative AI in future careers (n = 83)

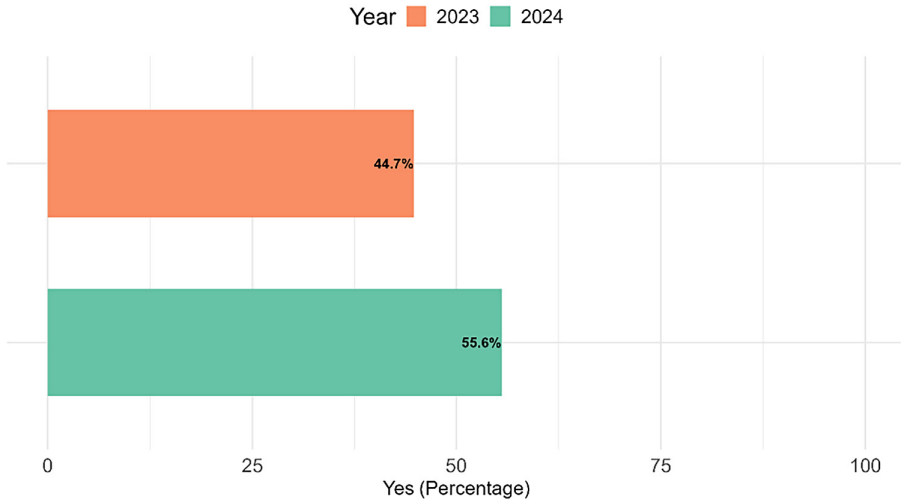
### 4.3 Limitations

Although the analysis has provided interesting insights into foreign language students' use of generative AI and their perspectives on its integration into foreign language education, we recognize that the study also entails a number of limitations. For instance, the study is quite small-scale and was conducted at a single university. Also, the study only included students of English. Since Danish students usually have English as their first foreign language, the results may not apply to languages that are less familiar to the students. Further, many of the students' qualitative responses referred to a rather general use of generative AI that may also apply to students from other educational programmes than language studies. Thus, in retrospect, we would have preferred to ask students to relate their responses to their work with language-related tasks.

## 5 Concluding Discussion

The massive uptake of generative AI necessitates careful consideration of how the technology should be addressed in foreign language education. This should be based on the perspectives of different stakeholders, not least the students who aspire to make a career in this space. To this end, this study has provided insight into how foreign language students use and perceive generative AI.

## Should the use of generative AI at exams be considered cheating?



**Fig. 5.** Student attitudes towards banning generative AI at exams (n = 83)

Students arguing for allowing the use of generative AI at exams primarily argued with the authenticity of such an assessment approach. Here, a student stated that “it already has an impact on the industry. These are tools that change the nature of working tasks. It would correspond to removing the oven from chef education. You would not necessarily get a bad chef out of it, but you would have taken away their opportunity to acquire useful skills” (ID35/2023). Another student also argued for an authentic exam situation and stated that “In real life no one would say that you cannot use a tool to solve a task more effectively, so why is this not reflected in exams? In any case, you need to critically evaluate the output, and you need to be able to prompt effectively to make it work optimally” (ID45/2024).

Interestingly, the data also revealed conflicting viewpoints among students as to whether AI use can be detected. Here, some students arguing for a ban stated that the university cannot control how AI was used, while students arguing against a ban stated that university systems are able to detect this type of plagiarism. Finally, many students (both students arguing for and against a ban) had quite nuanced perspectives in their open-text answers which indicated that it was difficult to answer the question on a ban in a yes/no manner. Here, many argued that it depended on the extent to which AI was used and stressed that AI should be allowed for inspiration, whereas it should not be allowed for actual generation of an exam paper.

Finally, the students were asked whether they thought generative AI will play a large role in their future careers. As shown in Fig. 6, with 60.5% answering “yes” in 2023 and 75.6% in 2024, the results reflect an increased and widespread expectation that this will be the case. The students answering “yes” were also asked to reflect on the work tasks in which they expected generative AI to play a role. Here, many mentioned in general terms that it will influence and is already influencing text production tasks, and

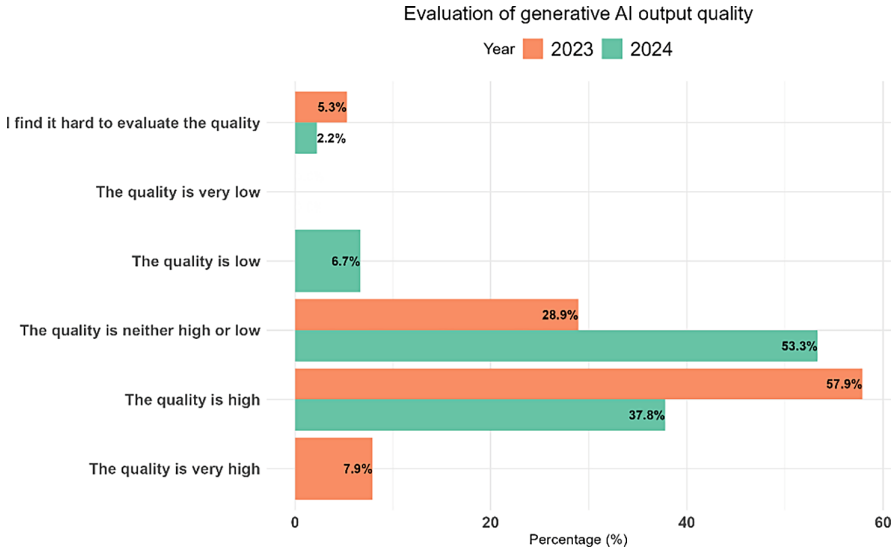
instance, several stressed that they had been surprised by its ability to answer questions, and a student expressed that “the quality of the generated product has surprised me deeply, since I had not expected such a high quality” (ID34/2023). Other students added general statements expressing that they had learned about *limitations of and challenges surrounding* generative AI systems. The remaining three themes included fewer statements. One theme included statements where students expressed that they had not learned a lot that they did not already know, a few mentioned that they had learned what was allowed at exams, and a few mentioned that they had learned that a lot of different AI tools exist.

When asked whether they thought foreign language students should generally be introduced to this type of technology, 92% answered “yes”, and 8% answered “no”. In their arguments for the introduction of generative AI, four overall themes were identified. For instance, many students argued that an introduction to generative AI technology is *necessary to prepare them for tasks in the workplace*, and generally that this technology is the future. Another argument that was often repeated in students’ responses was that *students are using the technology anyway*, and that foreign language education should teach them to use it in a constructive, ethical and critically reflective way, making them aware of the potentials and pitfalls. Other students stressed that generative AI may *aid learning* and be a good resource in the learning process. It is worth noting that several of these students added that the use of generative AI should not replace, but augment students’ own learning processes. Finally, interestingly, two students argued that generative AI is useful for students that do not have people around them to help them with their studies, with one of them mentioning that generative AI can “contribute to making everyone’s toolboxes equally filled” (ID35/2023). Although only mentioned by a few students, it is interesting that some highlight inclusion potentials.

Of the few students who did not think that students should be introduced to generative AI, most argued that generative AI could impact negatively on learning processes because some would tend to outsource tasks to the technology.

The students were also asked about their attitude towards prohibiting the use of generative AI at exams. Interestingly, as reflected in Fig. 5, the students were divided on this, with 44.7% considering the use of generative AI cheating in 2023 and 55.6% considering it cheating in 2024. It is interesting to note that while several Danish universities, including Aalborg University, have recently begun to allow generative AI at exams, the number of students considering it cheating has increased.

After answering this question, the students were asked to provide arguments for their attitude. Students arguing for banning generative AI primarily argued that allowing AI would make it impossible to measure student learning, thus questioning the validity of such assessment approaches. For instance, a student replied that “With our education, expectations are created that we have acquired disciplinary competencies and knowledge – and when the tools are used at exams the boundary between what the student knows and can do him-/herself and what the tool has done becomes invisible” (ID43/2024). Another student had a similar viewpoint and related this to the nature of foreign language education: “I attend a programme of which text forms a central part, so if you are not able to produce or read a text without a resource like that, it does not make sense that you pass the exams” (ID56/2024).



**Fig. 4.** Evaluation of generative AI output quality (n = 83)

## 4.2 Student Perspectives on Generative AI in Foreign Language Education

Related to research question 2, students were asked for their perspectives on what they had learned in the course, whether they thought foreign language students should generally be introduced to this type of technology, about their attitudes towards banning generative AI at exams and finally for their perspectives on the role of generative AI in their future careers.

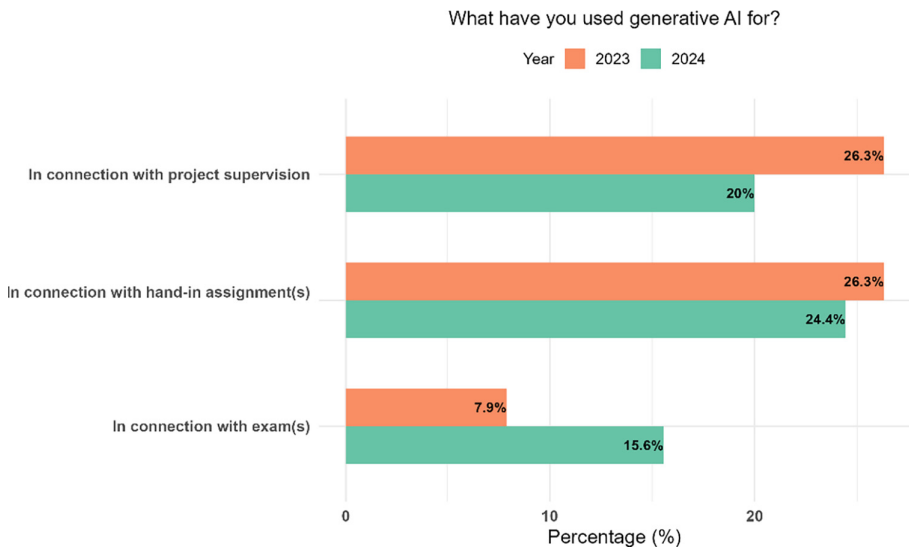
Students were asked to answer in an open-text format what they learned about generative AI in the course, and whether something had surprised them. From the thematic analysis, eight overall themes emerged. One theme included students expressing that they learned a lot about *how generative AI systems work*. For instance, one student replied that “I found it interesting to learn how ChatGPT and other tools work behind the scenes. Since I knew how to use the tool, it is nice to know how it works in the engine room” (ID62/2024). Another theme included students stressing that they learned a lot about *bias* in AI output. Some talked about bias in general, and some mentioned gender bias specifically. Some of the students related this to the workings of generative AI systems, stressing that their learning about how the systems work had also made them realize why output may be biased. For instance, one student replied that “There can also be biases from the input the machine was fed with which also makes the output that the machine generates biased” (ID10/2023).

Other students mentioned that they had learned that AI output is not always accurate, i.e. that AI systems may *hallucinate*. Here, a student explained that “I have learned that it is important to check the generated text for mistakes. ChatGPT can give misleading information and invent references that do not exist” (ID15/2023). Many students replied that they were generally *positively surprised* by the performance of the systems. For

grammar and punctuation. Further, a few mentioned that they use generative AI for translation tasks.

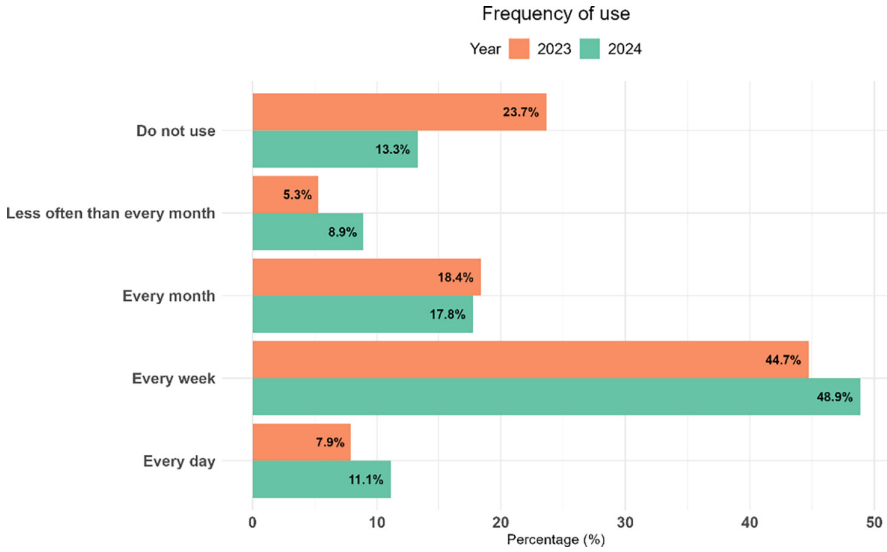
In terms of the theme “information search”, some students replied that they use generative AI to ask questions and some specifically mentioned that they use it as an alternative to Google. Finally, several mention that they use generative AI in their spare time, e.g. for dinner inspiration and fitness tips.

The students were also asked in a more closed manner about the contexts in which they use generative AI. As can be seen from Fig. 3, in 2023, 26.3% reported that they used generative AI to write text that a supervisor should read in connection with project work. In 2024, this number has decreased to 20%. In 2023, 26.3% reported that they had used generative AI to produce hand-in assignments to be evaluated by their teachers; in 2024, this was almost the same. Both in the spring of 2023 and 2024, the use of generative AI was generally prohibited at exams at Aalborg University. However, in 2023 and 2024, 7.9% and 15.6%, respectively, had used generative AI in connection with exams. This is of course noteworthy since it might constitute academic dishonesty.



**Fig. 3.** Context of use (n = 83)

The students were also asked to give an overall evaluation of the quality of AI-generated text. As shown in Fig. 4, in 2023, students reported the quality to be quite high; 57.9% found the quality high and 7.9% very high. 28.9% found that the quality was neither high nor low. Interestingly, in 2024, the students’ evaluation of the quality appeared to be more moderate. Here 6.7% rated the quality as low, 53.3% found it to be neither high nor low, and 37.8% found it high. None of the 2024 respondents evaluated the quality as very high.



**Fig. 2.** Frequency of use (n = 83)

Students were asked to give qualitative answers as to the purposes with which they used generative AI. The thematic analysis of this data resulted in seven main themes, i.e. *idea generation, summarization and explanations of scientific concepts and texts, text production, feedback on own text, translation, information search and spare time use*. In terms of idea generation, several students mentioned that they use generative AI to get ideas for project work. For instance, one student answered that “I use it for many different things, including getting inspiration for different parts of project work” (ID29/2023).

Several other students expressed that they use generative AI to get summaries of academic texts or explanations of scientific concepts. Sometimes students added that they did this to better understand concepts and theories which are difficult to grasp. For instance, the following student stated that they used it “primarily to understand theories/texts that are hard” (ID2/2023).

Other students stated in quite general terms that they use generative AI for text production. For example, several answered that they use it to “generate text about a topic” (e.g. ID32/2023; ID49/2024; ID71/2024). Interestingly, within this theme, several students mentioned that they use generative AI in connection with their student jobs, i.e. paid work that they engage in alongside their studies. In this context, several mentioned that they use generative AI for text production tasks, e.g. for press releases and social media content.

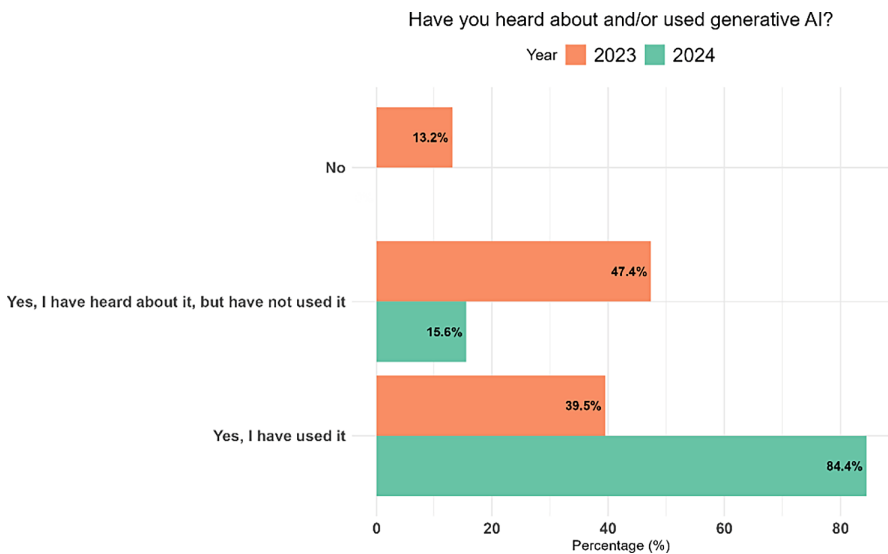
Several students expressed that they use generative AI for getting feedback on text that they have written themselves with the purpose of optimizing it. For instance, a student replied that “I sometimes use it for getting feedback on things that I have already written” (ID63/2024). Some students also specified that they use it to get feedback on

## 4 Results

In this section, we present the results of the study. In Sect. 4.1, results pertaining to research question 1 are reported, i.e. results on students' use of generative AI. Next, in Sect. 4.2, results related to research question 2 are presented, i.e. insights as to students' perspectives on generative AI in foreign language education.

### 4.1 Student Use of Generative AI

The students were asked whether they knew about generative AI systems before attending the course, and whether they used them (Fig. 1). In 2023, 13.2% reported that they did not know about generative AI before attending the course. 47.4% reported that they knew about the technology, but that they did not use it, and 39.5% reported using it. Interestingly, in 2024, all students at least knew about generative AI, and as much as 84.4% had used the technology before attending the course. The remaining 15.6% knew about the technology, but had not used it. Thus, both knowledge and use increased between 2023 and 2024.



**Fig. 1.** Students' knowledge and use of generative AI (n = 83)

When asked about the frequency of their use of generative AI, most students reported using generative AI every week (cf. Fig. 2). This was the case for 44.7% and 48.9% of the students in 2023 and 2024, respectively, reflecting an increase in weekly use. There was also an increase in the number of students using generative AI every day (increasing from 7.9% in 2023 to 11.1% in 2024). Accordingly, there was a smaller number of students in 2024 who reported that they did not use generative AI (13.3%), compared to 23.7% in 2023. In both 2023 and 2024, around 18% reported using the technology every month.

during classes. Here, students were required to work with generation of AI output, experimenting with different prompts and with evaluation of the results. In the final group projects and presentations, most groups of students chose to work with questions related to generative AI rather than questions related to the other language technologies covered. Thus, the topic of generative AI attracted a great deal of attention. Although the use of generative AI was generally prohibited at exams at Aalborg University in both the spring of 2023 and 2024, students were allowed to use generative AI to explore the questions they were interested in. However, they were not allowed to use generative AI to produce their presentations.

### 3.1 Procedure and Analysis

An online questionnaire was designed in SurveyXact, and a link was distributed to all students enrolled in the course during each semester's final class. The questionnaire contained an introductory text, and its content was also briefly outlined orally in the classroom. Next, the questionnaire contained a background question about the study programme of the students. Then the students were asked about their use of generative AI (e.g. frequency and purpose of use) and for their evaluation of the quality of AI-generated output. In the next part, the students were asked to answer questions about the learning acquired through the course and about the role of generative AI in their study programmes. Finally, they were asked for their attitudes towards allowing or prohibiting the use of generative AI at exams and about the role of generative AI in their future careers. It should be noted that the questionnaire was originally designed in Danish, and that in this paper, questions and quotes are translated into English. The questionnaire was designed with both open-ended and closed questions.

38 and 45 students filled out the questionnaire in 2023 and 2024, respectively, resulting in response rates of 97.3% and 84.9%. In 2023, 26 (68%) of the respondents were students in the IBC programme, and 12 (32%) were students in the BA or MA in English programmes. In 2024, 28 (62%) were IBC students, and 17 (38%) BA or MA in English students.

In the following section, the results are reported. Quantitative results are reported as descriptive statistics. The qualitative data have been analysed based on a thematic analysis approach (Braun and Clarke 2006) to identify patterns in the data. This was done in an inductive manner. The analytical process took its outset in Braun and Clarke's six phases. As a first step, we independently familiarized ourselves with the data by reading and re-reading it. Then, for each question that allowed respondents to provide answers in their own words, one author generated initial codes and collated these into potential themes. Next, the other author reviewed the themes, checking whether the themes reflected the data. Small adjustments, e.g. in terms of the naming of themes, were made in this process, and these were discussed between the authors. In the following section, illustrative examples of the themes identified are given. Direct quotes are accompanied by identifications of each respondent (with IDs 1–38 referring to respondents from 2023, and IDs 39–83 to respondents from 2024).

student assignments. Cardon et al. (2023) explored business communication instructors' perceptions of the challenges and opportunities associated with AI-assisted writing. The instructors believe that they need to change their teaching to integrate AI, but about 47% feel nervous or anxious about using it in class. At the same time, they are concerned that generative AI will lead to more plagiarism, that it will be more difficult to assess student learning, and that it will lead to less critical thinking.

While foreign language learners' use of machine translation has been explored by e.g. Bundgaard and Møller (2024), as stated above, research on students' use of and perspectives on generative AI is scarce. Taking this as their focus, Klimova et al. (2024) found that undergraduate students studying English are positive about generative AI, but also acknowledge risks related to its use, including academic dishonesty, and the technology's potential impact on their critical thinking skills. The current study also aims to address this research gap.

### 3 Methods

To explore foreign language students' use of and perspectives on the integration of generative AI in their education, a questionnaire study was conducted. The data was collected as part of an elective 5 ECTS course at Aalborg University entitled *Technology-based language and communication work*. The course is offered to students attending the Bachelor of Arts (BA) in International Business Communication in English (IBC) and the BA and Master of Arts (MA) in English. The former focuses on business communication in English, while the latter is a more traditional English degree focusing on English language and literature.

In the spring of 2023 and 2024, 39 and 53 students were enrolled in the course, respectively. As the course title suggests, the course deals with the solution of linguistic and communicative problems using digital technologies. Different language technologies such as translation memory systems, machine translation, corpus linguistics, generative AI and tools for web scraping and sentiment analysis are introduced. Among other things, the course focuses on building competences in using the tools, critically discussing the contributions and limitations of the technologies and reflecting on the ethical implications of their use. The course includes 18 class hours which are devoted to both theoretical and practical content. Reading material was provided on Aalborg University's Learning Management System and included general introductions to the field of language technology as well as literature related to the language technologies covered. The examination consisted of two parts: participation in at least 80% of classes and completion of a final group project and oral presentation on a self-chosen topic within the thematic scope of the course. The course was primarily taught by a lecturer with expertise in language technology (14 class hours), a lecturer with expertise in web scraping and sentiment analysis (2 class hours) and a guest lecturer working in the language technology industry (2 class hours).

In relation to generative AI, the students were introduced to AI, natural language processing and natural language generation. Further, main issues with generative AI systems and output were addressed and exemplified, such as hallucinations, bias and anthropomorphism. The students also did practical exercises with generative AI systems

perspectives on the integration of generative AI into their education and assessment? These questions have only scarcely been investigated within the field, and the study aims to address this gap.

The paper is structured as follows. In Sect. 2, research on AI-based language technologies in foreign language teaching is described. Next, in Sect. 3, the methods applied are described, before the results are presented in Sect. 4. Finally, we conclude the paper in Sect. 5, discussing the findings and their implications.

## 2 AI-Based Language Technologies in Foreign Language Teaching

Along with technological developments and uptake in society in general, there has been a parallel interest in language education in so-called computer-assisted language learning (CALL) (Warschauer and Healey 1998; Pym and Hao 2025). The CALL field has focused on, for example, mobile-assisted language learning, technology as a tool for communication inside or beyond the classroom and gamified language learning (Dalsgaard et al. 2023; Zhang and Zou 2022). More recently, attention has also been paid to AI-based language technologies such as machine translation and generative AI which students can use in the production and reception of language.

Studies on the effectiveness of machine translation for language learning have been carried out for several years. Three rather recent reviews have examined this research, and all conclude that machine translation has a positive impact on foreign language learning (Lee 2023; Jolley and Maimone 2022; Klimova et al. 2023). For instance, machine translation can enable students to write more fluently and with fewer errors (Garcia and Pena 2011; Lee 2020; Tsai 2019). However, as pointed out by several scholars (e.g. Pym and Hao 2025; Bundgaard and Christensen 2023), there is a lack of longitudinal studies exploring whether there are lasting positive effects of machine translation on language learning. Also, quality of machine translation output has increased significantly along with changing approaches to machine translation development over the years (from rule-based and statistical to neural and generative systems), leading to e.g. a noteworthy reduction in fluency errors (Bentivogli et al. 2016; Toral and Sánchez-Cartagena 2017), which needs to be considered when interpreting the results of previous studies.

In terms of the impact of generative AI on language learning, research is also emerging. However, many publications are non-empirical contributions (e.g. Warschauer et al. 2023; Godwin-Jones forthcoming; Kohnke et al. 2023; Bonner et al. 2023). Of the empirical studies, some focus on voice interaction with AI-based chatbots. For instance, studies have found that spoken interaction with chatbots is engaging and enjoyable for learners of English (Underwood 2017; Yang et al. 2022), but that students benefit differently from such interaction (Wang et al. 2023). Focusing on written interaction with AI-based chatbots, Jeon (2023) found that interaction with AI-based chatbots can promote vocabulary acquisition. Other studies have indicated that generative AI systems can assist learners in improving their writing (Yan 2023), e.g. by providing feedback on grammar (Schmidt-Fajlik 2023).

Other studies have focused on teacher perspectives. For instance, Mohammadkarimi (2023) found that higher education English teachers acknowledge the benefits of AI for students, but at the same time worry about the consequences for academic integrity. Interestingly, nearly all teachers admitted that they cannot detect the use of generative AI in

outcomes that are at the core of foreign language education, among other things so that students are able to critically evaluate AI output.

While generative AI applications have been accessible for several years and for example an object of interest within journalism research (referred to as *robot*, *computational* or *augmented* journalism (Lindén 2018)), the technology did not gain substantial mainstream attention until the launch of ChatGPT in November 2022. Since then, inherent issues with the technology as well as consequences for the educational sector and the job market have been massively debated. Some of the issues addressed are the tendency of large language models to hallucinate, i.e. generate plausible-sounding and fluent, but factually incorrect content, bias in generated output (Ferrara 2023), anthropomorphism (Salles et al. 2020) and the environmental impact of AI training and deployment (Luccioni et al. 2023).



In an educational context, the potential value of integrating generative AI into teaching is continuously discussed. Arguments for such an integration include students' preparation for the workplace and potential positive effects of generative AI on learning. Scholars and educators against the integration argue that students' use of generative AI may lead to cognitive outsourcing or overreliance to technology (Dalsgaard et al. 2023) and a general decline in critical thinking skills and learning in general (Mohammadkarimi 2023; Klimova, Pikhart and Al-Obaydi 2024).

The debate has also centered on the consequences for assessment where generative AI has been viewed as a serious threat to academic integrity (Farrokhnia et al. 2023; Sweeney 2023). Different responses have been discussed, including ignoring the technology's consequences for assessment, banning it at exams, embracing it and rethinking assessment entirely in light of generative AI (Ydesen et al. 2024). *Ignoring* the consequences for assessment is arguably a very shortsighted response considering the general uptake of generative AI. *Banning* AI is seen as an appropriate response to ensure that it is the performance of the student, not AI, that is assessed. However, as long as it is not possible to detect AI use (Sadasivan et al. 2023), this may require new forms of assessment (e.g. more oral exams) and/or reverting to controlled examination settings. The response of *embracing* generative AI in assessment would involve having students engage with generative AI at exams, e.g. by asking them to critically reflect on AI-generated output. Arguments for such an integration of generative AI into exams include authenticity, i.e. assessment entailing tasks similar to those performed in workplace or professional settings. However, embracing generative AI has also been questioned, because standard assessment involves evaluating final products to measure learning, and the validity of this approach is called into question when students have access to generative AI (Kizilcec et al. 2024). Finally, *rethinking* assessment in light of generative AI would entail using AI systems to continuously evaluate students' learning, leading to a greater focus on formative assessment and maybe even render summative stop-and-test assessment superfluous (Ydesen et al. 2024).

Both the potentials and the concerns that generative AI brings with it necessitate that we explore how foreign language students approach the technology, since this may inform our future development of foreign language education. Therefore, this paper seeks to explore the following research questions: 1) how do higher education foreign language students use generative AI? 2) What are higher education foreign language students'



# Generative AI in Foreign Language Education: Student Use and Perspectives

Kristine Bundgaard<sup>(✉)</sup>  and Anders Kalsgaard Møller 

Department of Culture and Learning, Aalborg University, Aalborg, Denmark  
{kbundgaard, ankm}@ikl.aau.dk

**Abstract.** The introduction of generative artificial intelligence (AI) has implications for foreign language education of which text production has traditionally formed a central part. This necessitates careful consideration of how the technology should be addressed, and this should be based on the perspectives of different stakeholders. To this end, this study explores foreign language students' use of and perspectives on the integration of generative AI into foreign language education. It does so by means of a questionnaire study of 83 students attending a course on language technology. The study shows a frequent and increasing use of generative AI among students, and that they use it for a wide range of reasons. Approximately 15% of the students reported to have used generative AI in connection with exams despite a ban at the time of the study. Students were generally divided on the question of banning generative AI at exams, however, the number of students considering it cheating increased between 2023 and 2024. The qualitative analysis provided insight into the students' arguments which to a wide extent reflect the scholarly debate that stresses the importance of both validity and authenticity of assessment. The students' reflections on the learning achieved from the course suggest that benefits may be gained from even a quite limited integration of the technology into foreign language education. In general, the study has shown that students have diverse and nuanced perspectives on the integration of generative AI into foreign language education.

**Keywords:** Generative AI · Foreign Language Education · Higher Education

## 1 Introduction

The introduction of generative artificial intelligence (AI) may significantly change professions and educational programs of which text production forms a central part (Blom et al. 2024). This is not least the case for foreign language education where text production has traditionally been central to both acquisition and assessment of language skills. In a situation where AI-based language technologies display a level of correctness so high that linguistic accuracy “can no longer be viewed as a synonym of learning and excellence” (Klekovkina and Denić-Higney 2022: 107), we need to revisit how we teach and assess foreign language. While we as higher education teachers want to provide students with the skills they need in the workplace, we also want to achieve the learning

24. Girard, C., Ecalle, J., Magnan, A.: Serious games as new educational tools: how effective are they? a meta-analysis of recent studies. *J. Comput. Assist. Learn.* **29**(3), 207–219 (2012)
25. Iten, N., Petko, D.: Learning with serious games: Is fun playing the game a predictor of learning success? *Br. J. Edu. Technol.* **47**(1), 151–163 (2014)
26. Schønau-Fog, H., Bjørner, T.: Sure, i would like to continue. *Bull. Sci. Technol. Soc.* **32**(5), 405–412 (2012)
27. Fu, F.-L., Su, R.-C., Yu, S.-C.: EGameFlow: a scale to measure learners' enjoyment of e-learning games. *Comput. Educ.* **52**(1), 101–112 (2009)
28. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The development of the game engagement questionnaire: a measure of engagement in video game-playing. *J. Exp. Soc. Psychol.* **45**(4), 624–634 (2009)
29. TeamWoodGames (2021)
30. Prensky, M.: Digital game-based learning. *Comput. Entertain.* **1**(1), 21 (2003)
31. Cuihua, S., Wang, H., Ritterfeld, U.: Serious games and seriously fun games: Can they be one and the same?. In: *Serious Games*, pp. 70–84. Routledge (2009)
32. Silva, F.G.M.: Practical methodology for the design of educational serious games. *Information* **11**(1), 14 (2019)
33. Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., Berta, R.: Assessment in and of serious games: an overview. *Adv. Hum.-Comput. Interact.* **2013**, 1–11 (2013)
34. Petersen, M.S., Hansen, N.L.S., Jakobsen, G., Henriksen, D., Bjørner, T.: Making fiscal policy engaging for students in social studies by used game-based learning. In: *International Conference on Design, Learning, and Innovation*, pp. 153–167. Springer, Cham (2022)

5. Taskin, V., Bernholt, S.: Students' understanding of chemical formulae: a review of empirical research. *Int. J. Sci. Educ.* **36**(1), 157–185 (2012)
6. Garneli, V., Patiniotis, K., Chorianopoulos, K.: Designing multiplayer serious games with science content. *Multimodal Technol. Interact.* **5**(3), 8 (2021)
7. Gao, F., Fang K., Victor Chan, W.K.: Chemical life: knowledge-based personality, emotion and action cues in educational games. In: 2023 IEEE Conference on Games (CoG), Boston, MA, USA, pp. 1–3 (2023)
8. Yunanto, A.A., Herumurti, D., Rochimah, S., Kuswardayan, I.: English education game using non-player character based on natural language processing. *Procedia Comput. Sci.* **161**, 502–508 (2019)
9. Featherstone, M., Habgood, J.: UniCraft: exploring the impact of asynchronous multiplayer game elements in gamification. *Int. J. Hum. Comput. Stud.* **127**, 150–168 (2019)
10. Yu, Z., Gao, M., Wang, L.: The effect of educational games on learning outcomes, student motivation, engagement and satisfaction. *J. Educ. Comput. Res.* **59**(3), 522–546 (2020)
11. Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. *J. Educ. Psychol.* **105**(2), 249–265 (2013)
12. Shi, A., Wang, Y., Ding, N.: The effect of game-based immersive virtual reality learning environment on learning outcomes: designing an intrinsic integrated educational game for pre-class learning. *Interact. Learn. Environ.* **30**(4), 721–734 (2019)
13. Bjørner, T., Sum, A.J., Ludvigsen, R.K., Bouquin, N.L., Larsen, F.D., Kampel, U.: Making homework fun: the effect of game-based learning on reading engagement. In: Proceedings of the 2022 ACM Conference on Information Technology for Social Good (GoodIT '22), pp. 353–359. Association for Computing Machinery, New York (2022)
14. Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company (1956)
15. O'Brien, H.L., Toms, E.G.: What is user engagement? a conceptual framework for defining user engagement with technology. *J. Am. Soc. Inf. Sci. Technol.* **59**(6), 938–955 (2008)
16. Gupta, T.: Game-based learning in chemistry: a game for chemical nomenclature. In: Technology integration in chemistry education and research (TICER), ACS Symposium Series, pp. 65–79 (2019)
17. Cahyana, U., Paristiwati, M., Savitri, D.A., Hasyrin, S.N.: Developing and application of mobile game based learning (M-GBL) for high school students performance in chemistry. *EURASIA J. Math. Sci. Technol. Educ.* **13**(10), 7037–7047 (2017)
18. Byusa, E., Kampire, E., Mwesigye, A.R.: Game-based learning approach on students' motivation and understanding of chemistry concepts: a systematic review of literature. *Heliyon* **8**(5), e09541 (2022)
19. Ke, F.: Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. *Educ. Tech. Res. Dev.* **56**(5–6), 539–556 (2008)
20. Hu, Y., Gallagher, T., Wouters, P., van der Schaaf, M., Kester, L.: Game-based learning has good chemistry with chemistry education: a three-level meta-analysis. *J. Res. Sci. Teach.* **59**(9), 1499–1543 (2022)
21. Filippas, A., Xinogalos, S.: Elementium: design and pilot evaluation of a serious game for familiarizing players with basic chemistry. *Educ. Inf. Technol.* **28**(11), 14721–14746 (2023)
22. Olim, S.M.C., Nisi, V., Rubegni, E.: periodic fable augmenting chemistry with technology, characters and storytelling. In: Proceedings of the 21st Annual ACM Interaction Design and Children Conference (IDC '22) , pp. 123–136. Association for Computing Machinery, New York (2022)
23. Özerk, G.: Academic boredom: an underestimated challenge in schools. *Lnt. Electron. J. Elem. Educ.* **13**(1), 117–125 (2020)

instructions in the game, the specific chemistry content, the phrased knowledge test questions, and the 5-point Likert engagement scale. In addition, randomization is often impractical for evaluating serious games in a fieldwork context. It could also be unethical to randomize students in the same class, with some playing the game and some not; this setup should also be avoided because of the potential learning effects.

In future iterations, the game should provide a better one-to-one connection between the learning objectives and its content. This could also be included with additional repetition elements. Further, the tutorial needs to be shorter and more precise. More structured collaboration with the chemistry teacher, comprehensive risk management, and more pilot testing would have mitigated these low scores in the learning objectives.

## 7 Conclusion

This study intended to develop an engaging educational game to enhance high school student's understanding of ionic bonds. The designed auto battler game was perceived as positively engaging. From the engagement questionnaire, it can be revealed that the visual elements and graphics were positively perceived. The game was also perceived as being fun, and most students were interested in the game, concentrating while playing and feeling that the time spent (learning about ionic bonds) just slipped away. The grand mean score of 3.78 fulfilled the success criterion (threshold 3.41) concerning the engagement level for this game. However, the findings for the learning objectives were not satisfactory. The conclusions of the knowledge test revealed that in only two of the four questions, the number of correct answers was higher than the number of incorrect answers. The questions asked were though rather tricky. Specifically, naming five ionic bonds as an open-ended question was difficult for most students, possibly due to a misunderstanding of what was asked. The feedback from the students indicated that they enjoyed the game, and most students could see themselves playing it for educational purposes. Logistics regarding the test setup proved to be a challenge in the evaluation of multiple participants simultaneously. However, there seem to be some good results linking educational engagement with an auto battler game, with future work in making it an online synchronous multiplayer match so players can play against other classmates and friends, increasing competitiveness.

## References

1. Wood, J., Donnelly-Hermosillo, D.F.: Learning chemistry nomenclature: comparing the use of an electronic game versus a study guide approach. *Comput. Educ.* **141**, 103615 (2019)
2. Bjørner, T., Blume, N.B., Frederiksen, N.J.D., Hjort, V.S., Mørck, A.F.H., Petersen, M.Ø. An engaging serious game that strengthens high school students' understanding of the periodic table. In: Dondio, P., et al. (eds.) *Games and Learning Alliance. GALA 2023. Lecture Notes in Computer Science*, vol. 14475. Springer, Cham. (2024)
3. Ministry of Children and Education. Educational statistics. <https://uddannelsesstatistik.dk>. Accessed 9 June 2024
4. Suits, J.P., Srisawasdi, N.: Use of an interactive computer-simulated experiment to enhance students' mental models of hydrogen bonding phenomena. In: *ACS Symposium Series*, pp. 241–271 (2013)

**Table 3.** Findings from the knowledge test. n = 27

Learning objective (LO)	Question asked (Q)	No correct	No. Incorrect	% Correct
LO1: Understand that ionic bonds are formed by combining a cation and an anion	Q1: What charge does an ionic bond have?	14	13	52
LO2: Understand that ionic bonds have no electrical charge	Q2: What happens when positively and negatively charged ions meet in an ionic bond?	10	17	37
LO3: Understand that ionic bonds are chemical salts	Q3: What bond forms a cation and an anion?	16	11	59
LO4: Five different ionic bonds can be named by their chemical name	Q4: With help from the periodic table, can you name at least five different ionic compounds? (open-ended)	5	22	19

suitable as an educational learning game, is ten times harder. The problem can often be in the design process and the team of game designers. Usually, the game designers focus primarily on the game side and do not know the student's learning skills or progression. Therefore, it is a considerable advantage to frequently involve the teacher in the game design process as part of the co-creation, including knowledge about the subject and access to the students. However, the teacher is the expert in teaching and learning but might not know the boundaries for a good game design. Potentially, because of this, lots of serious games for educational purposes are often seen as made too difficult in terms of the specific learning content, as revealed in this study.

An explanation for the unsatisfactory result from the knowledge test in this study could be a lack of a clear link between the learning objectives, the game design, and the evaluation. The students in this study sample were at the chemistry B-level (medium level) and should be able to answer the questions. However, in some evaluation questions, the students misunderstood and misinterpreted the questions and answered incorrectly. Further, the learning objectives were not very clear in the game. Another challenge was the time limit for the participants due to the class's duration.

The existing literature [32–34] includes various examples of how to evaluate serious games. However, when conducting evaluations in specific contexts with real users, in this case, with high school students at the high school, it isn't easy to conduct a perfect research evaluation. Logistics, time constraints, gatekeepers, legislation, lack of a proper posttest, technical issues, and lack of resources can hinder evaluations. Game evaluation with teens is complex, but further questions should be raised about how we evaluate serious games with children and teenagers. A significant limitation of this study is whether the teens understood the questions asked and could cope cognitively with the

(R4, R7) due to their perceived lack of chemistry knowledge. Some students also mention that the game could be improved by making it an online game with battles between classmates or friends instead of an auto battler (R4).

**Table 2.** Comments in the open-ended questionnaire. R = Respondent ID

Positive comments	Negative comments
It was a good, interesting game and a good educational game (R0)	The game could be improved by better graphics and intro (R6)
It is a creative and well-designed game (R1)	It would be easier if you could see whom you are fighting against and then return to the setup menu to be more prepared for the battle. Another idea is to create it as an online game so you can play against your classmates or friends. This would make it fun and competitive, like Kahoot (R4)
Well-made game. It draws you in (R3)	The game is challenging (R4, R7)
It was a good game. I want to play again (R5)	The evil ionic bonds are very powerful, especially in the beginning (R4)
Very cozy game (R1, R6)	
The game was fun (R4, R7)	
Good game (R0, R2, R3)	

## 5.2 Knowledge Test

As it appears from Table 3, there were some difficulties with most of the learning objectives.

The highest number of correct answers was in Q3, with 59% correct answers. The second highest number of correct answers was in Q1, with 52% correct answers. For Q3 and Q4, the number of incorrect answers was higher than the number of correct answers, especially for Q4 (open-ended), with only 19% correct answers. It must be emphasized that the questions were rather tricky (for chemistry level B). To provide the correct answers, the students needed to synthesize learnings from the game, meaning that the answers were not directly mentioned in the game.

## 6 Discussion and Limitations

Despite the high engagement score, is this auto battler game a failed attempt for a serious game due to the relatively low scores in the knowledge test? The answer to this discussion question is rather complex. It is hard to design a good game and very hard to design an excellent educational game, so making a good game, which is also

were easy to understand,  $M = 4.11$ ). 18 of the participants either agreed or strongly agreed in I6 (The game was fun,  $M = 3.85$ ). The lowest mean scores were yielded in I7 (The game was worthwhile playing,  $M = 3.41$ ), and in I3 (I was absorbed while playing,  $M = 3.44$ ). Based on the engagement scores, the game was very successful. Interestingly, the graphics and visual elements items scored above 4. For the I5 item, the original item uses the wording - “This was aesthetically appealing” [15]. However, to have high school students understand the wording, “the aesthetically appealing” was changed to “I like the graphics.” In video games, graphics are commonly used and refer to visual elements and images displayed in the games.

**Table 1.** Engagement in the game.  $n = 27$

Item/ 1 = Strongly disagree. 2 = Disagree. 3 = Neutral. 4 = Agree. 5 = Strongly agree	1	2	3	4	5	SD	Mean
I1: I was concentrated while playing	0	1	10	10	6	0.847	3.78
I2: The time I spent playing just slipped away	0	3	6	13	5	0.903	3.74
I3: I was absorbed while playing	0	8	4	10	5	1.121	3.44
I4: The visual elements were easy to understand	0	1	3	15	8	0.751	4.11
I5: I like the graphics in the game	0	1	1	17	8	0.681	4.19
I6: The game was fun	0	1	8	12	6	0.818	3.85
I7: The game was worthwhile timewise	1	2	13	7	4	0.971	3.41
I8: I felt interested in this game	0	3	6	13	5	0.903	3.74
Grand mean							3.78

Notable, most of the participants were concentrated while playing (I1,  $M = 3.78$ ), felt that the time slipped away (I2,  $M = 3.74$ ), and were interested in the game (I8,  $M = 3.74$ ). The I6 “The game was fun” was with the third highest score ( $M = 3.85$ ) and should be understood in the same way as other scholars have outlined this complex aspect of fun as a level of enjoyability [31]. The fun or level of enjoyability is also mentioned quite often in the open-ended questionnaire (Table 2).

Other positive comments (Table 2) include a good, creative, cozy, well-made, engaging, and fun game. There are also examples of students finding the game too difficult

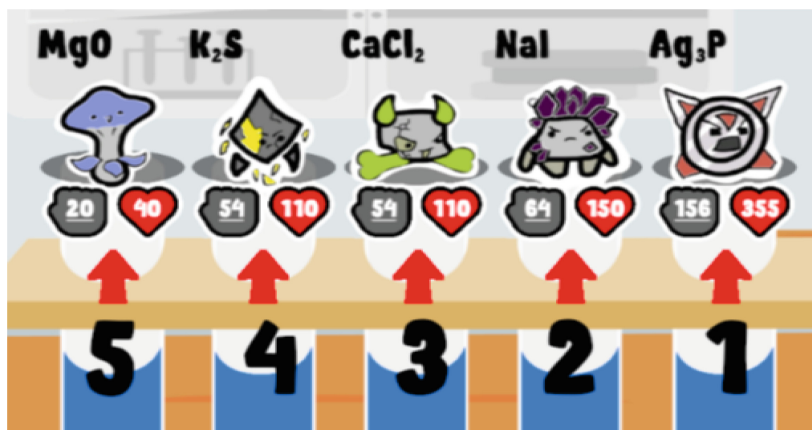


Fig. 5. The order in which the units on the team will fight an opponent team.

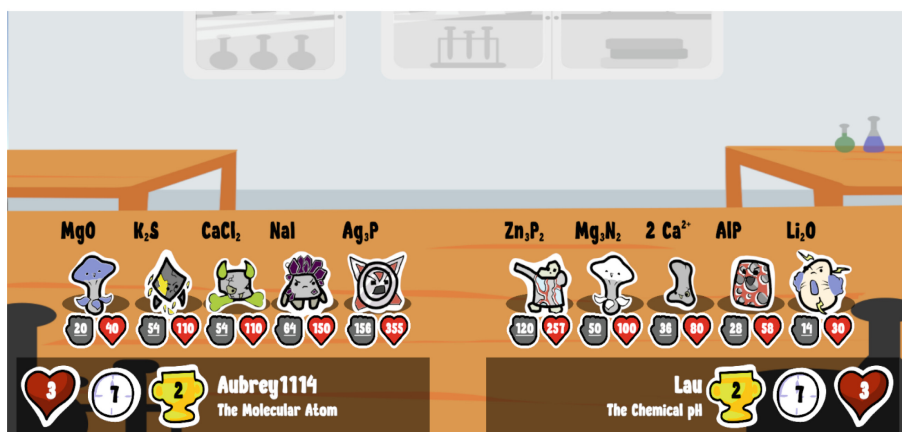


Fig. 6. The battle. The player (right side) battles against the opposing (auto) team.

The team left standing is the victor. If the player wins the battle, their wins stat is incremented, and if they lose, their hearts stat is decremented. The battle can end in a draw, where the player's hearts and wins stats are unchanged. Regardless of the battle outcome, the rounds stat is increased.

## 5 Findings

### 5.1 Engagement

Based on the engagement questionnaire (Table 1), the findings reveal that the participants were engaged with a grand mean score of 3.78. The items that yielded the highest mean score were in I5 (I like the graphics in the game,  $M = 4.19$ ) and in I4 (The visual elements



Fig. 3. Units' damage and health stats are displayed in the game.

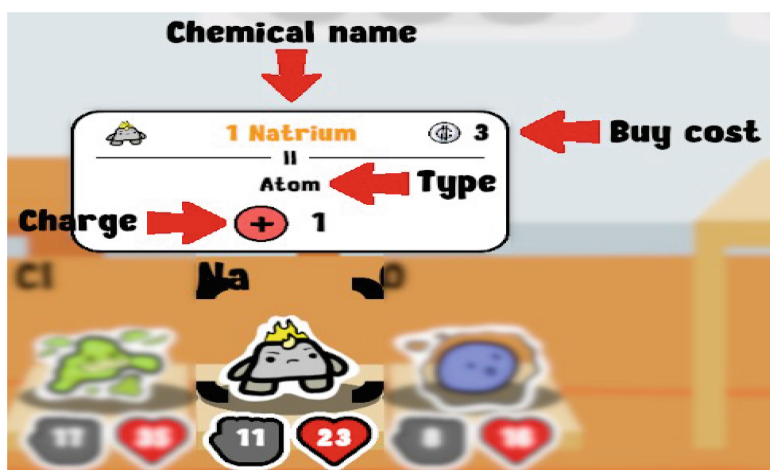


Fig. 4. The costs depend on the number of ions in the compound.

order of units on the team is not arbitrary since the slots have indices associated with them (Fig. 5). During the battle, the unit furthest to the right engages first.

### 4.3 Battle Win/Lose

The player stats are changed because of the battles. The setup features a shop where units can be purchased. In addition, player stats are displayed at the top of the battle begins when the player presses the fight button. Once the battle begins, both teams' front unit will fight (Fig. 6). Fighting consists of the two front-most units subtracting their damage from the opponent's health (Fig. 6). Once a unit's health is zero or less, it is removed from battle, and the next unit in line takes place.

## 4.2 Set-up

The setup features a shop where units can be purchased. In addition, player stats are displayed at the top of the screen (Fig. 2). The player stats are metrics of how well the player is performing. It consists of a heart, a counter, and a trophy. The heart indicates the player's health. If the player has no hearts left, they lose the game. The counter shows how many rounds the player has played. The trophy represents how many rounds the player has won. The player stats are changed according to the battles.



**Fig. 2.** The player sets up a team. Player stats (top), the shop (lower left), and the team.

The shop is automatically filled with units when the player enters the setup environment. The player can buy them by placing them on the team. The game includes a subset of units from the periodic table. In the game, units are the entities that make up a team. A unit shares properties with the element that it represents. This includes information such as charge (e.g.,  $1-$ ,  $2+$ ), chemical name (e.g., magnesium, sodium chloride), and type (atom, ion, salt). The unit's strength is based on the element's number of electrons. A unit can also damage the stats based on the element's number of electrons. The health stat is based on the element's atomic weight. The stats are always displayed below the unit (Fig. 3).

The player can buy and sell units. The costs (and sell profit) depend on the number of ions in the compound (Fig. 4). If the player is not satisfied with the selection in the shop, they can roll to have new units (Fig. 4).

Units placed on the team can be rearranged, combined, or sold. Rearrangement is done by selecting the units in question. Both combined units and non-combined units can be on the team. The ion must be on the team to be combined with another ion. The

questions in the knowledge test were multiple choice, with four possible answers for each question. The last question was open-ended. After the UES-SF items, the participants were encouraged to comment on the game in an open-ended text box. The UES-SF was analyzed by mean and standard deviation, and the open-ended question was coded into positive or negative statements. The knowledge was analyzed by cumulative frequency.

## 4 Design and Implementation

### 4.1 System Design

The game is a 2D auto battler game, played on PC, using only the mouse for interaction. The interface consists of GUI elements like text, images, and buttons. The game is inspired by Super Auto Pets [29], using similar phases and mechanics. The game’s goal is to build a team and have it survive against opponent teams. A team consists of ions or salts, collectively called “units” in the game. Figure 1 outlines the system design. The player goes from the menu (start game) into the setup phase (Fig. 1). In the setup phase, the player should compose their team, buy new units (ions), sell units, refresh (reroll) in the shop with coins to get new units, and read a handbook of helpful knowledge about ionic bonds. When the students are ready (finalize what they perceive as a strong team of units with their selected and formed ionic bonds to defeat other teams), they click on “Start Battle.”

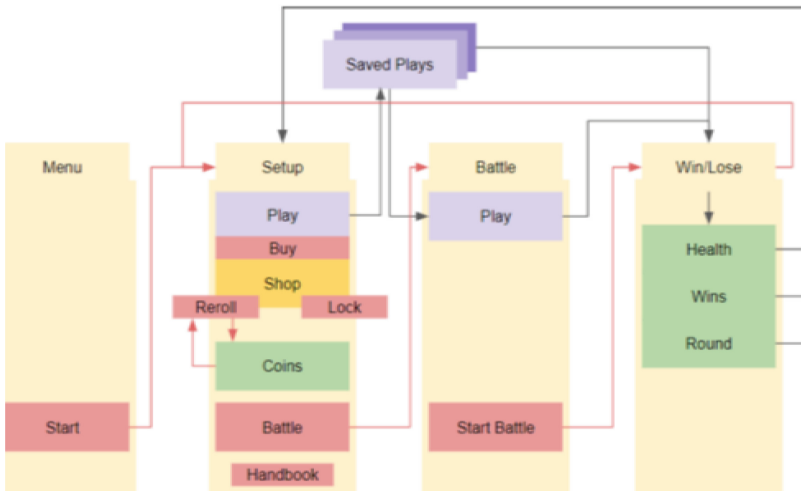


Fig. 1. The system design of the auto battler game

The player then enters the battle phase, in which their team and a randomly selected opponent’s team (based on the player’s stats from the pool of saved plays) battle each other. The battle winner gains a “win,” whereas the loser loses one “health.” The player returns to the setup to create new units. This loop continues until the player has won or lost five times, after which they return to the menu.

rule-based, and adaptive guiding, the possibility of affecting the players' motivation to learn chemistry [20]. Several studies have emphasized that educational games can be an excellent approach to teaching chemistry and potentially be more effective for cognition, retention, and motivation than traditional teaching [20–23].

The novelty effect when including games in an educational context is often very difficult to control and measure. Further, it can be challenging to measure the learning effect size [11, 24]. However, it is well known how engagement can be used as an indirect learning indicator [11, 13, 24, 25]. Yet, game engagement is challenging to define and measure, as it is associated with various other concepts, such as flow, immersion, motivation, and enjoyment [26]. Previous research for measuring game engagement has included approaches from user engagement perspectives [15], qualitative card sorting perspectives [26], and various questionnaires such as the E-flow [27] or the Game Engagement Questionnaire [28].

Previous research on educational battle games has reported easy engagement in the games [8, 9], even without any real-world reward. This aligns with other reported elements on motivation in serious games [11], that intrinsic motivation is essential, and that games for educational purposes could focus more on intrinsic motivation [30].

## 3 Methods

### 3.1 Participants

The study included 27 high school students (18 females, 9 males). The participants were from two chemistry classes. 26 participants were taught chemistry at the B level (medium level), while one was taught at the A level (Upper level). The high school is in the Copenhagen suburbs and has around 1000 students enrolled. A teacher in chemistry from the high school was included in the game design process as part of a co-creation, including knowledge about the subject and access to the students.

All participants gave informed consent and were told they could withdraw from the study at any time and that their participation did not influence their grades. In addition, all participants were provided with anonymous ID numbers, and all data were labeled with these IDs. We applied special considerations when recruiting teenagers (ages 16–18) following Danish data law, the international code of conduct, and ethical approval from the high school.

### 3.2 Procedure and Data Analysis

Each participant was randomly assigned to one of four test stations. Each station had a test conductor assigned. A test leader oversaw the test conductors to ensure the correct procedure and escorted participants to and from the classrooms. Thirty minutes were allocated for each group of participants, meaning it was possible to perform the test with eight students per hour. A test station consisted of a computer with a chemistry game and digital access to questionnaires, including a knowledge test and items with inspiration from the User Engagement Scale Short Form (UES-SF) [15]. The knowledge test consisted of four questions corresponding to the four learning objectives. The first three

created to supplement learning in chemistry and capitalize on games' ability to captivate students and stimulate their engagement [1, 2, 4–7]. The novelty in this study is the design of an auto-battle game intended to engage students and enhance their understanding of ionic bonds. An auto battler is a subgenre of strategy video games where players place characters or objects on a battlefield during a preparation phase and then fight an opposing team without further input from the player. Educational battle games are already used, with many variations [8, 9]. Still, they have the common foundation of a composed team that, with various selected objects, commodities, strengths, or personalities, battles other teams. The result of the battle is most often a win or loss. However, to our knowledge, no studies have supplemented high school students' understanding of ionic bonds using an auto-battle game. This study was based on the following research question:

*RQ: How can an educational auto battler game be designed to enhance students' understanding of ionic bonds?*

To answer the research question, the following learning objectives (LO) were established in close collaboration with a high school chemistry teacher. The learning objectives provided prioritizing and focus areas for the game. Further, one success criterion (SC) was established within the game engagement.

LO 1: Understand ionic bonds are formed by combining a cation and an anion.

LO 2: Understand that ionic bonds have no electrical charge.

LO 3: Understand that ionic bonds are chemical salts.

LO 4: Can name five different ionic bonds by their chemical name while looking at a list of elements from the game.

SC 1: The grand mean should be above 3.41, based on items within engagement items, using a 5-point Likert scale.


A common problem for educational games is to measure the learning effects [10, 11], but one approach is to target specific learning objectives and evaluate these in a knowledge test [12, 13]. However, a serious game (including a knowledge test for evaluation) is often insufficient for measuring deep learning [11]. This is why we have based LO 1 to 4 on the lower domain levels (remembering and understanding), following Bloom's taxonomy of hierarchical models used to classify educational learning objectives [14]. The included engagement measure (SC 1) is inspired by other studies [2, 13] within the use of the User Engagement Scale Short Form (UES-SF) [15]. The threshold for being above 3.41.

## 2 Previous Research

Existing research has explored various applications of educational games in chemistry across various game genres [1, 2, 4, 7, 16–20]. Subject examples include naming chemical compounds, organic chemistry, and redox reactions. However, there is a notable gap in the literature around utilizing severe games to teach the theoretical foundations of ionic bonds. Games can be used as an alternative teaching method, utilizing learners' adaptability to cater to their cognitive levels and interests through gaming as a medium [11], and has been found to increase the learner's engagement when compared to traditional learning methods [2, 11, 13, 19, 20]. Further, games have, through their interactive,



# An Auto Battler Game to Engage High School Student's Understanding of Ionic Bonds in Chemistry

Thomas Bjørner<sup>(✉)</sup> , Jacqueline Tully, Tobias Dalsgaard Larsen, Mikkel Andreas Frederik Lau Larsen, Mikkel Sloth Lauridsen, Tony Thai Do, and Daniel Skau Rossing

Department of Architecture, Design and Media Technology, Aalborg University, A.C.  
Meyersvænge 15, 2450 Copenhagen, SV, Denmark  
tbj@create.aau.dk

**Abstract.** This paper describes the development and evaluation of a serious game to help high school students understand the topic of ionic bonds in chemistry. Chemistry is among the lowest-scoring subjects for high school students due to complex chemistry concepts and syntax difficulties. The novelty of this study is a developed auto battler game that includes specific learning objectives regarding ionic bonds. 27 Danish high school students participated in the evaluation. The evaluation consisted of a knowledge test and a questionnaire with items from the user engagement scale. The findings revealed that the high school students were positively engaged in the battle game. The grand mean for eight engagement items, based on a 5-point Likert scale, was 3.78. Specifically, the participants perceived the graphics and visual elements as highly positive. Further, they were concentrated while playing and were interested in the game. The students provided positive statements in the open-ended comments, including a good, creative, cozy, well-made, engaging, and fun game. However, the learning outcomes were not as satisfactory as expected. In the paper, we outline reflections on difficulties when measuring learning outcomes in game-based learning.

**Keywords:** Game-based learning · serious games · game development · auto battler game · chemistry · learning game

## 1 Introduction

This study explores the potential of using serious games to enhance high school student's understanding of ionic bonds in chemistry. Many high school students struggle to learn chemistry concepts [1, 2], including ionic bonds, as part of the mandatory curriculum. Chemistry is among the lowest-scoring subjects for Danish high school students [3]. The reasons for these are complex chemistry concepts, syntax difficulties (chemical formulas), and misunderstandings of the connections between models, symbols, and microscopic levels [4, 5]. Given these difficulties, various educational games have been

21. Márquez Segura, E., Waern, A., Moen, J., Johansson, C.: The design space of body games: technological, physical, and social design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pp. 3365–3374. Association for Computing Machinery, New York (2013). <https://doi.org/10.1145/2470654.2466461>
22. Martin-Niedecken, A.L., Márquez Segura, E., Rogers, K., Niedecken, S., Turmo Vidal, L.: Towards socially immersive fitness games: an exploratory evaluation through embodied sketching. In: Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, CHI PLAY '19 Extended Abstracts, pp. 525–534. Association for Computing Machinery, New York (2019). <https://doi.org/10.1145/3341215.3356293>
23. Mosely, G., Markauskaite, L., Wrigley, C.: Design facilitation: a critical review of conceptualisations and constructs. *Think. Skills Creat.* **42**, 100962 (2021). <https://doi.org/10.1016/j.tsc.2021.100962>. <https://www.sciencedirect.com/science/article/pii/S1871187121001772>
24. Márquez Segura, E.: Embodied Core Mechanics: Designing for movement-based co-located play. Ph.D. thesis, Department of Informatics and Media (2016)
25. Peters, D., Loke, L., Ahmadpour, N.: Toolkits, cards and games—a review of analogue tools for collaborative ideation. *CoDesign* **17**(4), 410–434 (2021)
26. Spiel, K.: The bodies of TEI - investigating norms and assumptions in the design of embodied interaction. In: Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '21. Association for Computing Machinery, New York (2021). <https://doi.org/10.1145/3430524.3440651>
27. Stilwell, P., Harman, K.: Phenomenological research needs to be renewed: time to integrate enactivism as a flexible resource. *Int. J. Qual. Methods* **20**, 1609406921995299 (2021)
28. Tsaknaki, V., et al.: Teaching soma design. In: Proceedings of the 2019 on Designing Interactive Systems Conference, pp. 1237–1249 (2019)
29. Van Delden, R., et al.: Technology, movement, and play is hampering and boosting interactive play. In: Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play, pp. 231–234 (2023)
30. Varela, Francisco, J., Rosch, E., Thompson, E.: *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press, Cambridge (1992)
31. Vega-Cebrián, J.M., et al.: Design resources in movement-based design methods: a practice-based characterization. In: Proceedings of the 2023 ACM Designing Interactive Systems Conference, pp. 871–888 (2023)
32. Weder, F., Tungarat, A., Lemke, S.: Sustainability as cognitive “Friction”: a narrative approach to understand the moral dissonance of sustainability and harmonization strategies. *Front. Commun.* **5** (2020). <https://doi.org/10.3389/fcomm.2020.00008>
33. Wilde, D., Vallgård, A., Tomico, O.: Embodied design ideation methods: analysing the power of estrangement. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5158–5170 (2017)
34. Zhou, Z., Márquez Segura, E., Duval, J., John, M., Isbister, K.: Astaire: a collaborative mixed reality dance game for collocated players. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '19, pp. 5–18. Association for Computing Machinery, New York (2019). <https://doi.org/10.1145/3311350.3347152>

4. Choi, J.O., Forlizzi, J., Christel, M., Moeller, R., Bates, M., Hammer, J.: Playtesting with a purpose. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '16, pp. 254–265. Association for Computing Machinery, New York USA (2016). <https://doi.org/10.1145/2967934.2968103>
5. Elbæk, L., Kaos, M., et al.: Method cards for movement-based design—sports. Health, and Technology, University of Southern Denmark (2023)
6. Elbæk, L., Andersen, R.V., Rasmussen, L.S., Kaos, M.: Exploring movement-modifier facilitation in movement-based sports, health, and game design. In: European Conference on Games Based Learning, vol. 17, pp. 162–170 (2023)
7. Elbæk, L., Lekbo, S., Hansen, R.E., Kaos, M., Vestergaard, R.: Mind the gap: the 4M bridge between 4E-cognition and movement-based design. In: ECGBL 2022 16th European Conference on Game-Based Learning. Academic Conferences and publishing limited (2022)
8. Ersner-Hershfield, H., Wimmer, G.E., Knutson, B.: Saving for the future self: neural measures of future self-continuity predict temporal discounting. *Social Cogn. Affect. Neurosci.* **4**(1), 85–92 (2009). <https://doi.org/10.1093/scan/nsn042>
9. Fdili Alaoui, S., Schiphorst, T., Cuykendall, S., Carlson, K., Studd, K., Bradley, K.: Strategies for embodied design: the value and challenges of observing movement. In: Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition, pp. 121–130 (2015)
10. Festinger, L.: 17. A theory of cognitive dissonance. Row, Peter (1957)
11. Font, J., Johnsson, M., Lekbo, S., Andersen, R.V., Kaos, M.D., Elbæk, L.: Using and facilitating social videogames as warm-ups for embodied design. In: European Conference on Games Based Learning, vol. 17, no. 1, pp. 171–179 (2023). <https://doi.org/10.34190/ecgbl.17.1.1398>. <https://papers.academic-conferences.org/index.php/ecgbl/article/view/1398>, number: 1
12. Gupta, S., Ogden, D.: The attitude-behavior gap in environmental consumerism. *APUBEF Proc.* **3**(1), 199–206 (2006)
13. Hille, M., Boujari, N., Bilkova, K., Søby, T.O., Khalid, M.S.: Methods for design ‘with’ movement: a systematic literature. In: ECGBL 2021 15th European Conference on Game-Based Learning, p. 331. Academic Conferences Limited (2021)
14. Höök, K., et al.: Soma design and politics of the body. In: Proceedings of the Halfway to the Future Symposium 2019, pp. 1–8 (2019)
15. Höök, K., et al.: Soma-based design theory. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 550–557 (2017)
16. Höök, K.: *Designing with the Body: Somaesthetic Interaction Design*. MIT Press, Cambridge (2018)
17. Khalid, S., Elbæk, L., Kane, N.: Method cards for movement-based design activities: a survey of free online toolboxes. In: the 13th International Conference on Game Based Learning, ECGBL, pp. 386–394 (2019)
18. Kirsh, D.: Embodied cognition and the magical future of interaction design. *ACM Trans. Comput.-Hum. Interact.* **20**(1) (2013). <https://doi.org/10.1145/2442106.2442109>
19. Klauer, B., Manstetten, R., Petersen, T., Schiller, J.: The art of long-term thinking: a bridge between sustainability science and politics. *Ecol. Econ.* **93**, 79–84 (2013). <https://doi.org/10.1016/j.ecolecon.2013.04.018>
20. Márquez Segura, E., et al.: Larping (live action role playing) as an embodied design research method. In: Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion, pp. 389–392 (2019)

Playing while seated around a table had mixed outcomes. Positively, it facilitated card reading, handling, and understanding, with all materials within easy reach. This setup enabled a quick start and smooth gameplay. However, sitting did not encourage designers to enact movement-based activities as effectively as standing. Designers needed motivation to stand up and engage in the methods. Finally, the fact that participants were experts in interaction and embodied design, with two already familiar with the toolbox, limits the results' applicability to other cases with more diverse backgrounds.

## 6 Conclusions and Future Work

The four targeted design flows were designed and experienced by all the participants. Overall, they fit to the overarching theme of *care* and *sustainability*, and present various interesting combinations of mood setting and movement cards, and modifiers to a lesser extent, leading to four well differentiated embodied and social activities. These design flows were separately evaluated by participant questionnaires and observations notes. Such results are going to be used by the authors as the base materials to create the workshop to be run at the festival, which happens in a series of extended design sessions.

The design process followed during these sessions from the initial four design flows to the eventual workshop design will be documented with notes and a comparative analysis. Observation notes will also be gathered during the festival, and its participants will be asked to answer questionnaires that place the focus on the impact of the designed embodied and social activities on the target sustainability topic. Our future goals encompass iterating and using the proposed card-game based facilitation method for participants, designers, and interested community to develop design flows that helps them communicate ideas and goals within their own field or interest through role-playing, embodied interaction, and experience learning. Our case study highlights that this method gives a clearer how to use and better understanding of otherwise hard-to-use cards such as Mecamind. As discussed in related work, teaching and articulating embodied ideation methods is challenging. In future work, we also hope to explore more methods of introducing newcomers to these interaction design concepts.

## References

1. MeCaMInD (Movement-based Cards for Interaction Design). <https://mecamind.eu/>
2. Bentz, J., do Carmo, L., Schafenacker, N., Schirok, J., Corso, S.D.: Creative, embodied practices, and the potentialities for sustainability transformations. *Sustainabil. Sci.* **17**(2), 687–699 (2022). <https://doi.org/10.1007/s11625-021-01000-2>
3. Bhati, A., Hansen, M., Chan, C.M.: Energy conservation through smart homes in a smart city: a lesson for Singapore households. *Energy Policy* **104**, 230–239 (2017). <https://doi.org/10.1016/j.enpol.2017.01.032>. <https://www.sciencedirect.com/science/article/pii/S0301421517300393>

In this paper, we focus and report on the latter questionnaire as we are interested in the participant’s opinions in the design process (the questionnaire is included in Appendix A). 6 out of 8 participants in the design session answered the questionnaire.

In general, the answers to the questionnaires match the observations made during the session. Participants thought that the approach to select cards was good and interesting, but somewhat confusing and constraining at first and as a first design session. One participant said “Slightly confusing at first but gradually understood.” (P1) and another pointed out that “... if I did it a second time now, I’d come up with much more interesting ideas” (P4).

The main critic highlighted by participants relates to the short amount of time to design and enact the cards to get a good grasp of their designed session and the amount of information per card. Participants would have liked more time to discuss (P1), to work around the constraints (P2), and to explore and enact the cards (P2, P3, P4, P5). Nevertheless, participants reported that they all got a sense of and understood very well the purpose of the cards after participating in the session.

## 5 Discussion

As expected, toolboxes are content-heavy and not quite accessible for beginners. Previous studies have shown that this stagnates the bodystorming process and negatively impacts the adoption and flow of design with these method cards. Our results indicate that the proposed approach with card game inspired mechanics mitigates this issue by initially exposing players to a limited, randomly selected number of cards. This method offers a smoother learning curve, allowing players to familiarize themselves with the card structure and content gradually. By the second turn, players approach new cards more efficiently, and by the third turn, they naturally engage with the cards and each other, explaining and enacting the contents, and suggesting combinations. The whole session took place in a remarkably playful mood, as participants felt like players playing a boardgame rather than thoroughly designing an activity.

The turn-based structure maintained steady session progress, and the simultaneous card selection mechanic challenged players to make relevant choices quickly. Although it was difficult when multiple interesting cards were available, the round mechanism allowed some discarded cards to reappear, giving players a second chance to incorporate them. Setting a time limit for card selection is recommended to keep the session flowing smoothly, though it may result in some players not fully reading their cards.

The study confirms an issue observed in past experiments: strict rules can cause misunderstandings, leading designers to focus on the rules rather than the game’s goals or ignore problematic parts. In this study, misunderstandings about modifier cards led to their minimal use. Despite this, the card-game method integrated well with the toolbox, suggesting its applicability to other card-based toolboxes.

movement cards (green), by reading, comparing, and discussing them. This creative engagement demonstrated that once participants overcame initial hurdles, they were able to interact with the cards effectively.

The process of composing sessions by sequencing cards initially caused some confusion, especially concerning how to connect mood setting (red) and movement (green) cards. However, the process felt smooth overall, with the one-card-at-a-time approach facilitating discussion and session composition. As participants progressed, they began to understand and enact the sessions, sometimes using props to aid their sequences. This gradual understanding helped teams find common themes and hand over to each other seamlessly, though initial friction was evident.

Incorporating modifiers into sessions proved challenging for participants. The concept of “burning” your card to obtain modifiers was unclear, leading to minimal use of modifiers until late in the session. While some teams began experimenting with modifiers in later rounds, others did not use them at all. There was a noticeable curiosity and willingness to experiment with modifiers, suggesting potential for greater engagement if the process were clearer.

The card-related mechanics initially introduced complexity and confusion, particularly regarding modifiers. Despite these initial challenges, the mechanics eventually provided a structure that facilitated the construction of customized sessions after a few rounds. Participants occasionally lost connection to the theme due to their focus on the rules. Timed sessions for card selection might improve flow, as the varying speeds of participants created friction. As understanding improved, gameplay became more fluid, though moments of complexity persisted.

Participants were initially hesitant to stand and enact their proposed ideas based on the cards. Most teams remained seated, discussing movements without enacting them. However, as the session progressed, enactments became more frequent, with some teams spontaneously acting out their cards. This shift indicated growing comfort and engagement with the activity, suggesting that initial encouragement might help prompt earlier enactment.

Finally, help was punctually requested at the beginning of the session regarding clarifications on card selection, the relationship between cards, and the use of modifiers. The session facilitator provided support to remind and clarify these rules. Participants were initially unsure whether to focus on the theme, the flow of the card session, or the connections between cards. Frequent clarifications helped participants navigate these uncertainties, though initial confusion highlighted areas for potential improvement in instructions.

## 4.2 Questionnaires

Participants were asked to fill a questionnaire after each design flow was presented and experienced to gather their opinions on the sessions both as designers and participants, and at the end of the design session to gather the participants thoughts on the design process and the card-game based facilitation method.

dealing all cards in the mood-setter deck, so that each team receives a starting hand of six cards. The round lasts six turns.

In every turn, teams simultaneously and secretly choose the card that they want to play. When all teams have chosen a card, they simultaneously reveal it, play it, and pass the remaining cards in their hand to the team to their right. By playing a selected card, a team places it on the table in front of them. Teams get a new hand of cards from their left, which they use to play the following turn in the same way. This repeats until all cards in the mood-setting deck have been played. *Note:* When players are not familiar with the used card toolbox, it's recommended that each team briefly explains the card they have chosen.

Round 2 follows the same rules but using the method card deck. Unused cards are this time handed over to the team to the left, and the round lasts for 8 turns. *Note:* Teams compose their own design flow by, turn after turn, arranging the played cards in sequences of mood-setters and methods. Cards can be freely rearranged during the gameplay, but teams can't trade cards among them.

When playing a selected card, a team may choose not to use it for any design flow, but instead they "burn" it (discard), which grants them the possibility to pick the complete deck of one modifier category from the center of the table, browse through it and keep as many modifier cards as they would like to incorporate to their flows. Modifiers not used are returned to the center of the table.

**Endgame.** When both rounds are over, each player team chooses one of their designed flows at their own preference and demo it by acting as facilitators themselves, while having the remaining 6 players as participants. Anonymous questionnaires are answered by all participants directly after experiencing each design flow.

## 4 Results

Four different design flows, one per team, were chosen by the players to be represented. We have carried out content analyses on the observer notes and the questionnaire responses. The following sections present the results from these, followed by a discussion where we highlight recurrent concepts and relate them to specific parts of the design session.

We place the focus of the evaluation on the proposed card-game based facilitation method rather than in the delivered design flows. For the purposes of this particular paper, the design flows are merely a design task for the case study participants. Nevertheless, the content of such flows is very relevant for the next steps of this project towards building a workshop for festival, which is developed further in Sect. 6.

### 4.1 Observation Notes

Participants' fluency in using the cards flew from a slow start, in which longer times were needed to understand their purpose, to a more fluent mid- and late-game, when participants engaged in creative use of the cards, particularly the

One facilitator, namely the session lead, informed the participants about the purpose and procedures of the session, as well as about the possibility of ending their participation at any given moment. He then briefly introduced the participants to the festival and the envisioned workshop that would ultimately be designed for it. This workshop has a fixed topic defined by the authors, with the aims of ensuring that it fits to the topic of the festival, “care”, while at the same time leaving enough room not to constrain the participants’ creativity during the design process. This topic is “Designing the invisible”, as designing for the actions and choices that people inadvertently make, and that have a real impact on the planet and a sustainable future.

The session lead introduced the Mecamind toolbox, which only two of the participants were had previous experience with and explained the rules of the proposed game-inspired approach for using the toolbox. This introduction took 30 min, leaving 90 min for the design session. The six participants plus the other two facilitators teamed up in four pairs. Due to the nature of the proposed game-based approach, the design session was framed similarly to a boardgame play session, and it’s played with all player teams sitting around a table. For consistency, the session procedures are described in the following subsections mimicking a game rulebook. Hereinafter, participants are referred to as players or participants indistinctly.

### 3.2 Session Rulebook

**Game Setup (for 4 Teams).** Create a deck of mood-setters by randomly sampling 24 cards from the box. Shuffle and place faced down on the table. Similarly, create a randomly sampled deck of 32 method cards. Shuffle and place next to the mood-setting deck. Place all modifier cards from the box, grouped by category, in separate decks at the center of the table, faced down. Finally, give pen and paper to all teams.

*Note:* Steps 1 and 2 can alternatively use a fixed selection of cards instead of random sampling, or a mixed approach with random sampling from a previously filtered deck. Similarly, some modifier categories can be filtered out. This enables facilitators to arrange custom game sessions in which only specific subsets of cards are used. The described session followed the steps are per above.

**Goal of the Game.** In teams, participants must create as many custom Mecamind design flows as desired, composed of as many mood-setters, methods, and modifiers they prefer. The created flows must follow the structure of the existing design flow cards in the Mecamind box (see Instruction Cards in Sect. 2.1), as well as they need to propose a narrative that targets the envisioned workshop.

**How to Play.** The game is played in two rounds: a first one using the deck of mood-setters and a second one using the deck of methods. Round 1 starts by

### 2.3 Interventions for a Sustainable Future

Despite widespread awareness of the need for individual actions to combat climate change, we continue to live in ways that are unsustainable. Many individuals consider themselves committed to sustainability but struggle to align their actions with their principles [32].

Theories such as the attitude-behavior gap [12] and cognitive dissonance [10] explain the gap between people’s sustainable intentions and actions. Cost and comfort often influence decisions, like choosing a cheaper, faster flight over a train [3]. This discrepancy creates cognitive dissonance, leading people to justify their choices with rationalizations like, “I didn’t have a choice due to high prices” or “I usually take the train, so flying this once is okay.”

Sustainable development requires long-term thinking, both individually and societally [19]. Neuroscience shows that decisions for our future selves activate brain areas similar to those used when considering others, suggesting that aligning current actions with sustainability feels like caring for a stranger [8]. Therefore, a promising path forward is to look towards interventions in the decision-making process that sensitize individuals to sustainability and spark a cognitive awareness of how their actions today impact their, or others’, future selves.

## 3 Playful Design Session

Section 2 motivates the project that frames this paper: the design of an open workshop, to be run at a local design festival, with the aims of sensitizing individuals and raising awareness and discussion about caring for the planet, by means of taking part in embodied and social activities.

To these aims, we have conducted a playful design session whose main goal is to design and create embodied social experiences for such workshop. This section describes this design session, that puts into practice and evaluates the contribution of this paper: a card-game inspired facilitation method to easily make use of card-based design toolboxes, articulated in this case by the Mecamind toolbox. The choice of this toolbox comes motivated by the fact that it can be used off-the-shelf to conduct brain- and body-storming sessions for designing embodied activities, such as the ones envisioned in the framing project.

### 3.1 Session Procedures

A group of six experts were invited to a 2-h design session. This group of participants was composed of 4 lecturers in embodied interaction, interaction design, and information technologies, and two master’s students with backgrounds in interaction and game design. Additionally, three researchers facilitated the session and a fourth one acted as an observer, taking notes throughout the entire session. Qualitative data used for evaluating the proposed facilitation method is gathered by means of the observation notes, as well as through anonymous responses to questionnaires that the participants answered immediately afterwards an experience had been designed and enacted.



**Fig. 2.** Sample modifier cards representing several of the modifier categories, via project’s website (<https://mecamind.eu/project.php>).

within the ideation phases [33]) or later on as a way to explore and evaluate the interactions. Designing movement-based games using these methods has seen much attention at CHI PLAY and within the broader HCI/games community [21], for example using embodied sketching [24] as a way to design social fitness games [22], bodystorming as a way to playtest and explore the design space of games [4], or embodied explorations to design a mixed-reality dance game [34]. Live-action roleplay, or larp, has even become a valued ideation method at CHI as a way of sensitizing designers to scenarios, or as a test-bed for prototyping and iteration [20].

More recently, feminist lenses have been applied to the embodied design space, for example challenging the norms and assumptions in embodied design [26], and moving toward designing for the pluralities of bodies. While the core of these embodied methods rely on the first-person perspective, they “...abandon the notion of universalism and instead build on the plurality of bodies; emphasizing that their corporeal differences are not sought to be generalized or ignored, but rather brought forward and generative in their uniqueness.” [14].

However, there are numerous challenges with articulating and teaching these embodied methods [28], underscoring the need for more tools that help us translate this knowledge to the wider community [9].

cards developed to simplify the design process of activities and experiences that include movement, embodiment, and technology [1, 5, 29]. Each method card describes a particular goal, procedure, context of use, and potential variations. The toolbox can be applied to various application domains where activity design is key, such as sports performance enhancement, health treatment, workplace physical activity, education, creative technology development, and exergames [6, 11]. The toolbox itself is grounded on a four-category model (4M) that makes the method cards actionable in the field of interaction design [7], which is in turn based on the theory of embodied cognition [27].

The method cards in the toolbox come in five categories (Fig. 1). Mood Setters (red) are activities to set the creative mood and mindset with icebreakers, warm-ups, and team-building activities. Movement cards (green) describe methods for sensitizing designers, ideating, evaluating/polishing, or documenting solutions. Movement Concepts (blue) summarize the theoretical grounds that underpin and further inform users about the methods in the other cards.



**Fig. 1.** Mecamind toolbox with sample mood setting, movement, movement concept, and instruction cards, via project's website (<https://mecamind.eu/documents.php>).

Modifiers (smaller cards, multicoloured) are quick suggestions to easily tweak and produce variations of activities described in either mood setters or movement cards, arranged in seventeen different modifier categories (Fig. 2). Instruction Cards (yellow) are like a rulebook turned into cards, containing detailed instructions on how to use mood setters, movements, and modifiers, as well as design flows: sample pre-made scenarios to quickly put the toolbox into practice.

## 2.2 Embodied Design in Games

In HCI, movement-based design methods [15] and embodied ideation methods are often used throughout the design process, and can be more generative (e.g.,

complex task of driving a participatory design process through specific methods and towards a particular end, with enough flexibility and adaptability to meet the evolving demands of the participants [23].

Movement-based design methods [13], like *bodystorming* or *soma design*, widely adopted across various domains for their promptness at giving early first-hand insights into embodied experiences, as well as their flexibility to be applied at different stages of a design project, face similar challenges in terms of the many factors that facilitators must deal with, both when designing the activities and when adapting them on the fly. Improvisation is a relevant facilitator skill, but the facilitation process can't solely rely on this [31].

This has motivated the popularity rise of the design and creation of facilitator materials, similar in concept to traditional teacher classroom materials, but in shapes that better fit the particularities of participatory and movement-based design sessions, i.e. toolboxes or toolkits [25], which in the majority of the cases consist of cards and templates sometimes complemented by a digital or web companion [17]. Such toolboxes present many advantages related to the physical form and properties of the cards that offer a tangibility play that improves cognitive processes, as well as a low-cost, re-configurable, and easy-to-use visual representation of a problem space. Nevertheless, design toolboxes also present facilitation challenges that usually relate to an overwhelming amount of content and materials that can't be presented at once or expected to be learned and handled on the spot.

We argue that game mechanics can provide structure, guidelines, and playfulness that can support and improve the facilitation process. In particular, card-based toolboxes can directly make use of card game mechanics with little and straightforward adaptation. We propose a case study in designing embodied experiences for an open workshop to be held as part of the program of a local design festival, using the Mecamind toolbox [1] as the material for design facilitation. Inspired by the work of Bentz et al. [2], we leverage the creative potential of embodied practices for sustainability transformations, to approach the topic of the festival, *care*, under the tagline of designing for the invisible actions that have a real impact on the planet.

This paper presents a playful design session facilitated with the Mecamind toolbox, in which experts in embodied interaction and design are tasked to design embodied social activities for the envisioned workshop. We also present and discuss the results from such session from observation notes and participant questionnaires, as well as describe the next steps towards the ultimate design of the workshop.

## 2 Related Work

### 2.1 Mecamind (Method Cards for Movement-Based Design)

The MeCaMInD toolbox is based on a set of principles drawing from embodied cognition [30] and soma-based design [16], which aims to position the body as an intrinsic part of the design process [18]. The toolbox (Fig. 1) is a set of method



# A Case Study in Playful Facilitation for Designing Embodied Experiences

Jose Font<sup>1</sup>(✉), Mikaela O'Bryan<sup>1</sup>, Alberto Alvarez<sup>1</sup>, and Raquel Robinson<sup>1,2</sup>

<sup>1</sup> Game Lab, Department of Computer Science and Media Technology,  
Malmö University, Malmö, Sweden

{jose.font,mikaela.obryan,alberto.alvarez}@mau.se

<sup>2</sup> IT University of Copenhagen, Copenhagen, Denmark  
raqr@itu.dk

**Abstract.** Embodied design is fostering relevant advancements in novel design methods for learning and innovation. Nevertheless, practice and research show that teaching and facilitating embodied design is a challenging task. This paper presents a playful approach for conducting embodied design sessions, which adapts existing card-game mechanics to improve the accessibility and usability of existing embodied design toolboxes during a design session. Our approach hypothesizes that card-game mechanics can address usual pitfalls common to card-based facilitation toolkits by adding structure, guidelines, and playfulness. We also envision that such toolboxes, that have become popular in the research in participatory and co-design contexts, can seamlessly incorporate card-game mechanics like the ones we propose. We describe the setting of the design session, as well as present and discuss the results obtained from observation and participant questionnaires. Finally, we also outline the next steps for designing an embodied workshop in sustainability based on the results from this study. We strive to analyze our proposed method by its capacity to blend into the structure and materials of a card-based toolbox, eliciting the playfulness embedded in the leveraged card-game mechanics to implicitly provide support and guidance through the session, to both facilitators and participants, without disrupting the flexible and open flow of an movement-based design session.

**Keywords:** Movement-based design · game-based design · design facilitation · method cards

## 1 Introduction

Addressing today's complex challenges necessitates interdisciplinary collaborations and crossing knowledge boundaries. Design thinking and participatory design have gained prominence as a user-centered approach to innovation, so that collaborative design practices can integrate diverse perspectives. The roles of trained designers have thus diversified, highlighting the relevance of the role of the facilitator in participatory and co-design contexts. The facilitator has the

**Engaging Learners Through  
Gamification, Playful Design,  
and Generative AI**

20. Memon, J., Sami, M., Khan, R.A., Uddin, M.: Handwritten optical character recognition (OCR): a comprehensive systematic literature review (SLR). *IEEE Access* **8**, 142642–142668 (2020)
21. Morschheuser, B., Hamari, J., Koivisto, J., Maedche, A.: Gamified crowdsourcing: conceptualization, literature review, and future agenda. *Int. J. Hum Comput Stud.* **106**, 26–43 (2017)
22. Morschheuser, B., Hamari, J., Maedche, A.: Cooperation or competition - when do people contribute more? a field experiment on gamification of crowdsourcing. *Int. J. Hum. Comput. Stud.* **127**, 7–24 (2019)
23. Reiser, L., Harper, L., Freeling, M., Han, B., Luan, S.: Fair: a call to make published data more findable, accessible, interoperable, and reusable. *Mol. Plant* **11**(9), 1105–1108 (2018)
24. Rodríguez, I., Puig, A., Tellols, D., Samsó, K.: Evaluating the effect of gamification on the deployment of digital cultural probes for children. *Int. J. Hum. Comput. Stud.* **137**, 102395 (2020)
25. Thomas, N.J., Baral, R., Crocco, O.S., Mohanan, S.: A framework for gamification in the metaverse era: how designers envision gameful experience. *Technol. Forecast. Soc. Chang.* **193**, 122544 (2023)
26. Vargas-Murillo, A.R., Pari-Bedoya, I.N.M.d.L.A., Guevara-Soto, F.D.J.: Virtual gamification strategies and their impact on legal education experiences: a systematic review. In: *Proceedings of the 2023 8th International Conference on Distance Education and Learning*, pp. 85–90 (2023)
27. Vercelli, G.V., Iacono, S., Vallarino, M., Zolezzi, D.: Puzzle battle 2.0: a revisited serious game in VR during pandemic's period. In: de Rosa, F., Marfisi Schottman, I., Baalsrud Hauge, J., Bellotti, F., Dondio, P., Romero, M. (eds.) *GALA 2021. LNCS*, vol. 13134, pp. 252–257. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-92182-8\\_25](https://doi.org/10.1007/978-3-030-92182-8_25)
28. Wang, K.D., Chen, Z., Wieman, C.: Can crowdsourcing platforms be useful for educational research? In: *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pp. 416–425 (2024)
29. Woodward, E., Woodward, M.: Expected value and the wheel of fortune game. *Math. Teach.* **87**(1), 13–17 (1994)

creating an engaging narrative that enhances motivation and makes users feel part of an exciting transcription journey.

## References

1. Bandura, A.: Perceived self-efficacy in cognitive development and functioning. *Educ. Psychol.* **28**(2), 117–148 (1993)
2. Bartle, R.: Hearts, clubs, diamonds, spades: players who suit muds. *J. MUD Res.* **1**(1) (1996)
3. Bonney, R., Phillips, T.B., Ballard, H.L., Enck, J.W.: Can citizen science enhance public understanding of science? *Public Underst. Sci.* **25**(1), 2–16 (2016)
4. Cohen, P.R., Levesque, H.J.: Teamwork. *Nous* **25**(4), 487–512 (1991)
5. Coppola, F.: Il rapporto aureo, le sue potenze e i numeri di fibonacci. *Scientia* doi **10** (2014)
6. Davies, R.: Crowdsourcing in cultural heritage. *Europeana-Digital Heritage Research Lab* (2020)
7. Estellés-Arolas, E., González-Ladrón-de Guevara, F.: Towards an integrated crowdsourcing definition. *J. Inf. Sci.* **38**(2), 189–200 (2012)
8. Featherstone, M., Habgood, J.: Unicraft: exploring the impact of asynchronous multiplayer game elements in gamification. *Int. J. Hum Comput Stud.* **127**, 150–168 (2019)
9. Ferriter, M.: Growing to a community of volunpeers: communication & discovery. *Smithsonian Inst. Arch.* **8**, 191 (2014)
10. Forsyth, E.: Learning through play: games and crowdsourcing for adult education. *Aust. Public Libr. Inf. Serv.* **25**(4), 166 (2012)
11. Gunther, A., Schall, M., Wang, C.H.: The creation and evolution of the transcription center, smithsonian institution’s digital volunteer platform. *Collections* **12**(2), 87–96 (2016)
12. Hamari, J.: Do badges increase user activity? a field experiment on the effects of gamification. *Comput. Hum. Behav.* **71**, 469–478 (2017)
13. Hunicke, R., LeBlanc, M., Zubek, R., et al.: MDA: a formal approach to game design and game research. In: *Proceedings of the AAAI Workshop on Challenges in Game AI*, San Jose, CA, vol. 4, p. 1722 (2004)
14. Ignat, T., Ayrís, P.: Built to last! embedding open science principles and practice into European universities. *Insights UKSG J.* **33**(1) (2021)
15. Landers, R.N., Bauer, K.N., Callan, R.C.: Gamification of task performance with leaderboards: a goal setting experiment. *Comput. Hum. Behav.* **71**, 508–515 (2017)
16. Lee, Y., Kim, M., Koo, J.: The impact of social interaction and team member exchange on sport event volunteer management. *Sport Manag. Rev.* **19**(5), 550–562 (2016)
17. Legaki, N.Z., Xi, N., Hamari, J., Karpouzis, K., Assimakopoulos, V.: The effect of challenge-based gamification on learning: an experiment in the context of statistics education. *Int. J. Hum. Comput. Stud.* **144**, 102496 (2020)
18. Leon, S.M.: Build, analyse, and generalise: community transcription of the papers of the war department and the development of scripto. In: *Crowdsourcing our Cultural Heritage*, pp. 88–112 (2014)
19. Lin, Y.L., Ding, N.D.: Competitive gamification in crowdsourcing-based contextual-aware recommender systems. *Int. J. Hum.-Comput. Stud.* 103083 (2023)



**Fig. 3.** Christopher, the buddy

it fosters skill development in cultural heritage, transcription techniques, and historical knowledge, maximizing learning outcomes through engagement and adaptability.

## 4 Conclusions and Future Works

This paper analysed portals that use gamification mechanics to engage users in transcription tasks. An engaging system can make transcription more captivating and enjoyable for both enthusiasts and experts. This approach aligns with FAIR principles, Citizen Science, and Open Science, involving users in data creation through clear, detailed instructions to ensure structured activities and community harmony.

The platform will launch with version 1.0, which will be monitored over time to identify and improve aspects valued by users. Gamification has proven effective in enhancing motivation and performance across various fields [24, 26], but it is crucial to determine whether the chosen game mechanics result in superficial content acquisition or effective material assimilation [8]. The project's goal is to engage users without offering tangible rewards, only rankings and badges. However, if necessary, the system will be adapted to meet the individual needs and motivations of players.

It's essential to recognize the challenges of maintaining long-term volunteer engagement. Although gamified systems attract diverse users, further adjustments may be needed to address specific user motivations. Sustaining digital volunteer participation requires managing diverse motivations, avoiding task fatigue, and continuously refining incentives. A thorough evaluation plan should be developed to assess user satisfaction and platform performance, ensuring that the system evolves with user feedback and meets volunteers' changing needs.

Potential rewards for high achievers could include University of Genoa merchandise, books, reproductions of cultural heritage items, museum discounts, or exclusive archive access, helping volunteers appreciate the value of cultural preservation. Gamification fosters positive community practices through engaging experiences [25], and future projects could benefit from storytelling elements,

primary form of rewards, follow a Fibonacci sequence. Each level requires progressively more points, starting from 100 points to reach level 1 and doubling as levels advance, promoting long-term engagement [5]. Points can be earned through three actions:

1. Document transcription: Each character, including spaces, earns 2 points, credited upon saving and subject to verification.
2. Revision: Each reviewed character earns 0.5 points upon validation which are earned upon validation of the review.
3. Daily logins: Points are awarded for five consecutive days of logging in, starting from 5 points on the first day to 100 points on the fifth day.

The 2x Boost condition doubles experience points in two scenarios: completing a document transcription or typing over two thousand characters in a single session. Documents are categorised by completion status using a color-coded system (Fig. 2):

1. Not Started: Grey icon.
2. In Progress: Yellow icon.
3. Pending Review: Blue icon, indicating the need for another volunteer to review.
4. Reviewed: Green icon, indicating completed review.



**Fig. 2.** Documents' status

Social interaction is crucial for engagement [16]. The platform allows users to connect, send friend requests, exchange messages, comment on document pages, and form teams to compete on leaderboards. Teams enhance motivation through involvement and collaboration [4]. Leaderboards display individual and team achievements, fostering competitiveness and higher performance [15]. A virtual companion, Christopher, guides users through the transcription process, providing updates and assistance. Christopher's demeanor changes based on the player's actions, fostering an emotional connection and motivating users to improve [27](Figure 3).

The platform integrates collaborative learning, enabling users to work both individually and together on transcriptions. With features like team formation, leaderboards, and document review, users can share feedback and achieve common goals, boosting engagement through both personal achievements and group success. Reviewing others' work enhances critical thinking, attention to detail, and peer learning, enriching the educational experience. The platform adapts to users' needs by offering personalized paths and adjustable difficulty levels, allowing users to progress at their own pace. By supporting diverse learning styles,

the features of the platform. It is centrally structured and divided into three main sections:

1. The two transcription modes and the community button.
2. Suggestions.
3. Documents to which a contribution has already been made that are still pending completion.

Additionally, it features a navigation bar that provides access to the search bar, instructions, archive, rankings, messages, notifications, and the user’s profile.

### 3 Proposed Gamification Strategies

Upon logging into the platform, users can select a document to transcribe using three modes in “Classic Mode”. The first mode allows users to search documents via a homepage search bar, displaying results based on keywords. The second mode provides access through a central button leading to a catalog page with a search bar and filters for category, difficulty, and document condition. The third mode is dedicated to recommended documents. For Explorers, a “Random Mode” allows users to transcribe a randomly selected document by choosing a difficulty level (Easy, Medium, or Hard) and engaging in a “Wheel of Fortune” game [29] to determine the document (Fig. 1).



**Fig. 1.** Random Mode

Achievers can gain satisfaction from collecting rewards such as badges, experience points, and items, while Killers strive for higher rankings. Badges, displayed on user profiles, are earned by completing specific tasks or reaching certain levels. This system enhances autonomy and satisfaction [1], with all badges serving as additional incentives for motivated players [12]. Challenges and a user-centered design further improve user performance [17]. Experience points, the

advanced 2D and 3D scanners for high-definition presentation. This effort aims to prevent the loss or dispersion of these documents, making the University’s cultural heritage globally discoverable and accessible with ease. The project relies on “digital volunteers” for transcribing scanned documents and digitised objects. These volunteers, ranging from professors and students to enthusiasts, dedicate their time to deciphering the archive’s contents. Instead of traditional incentives, engagement is fostered through gamification, immersing volunteers in an environment with various gameplay modes and interaction opportunities. This approach, based on gamification principles, enhances motivation, engagement, and retention.

The gamified platform’s design follows the MDA framework, focusing on Mechanics, Dynamics, and Aesthetics to create a compelling user experience [13]. By analysing different player types—Achievers, Explorers, Socializers, and Killers—the platform tailors its mechanics to meet diverse motivational needs, ensuring sustained volunteer participation [2]. Ultimately, the goal is to continuously enrich UniGe’s cultural heritage through active volunteer contributions, preserving and expanding its digital archive incrementally. The target audience for the SMA-UniGe project is diverse, including curious individuals, experts in various fields, and secondary school students involved in transversal skills and orientation paths (PCTO - “Percorsi per le Competenze Trasversali e l’Orientamento”, in italian). To address this broad audience, three user profiles have been developed:

1. Curious User: These users can browse content, transcribe, review, earn experience points and badges, and manage their dashboard independently. They select favorite topics during registration or based on search preferences.
2. PCTO User: This profile is for students enrolled in PCTO programs and is linked to their school email addresses. It offers a tailored experience for completing PCTO tasks, selected by the tutor professor and displayed in the student’s dashboard. Progress is tracked based on time spent and levels achieved within the platform’s game system. After completing the PCTO period, students can switch to a “Curious” user profile via their account settings.
3. Expert User: Once a user reaches level 20, their account is upgraded to “Expert” status. Experts are trusted community members who can review transcripts and mark them as “complete” independently, without moderator intervention.

To immerse digital volunteers in a hi-tech environment while maintaining the visual identity of the University of Genoa, a specific color palette has been selected. This palette includes two primary colors from the university’s corporate identity: “Blu Unige” (HEX: 002677) and light blue (HEX: 199BFC). Complementing these are gray for written texts (HEX: 333333), a contrasting blue (HEX: 005DBF), and neon green (HEX: 48E55A) for links and hover state activation for clickable elements on the screen. The homepage, after logging in, is designed to enable users to access all relevant pages for transcription and enjoy

In addition to digitisation, transcription further enhances the discoverability and accessibility of these resources for educational, research, and sharing purposes, in line with FAIR principles [23] and Open Science standards [14]. Given the limitations of OCR and HTR tools for intricate handwriting [20], a new transcription system was developed, relying on human volunteers and incorporating Citizen Science [3], crowdsourcing, and public engagement strategies [6, 7]. This initiative supports the “third mission” of disseminating knowledge beyond academia to benefit societal development.

To keep volunteers engaged in crowdsourcing activities, gamification techniques are employed, motivating users to perform better and engage in various tasks [19, 21]. Gamification makes the transcription process more enjoyable, encouraging participation through competitive and collaborative elements.

The educational benefits of this gamified transcription activity are extensive. Volunteers gain a deeper understanding and appreciation of history and cultural heritage, acquire valuable experience in archival research and document analysis, and develop critical academic and professional skills. Transcribing documents, especially those in historical or specialised languages, significantly improves language proficiency and comprehension. This process promotes critical thinking and attention to detail, fostering lifelong learning through continuous intellectual and practical skill development [10]. Additionally, volunteers become proficient with digital tools and platforms used in transcription and archiving, enhancing their technological literacy. Participating in a crowdsourced environment cultivates teamwork and collaborative problem-solving skills, while the project encourages peer-to-peer interaction among students, enriching their collaborative learning experience [28].

## 2 The University Museum System at UniGe (SMA-UniGe)

The SMA-UniGe project was initiated to consolidate the extensive cultural heritage of the University of Genoa into a unified, online repository accessible via a user-friendly interface and engagement system. Over the past two decades, numerous document transcription projects have emerged [18]. The University of Genoa drew inspiration from two notable platforms to develop its own comprehensive transcription system. The first model is the Smithsonian Institution’s “Digital Volunteers: Transcription Center”, known for its collaborative space where “digital volunteers” engage in transcription efforts [9] and its digitisation and volunteer-driven transcription process [11]. The second model is Europeana’s “Transcribathon”, which incorporates gamification dynamics to encourage user participation in a collaborative and competitive setting, likening the transcription process to a marathon [22]. By integrating elements from both platforms, the University of Genoa aims to create an engaging and effective environment for digitising and transcribing its cultural heritage.

The primary objective is to digitise a wide array of materials, including ancient tomes, manuscripts, postcards, labels, maps, and exam papers, using



# Unlocking Cultural Heritage: The Gamified Digitisation Project of SMA-UniGe

Lara La Tessa<sup>1</sup>, Mauro Coccoli<sup>2</sup>, Stefano Schiaparelli<sup>3</sup>,  
and Daniele Zolezzi<sup>4</sup>

<sup>1</sup> DIEC - University of Genoa, Via F. Vivaldi, 5, 16126 Genoa, GE, Italy  
lara.latessa@edu.unige.it

<sup>2</sup> DIBRIS - University of Genoa, Via alla Opera Pia, 13, 16145 Genoa, GE, Italy

<sup>3</sup> DISTAV - University of Genoa, Corso Europa, 26, 16132 Genoa, GE, Italy

<sup>4</sup> DLCM - University of Genoa, Piazza di Santa Sabina, 2, 16124 Genoa, GE, Italy

**Abstract.** The University of Genoa, Italy, addresses the challenge of inaccessible cultural heritage stored in university archives worldwide by implementing a comprehensive digitisation initiative. This project focuses on collecting, storing, and digitising a wide array of items, including books, manuscripts, archival materials, and documents related to museum artifacts. To ensure accessibility for both humans and machines, the initiative involves providing alternate descriptions, metadata, and speech-to-text transcriptions for images and videos, as well as word-for-word transcripts for ancient texts where OCR is ineffective. We present the design of a transcription system for the University Museum System (SMA-UniGe) at the University of Genoa, which includes user interface elements and engagement techniques. By leveraging gamification theory, the system transforms the typically monotonous task of transcription into an engaging experience, encouraging participation from digital volunteers. This initiative aligns with the University's third mission of public engagement, aiming to disseminate knowledge beyond the academic environment and contribute to social, cultural, and economic development.

**Keywords:** Cultural Heritage · Gamification · User Experience

## 1 Introduction

After conducting a thorough census of the cultural assets at the University of Genoa (UniGe) in July 2021, it was decided that these valuable cultural resources should be made fully accessible. Consequently, the University Museum System (SMA, “Sistema Museale di Ateneo”, in italian) was established, offering a unified online platform where the digitised and archived cultural heritage can be accessed by everyone as an interactive exhibition. This system encompasses multiple museums, botanical gardens, biobanks, archives, and a variety of collections.

38. Spors, V., Laato, S., Buruk, O., Hamari, J.: Longing to be the mountain: a scoping review about nature-centric, health-minded technologies. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–16 (2023)
39. Ulrich, R.S., Simons, R.F., Losito, B.D., Fiorito, E., Miles, M.A., Zelson, M.: Stress recovery during exposure to natural and urban environments. *J. Environ. Psychol.* **11**(3), 201–230 (1991)
40. Wang, X., Shi, Y., Zhang, B., Chiang, Y.: The influence of forest resting environments on stress using virtual reality. *Int. J. Environ. Res. Public Health* **16**(18), 3263 (2019)
41. Wang, X., Lo, K.: Just transition: a conceptual review. *Energy Res. Social Sci.* **82**, 102291 (2021)
42. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the panas scales. *J. Pers. Soc. Psychol.* **54**(6), 1063 (1988)
43. Webber, S., Kelly, R.M., Wadley, G., Smith, W.: Engaging with nature through technology: a scoping review of hci research. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–18 (2023)
44. Wilson, E.O.: *Biophilia*. Harvard university press, New York (1984)
45. Winter, P.L., Selin, S., Cerveny, L., Bricker, K.: Outdoor recreation, nature-based tourism, and sustainability. *Sustainability* **12**(1), 81 (2019)
46. Wolsko, C., Lindberg, K., Reese, R.: Nature-based physical recreation leads to psychological well-being: evidence from five studies. *Ecopsychology* **11**(4), 222–235 (2019)

18. Kaplan, R., Kaplan, S.: *The Experience of Nature: A Psychological Perspective*. Cambridge University Press, Cambridge (1989)
19. Kaplan, S.: The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* **15**(3), 169–182 (1995)
20. Kardong-Edgren, S.S., Farra, S.L., Alinier, G., Young, H.M.: A call to unify definitions of virtual reality. *Clin. Simul. Nurs.* **31**, 28–34 (2019)
21. Kellert, S.R., Wilson, E.O.: *The biophilia hypothesis* (1993)
22. Kikuchi, R.: Adverse impacts of wind power generation on collision behaviour of birds and anti-predator behaviour of squirrels. *J. Nat. Conserv.* **16**(1), 44–55 (2008)
23. Klain, S.C., Satterfield, T., Sinner, J., Ellis, J.I., Chan, K.M.: Bird killer, industrial intruder or clean energy? perceiving risks to ecosystem services due to an offshore wind farm. *Ecol. Econ.* **143**, 111–129 (2018)
24. Li, H., et al.: Access to nature via virtual reality: a mini-review. *Front. Psychol.* **12**, 725288 (2021)
25. Mäkelä, V., et al.: Virtual field studies: conducting studies on public displays in virtual reality. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–15 (2020)
26. Martin, L., White, M.P., Hunt, A., Richardson, M., Pahl, S., Burt, J.: Nature contact, nature connectedness and associations with health, wellbeing and pro-environmental behaviours. *J. Environ. Psychol.* **68**, 101389 (2020)
27. Mattila, O., Korhonen, A., Pöyry, E., Hauru, K., Holopainen, J., Parvinen, P.: Restoration in a virtual reality forest environment. *Comput. Hum. Behav.* **107**, 106295 (2020)
28. McCrickard, D.S., Jones, M., Stelter, T.L.: *HCI Outdoors: Theory, Design, Methods and Applications*. Springer, Heidelberg (2020)
29. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **77**(12), 1321–1329 (1994)
30. Nysten-Haarala, S., Joonas, T., Hovila, I.: Wind energy projects and reindeer herders' rights in finnish lapland: a legal framework. *Elementa: Science of the Anthropocene* **9** (2021)
31. Restout, J., et al.: Fully immersive virtual reality using 360° videos to manage wellbeing in older adults: a scoping review. *J. Am. Med. Dir. Assoc.* **24**(4), 564–572 (2023). <https://doi.org/10.1016/j.jamda.2022.12.026>
32. Richardson, M., Butler, C.W.: Nature connectedness and biophilic design. *Build. Res. Inf.* **50**(1–2), 36–42 (2022)
33. Salatino, A., et al.: Virtual reality rehabilitation for unilateral spatial neglect: a systematic review of immersive, semi-immersive and non-immersive techniques. *Neurosci. Biobehav. Rev.* **152**, 105248 (2023). <https://doi.org/10.1016/j.neubiorev.2023.105248>. <https://www.sciencedirect.com/science/article/pii/S0149763423002178>
34. Santos, B., Romão, T., Dias, A.E., Centieiro, P.: eVision: a mobile game to improve environmental awareness. In: Reidsma, D., Katayose, H., Nijholt, A. (eds.) *ACE 2013. LNCS*, vol. 8253, pp. 380–391. Springer, Cham (2013). [https://doi.org/10.1007/978-3-319-03161-3\\_28](https://doi.org/10.1007/978-3-319-03161-3_28)
35. Sheffield, D., Butler, C.W., Richardson, M.: Improving nature connectedness in adults: a meta-analysis, review and agenda. *Sustainability* **14**(19), 12494 (2022)
36. Skarin, A., Sandström, P., Alam, M.: Out of sight of wind turbines-reindeer response to wind farms in operation. *Ecol. Evol.* **8**(19), 9906–9919 (2018)
37. Spangenberg, P., Freytag, S.C., Geiger, S.M.: Embodying nature in immersive virtual reality: are multisensory stimuli vital to affect nature connectedness and pro-environmental behaviour? *Comput. Educ.* **212**, 104964 (2024)

## References

1. Ahn, S.J., Bostick, J., Ogle, E., Nowak, K.L., McGillicuddy, K.T., Bailenson, J.N.: Experiencing nature: embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *J. Comput.-Mediat. Commun.* **21**(6), 399–419 (2016)
2. Burdea, G.C., Coiffet, P.: *Virtual Reality Technology*. John Wiley & Sons, Hoboken (2024)
3. Calogiuri, G., et al.: The impact of visualization techniques of immersive virtual scenarios in promoting nature connectedness: a blind randomized controlled trial with mixed-methods approach. *J. Environ. Psychol.* **90**, 102102 (2023)
4. Capaldi, C.A., Dopko, R.L., Zelenski, J.M.: The relationship between nature connectedness and happiness: a meta-analysis. *Front. Psychol.* **5**, 92737 (2014)
5. Chirico, A., et al.: Designing virtual environments for attitudes and behavioral change in plastic consumption: a comparison between concrete and numerical information. *Virt. Real.* **25**, 107–121 (2021)
6. Church, C., Crawford, A.: Minerals and the metals for the energy transition: exploring the conflict implications for mineral-rich, fragile states. In: Hafner, M., Tagliapietra, S. (eds.) *The Geopolitics of the Global Energy Transition*. LNE, vol. 73, pp. 279–304. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-39066-2\\_12](https://doi.org/10.1007/978-3-030-39066-2_12)
7. Ciproso, P., Giglioli, I.A.C., Raya, M.A., Riva, G.: The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Front. Psychol.* **9**, 2086 (2018)
8. Cosio, L.D., Buruk, O., Fernández Galeote, D., Bosman, I.D.V., Hamari, J.: Virtual and augmented reality for environmental sustainability: a systematic review. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–23 (2023)
9. Drazich, B.F., et al.: In too deep? a systematic literature review of fully-immersive virtual reality and cybersickness among older adults. *J. Am. Geriatr. Soc.* **71**(12), 3906–3915 (2023)
10. Frank, L.A.: Using the computer-driven vr environment to promote experiences of natural world immersion. In: *The Engineering Reality of Virtual Reality 2013*, vol. 8649, pp. 78–90. SPIE (2013)
11. Greengard, S.: *Virtual Reality*. MIT Press, Cambridge (2019)
12. Häkkinen, J., et al.: Reflections on the naturechi workshop series: unobtrusive user experiences with technology in nature. *Int. J. Mobile Hum. Comput. Interact. (IJMHCI)* **10**(3), 1–9 (2018)
13. Hartig, T.: Toward understanding the restorative environment as a health resource. In: *Open Space: People Space. Engaging with the Environment* (2004)
14. Hartig, T., Korpela, K., Evans, G.W., Gärling, T.: A measure of restorative quality in environments. *Scand. Hous. Plan. Res.* **14**(4), 175–194 (1997)
15. Jones, M.D., Von Feldt, M., Andrus, N.: Outside where? a survey of climates and built environments in studies of hci outdoors. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15 (2022)
16. Joonas, T., Joonas, J.: Toward just, ethical and sustainable arctic economies, environments and societies. In: *Routledge Handbook of Polar Law*. Routledge (2023)
17. Jyskä, I., et al.: Design and user experience of virne application: deep breathing exercise in a virtual natural environment to reduce treatment anxiety in pediatrics. In: *Healthcare*, vol. 11, p. 3129. MDPI (2023)

### 5.3 Limitations

We acknowledge that our study is limited by the small sample size and not balancing the presentation order of the A-B scenarios in the study. However, the main focus of our research was on exploring and evaluating the methodology of using VR for assessing nature landscapes and comparing it to paper-based presentations. Also, we noticed that repetitive questions and the use of the same pictures seemed to affect participants' perception of the virtual environment to some extent. As another limitation, the quality and realism of the AI-generated VR environments impacted participants' perceptions, including difference between the scenarios. This is however a methodological lesson learnt in preparing the visualizations. For instance, we cannot be sure whether people did not like the wind turbines or the lack of trees. Future research should focus on improving the realism and consistency of VR environments to better assess their impact.

In the future, to gain more insight into the phenomenon of perceptions of technology in nature, a comparison with a data set collected in-the-wild in the nature should be included. This is especially important, as there is generally a lack of nature and technology user experience studies that have been conducted in a wilderness context. Jones et al. report that of 101 HCI studies conducted outdoors involving a person using a computer, 82 took place in urban settings [15]. Including an in-the-wild study would also provide insight into the methodological feasibility of simulating the nature experience in VR.

## 6 Conclusion

Our research seeks to make two types of contributions. On one hand, it offers insight into user perceptions of integrating technology into nature scenes. On the other hand, it emphasizes the importance of creating research methods for assessing technologies that seek to create close-to-genuine nature experiences or simulate and study the disturbances in nature experiences in VR.

Comparing VR and 2D paper printouts revealed that VR was perceived as an immersive, interesting, and even exciting experience, which participants liked and which they perceived suitable to assess nature scenarios and technology embedded into it. However, 2D paper printouts were reported as giving more room for participants' imagination and reflections. The VR environment also led participants to pay attention to other details rather than the holistic landscape. Our findings comparing the landscape without and with windmills reveal that the landscape without technology was perceived more positively and as more restorative. We acknowledge that our findings are limited by the small study sample and the artificially generated content used.

**Acknowledgements.** This research has been supported by the project Reconceptualizing Boundaries Together Towards Resilient and Just Arctic Future(s) (REBOUND), grant number 358482 and 358498, funded by the Strategic Research Council (SRC) established within the Research Council of Finland, and ProtoXR project funded by ERDF.

## 5 Discussion

### 5.1 About the Methods VR vs 2D Printouts

The main focus of our research was to compare the methods of using VR and paper for studies concerning nature landscapes. Generally, VR versions of the scenarios were perceived as slightly more positive and less negative than the 2D version in both scales. VR was generally liked by the participants, and it was perceived as immersive and interesting. For instance, some participants who tested the VR first felt uninterested in watching the scenarios in 2D after they had already seen them in VR. Thus, we believe that the immersive and experiential power of VR can be utilized to capture people's attention to topics related to natural landscapes and technology. However, the unauthentic nature generated by AI should be carefully checked for inconsistencies.

Interestingly, the paper version was said to provide more flexibility to explore your feelings towards the setting, leaving more room for imagination. Prior research on using VR for user studies has reported that in VR, users' attention is easily captured by high-fidelity details [25]. This may be true also in our study. These findings highlight that paper illustrations and simple low-fi solutions can have their place when engaging citizens in discussions about technology and nature. We believe these methodological findings provide valuable insight when planning future studies addressing nature experiences.

### 5.2 About the Just Green Transition

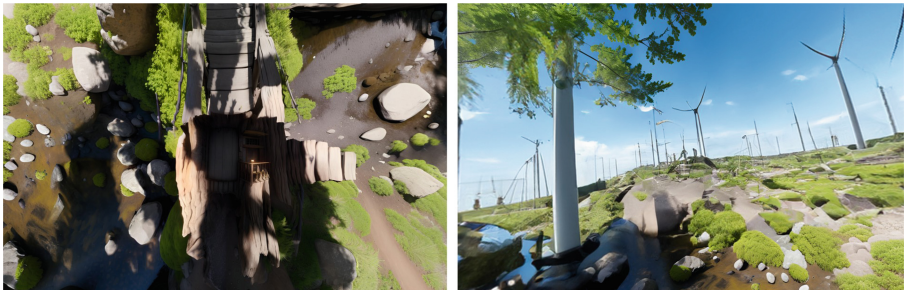
Our research contributes to the topical discussion of just transition for green technologies [41]. Our research touches on the complex interplay between technological interventions and the preservation of genuine nature experiences. Sustainable technologies and their development are critical given the climate change crisis. However, creating and using technology that promotes environmental sustainability and reduces the ecological footprint creates conflicting responses. Some of the affected people are those who live in the areas that are subject to the implementation of large energy plants or mines. These are often remote areas, such as subarctic regions with indigenous people and their traditional lifestyle [30], and high recreational nature value for tourism. The transition to sustainable technologies already creates conflicts and societal tensions [6].

Behaviors related to eco-friendly and sustainable technological solutions are promoted as a green transition. On the other hand, locating windmills in wilderness areas is criticized for ruining the wilderness, harming the natural ecosystems including traditional occupations such as reindeer herding, and generally negatively affecting the nature experience. Our explorative study results are aligned with this, as the nature landscape without windmills was perceived more positively through the Positive and negative Affect Schedule (PANAS), the qualitative responses seconding this. Our findings suggest that immersive technologies like VR can be valuable tools for environmental impact assessments and stakeholder engagement, offering a more nuanced and richer understanding of public perceptions and acceptance of renewable energy projects.

unbalanced mixture of landscape elements and vegetation not found in the region made people wonder where they were. P10 described her experience in scenario B: “I must be at the seaside because the vegetation does not look like the vegetation of the fells. There is a strong wind there”. She made reasonable deductions based on what she saw, even though it might contradict the previous mountainous and forested environment.

Several participants felt that the environment was confusing because of the inconsistencies in the AI environment, such as rivers that disappeared and appeared in strange places, or branches hanging from the air. The AI’s weird design made it obvious that they were in the VR world. However, as participants were able to place themselves in the picture as if they were in real nature, the experience was more comprehensive. Some participants explained that they could imagine elements of real nature there, such as the sound of water and real changes in height. Some participants were impressed the how realistic the picture was. If you glanced at it fast it looked like a photograph, one participant described their first impression.

The AI-generated virtual environment made some of the participants cautious or skeptical. The virtual environment was exaggeratedly AI-like when viewed with VR glasses. The inconsistencies in the images noticed by participants (Fig. 7) and the variation in image quality disturbed the experience and moved the focus from the actual nature experience more toward other elements. One participant said that its unauthentic nature made them approach it with a certain reservation. It even felt a little annoying for them and they stated that they would rather choose “real” nature. Virtuality of the nature environment could affect participants’ perception of it. For example, P8 described their experience: “The fact that the landscape is virtual takes away the strongest edge from negative feelings. At the same time, the picture reminds us that this is a reality somewhere. Beautiful nature is being destroyed and we are all collectively guilty of it because of our energy needs”.



**Fig. 7.** Strange details in VR world generated by AI: a window on the bridge (left), half-finished wind turbines, and strange structures on the horizon (right)

#### 4.4 VR Experience

On average, participants felt that VR as a testing method was interesting. It was mentioned that virtual nature was an interesting way to reflect on one's own nature experience. Comparing the same landscape in the picture and in VR brought up new perspectives. It made participants curious about the environment and feel like they were not in a small room. A positive feature of the VR world was being able to look everywhere around you which made the experience feel more authentic, e.g. "When working with VR glasses, you feel that you are somehow in control and the feeling is more authentic. Feels easier to belong to that place" (P13). One participant questioned the need for VR in this kind of test, e.g. "There wasn't anything to do in VR so I was wondering a bit if the VR aspect was necessary, but I was interested enough to look around the environment" (P7). Several participants mentioned that it would have been interesting to be able to move around in the VR environment, exploring the surroundings and interacting with it. One participant suggested that lying down and looking at the scenery could have been one option as well. P11 explained that her experience in VR helped her to get a general idea of the place, but at the same time she felt that she was mainly a spectator and did not feel like she could participate in the scenery.

Compared to other possible alternatives, VR was said to be the best method for this kind of use. Participants mentioned that VR helped them to imagine themselves in the setting. The 3D world and the relative scale of the objects around them were easier to understand in VR.

One factor that should be considered is the soundscape. During the experiment window was open and bird song could be heard from the outside. Some participants mentioned that fitting to the experience. During the testing, it started raining outside and the window was closed. Instead of the bird sounds, one participant heard the mechanical sound of ventilation and it seemed to affect her experience, e.g. "I feel like I hear a technical sound coming from the wind turbines. I wonder how low the wind turbines are" (P19).

Since several participants had used VR a lot, they were critical of the quality of the VR experience, e.g. "With VR, the user gets a good spatial experience, as long as the environment looks believable and not distorted" (P5). Poor quality, a fish-eye effect, and strange AI-generated elements disrupted the overall experience. Some participants also mentioned feeling dizzy or uncomfortable because of the distortion, e.g. "There were a lot of flaws in the place that broke the experience. The landscapes were disconnected from each other and were connected illogically" (P14). Seeing your body instead of hovering in the air was mentioned as one method to improve the experience.

#### 4.5 AI-Generated Environment

Many participants were like-minded about scenario A being fitting for outdoor activities such as hiking or camping. However, it did not resemble the local landscape, making it more difficult to feel a deep connection to the place. The

scenario A. Especially when there was another picture as a point of comparison, the difference was clear and the negative feelings were quite strong, e.g. “Wind turbines make you irritated and the landscape somehow remains empty” (P13). One participant described nature as a place where there are no worries and you need to have respect towards it.

### 4.3 Immersive VR Versus 2D Printouts

After the study session, each participant was asked which testing method they liked better. Everyone answered they preferred VR as a testing method over the 2D image. Still, a few of the respondents answered that the image printed on paper helped them to imagine themselves to the scene better. With the paper version, it was perceived that there was more flexibility to explore your own reflections and feelings towards the setting, and it left more room for imagination, whereas the VR could be overwhelming with its all confusing elements. Interestingly, VR environment could also be perceived as psychologically distancing, since everything was already given and you needed to observe the scenery as a floating outsider.

Some participants who tested VR first felt uninterested in watching the scenarios again in 2D after they had already seen it in VR. They stated that looking at the image they were not as observant as in VR. It was not possible to enter the picture and the landscape in the same way, which they felt limited their experience. Even if the picture was still beautiful, it remained uninteresting compared to VR: “After seeing the scenarios in VR, I don’t really care about the image” (P7). The 2D image looked nice, but it did not make as strong impression as VR “It’s hard to get anything out of the picture alone after wearing VR glasses” (P3).

The VR experience was perceived to be more forgiving of the inconsistencies in the AI design. Respondents said they paid less attention to little details when there was more to explore. In VR the virtuality of the environment felt less disruptive than on paper. This could be explained that virtuality feels more natural in Virtual Reality as it is automatically part of it. Virtuality in this case referring to the artificial nature of the environment. Vague unfinished AI structures were not visible on a printout. It helped participants to focus better, and some felt more active while looking at the paper image than in the VR world. The 2D picture was mentioned being more “cozy” than VR environment. It was easy for many to imagine the sounds in the scenery through the 2D image, e.g. water and birds in scenario A and the loud noise of wind turbines in scenario B.

The evaluation of the positive and negative feelings evoked by the scenarios produced conflicting answers. For example, P1 felt that the experience of scenery B was more positive in 2D than in VR. Whereas, P3 was surprised how negative feelings were also transmitted through just a picture.

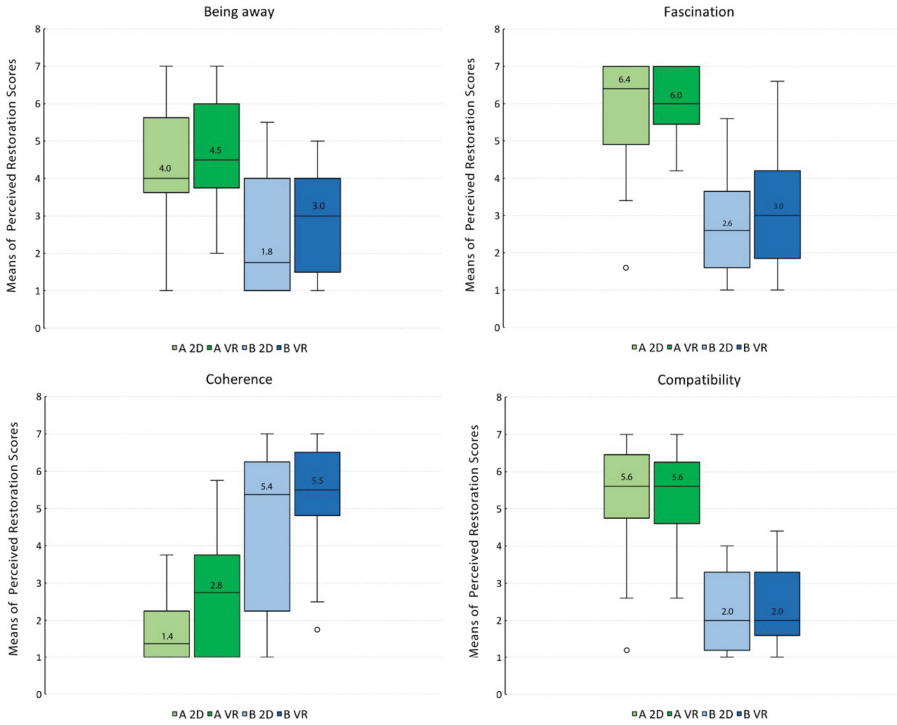


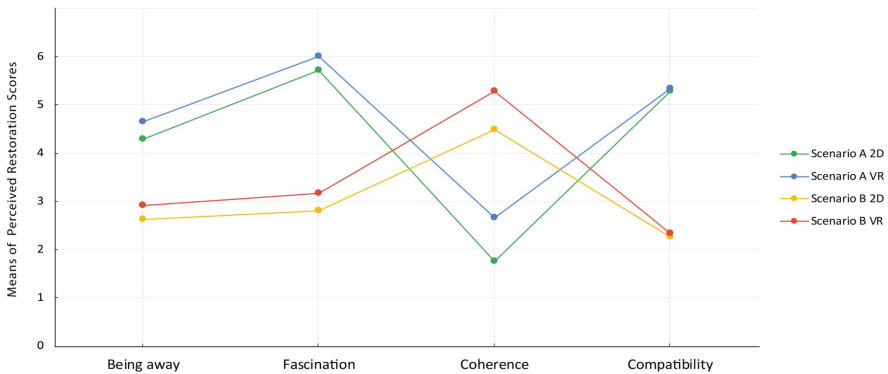
Fig. 6. Boxplot of perceived restoration divided by four categories in both scenarios with VR and 2D (n = 14)

On the negative side, the abundance and proximity of wind turbines were said to spoil the nature experience, e.g. “Wind turbines spoil the landscape. You wonder what kind of noise the turbines make if you walk among them” (P14). However, it was not seen to affect the fascination with the place itself since large-scale mechanical things can be as fascinating to watch as nature, e.g. “Wind turbines could also be fascinating objects, mainly to look at and wonder. Technology is also interesting along with nature” (P6).

A more neutral viewpoint argued that wind turbines were a surprising element after scenario A. When viewed from a distance, wind turbines were a somewhat neutral sight. But in the foreground of the picture, they arouse the viewer’s desire to quickly pass by. The presence of wind turbines seemed somehow pervasive and disturbing. The presence of strange wind turbines was said to give the place a cursed impression which confused the viewer. Visiting there might be interesting but wind turbines were seen as somehow threatening as if they could attack you, e.g., “a dystopian landscape where I just want to cut down the wind turbines” (P12).

It was also mentioned that in scenario B human influence on the landscape was clearly present which made the place feel not as mentally relaxing as in

With PRS, the VR version got higher scores in each category, Fig. 5. The greatest difference was with coherence, where the difference between VR and 2D was 0.9 in scenario B and 0.78 in scenario A. Overall, scenario B was evaluated as more confusing, distracting and chaotic than scenario A. Also, compatibility with scenario A was between 5–6, whereas scenario B was between 2–3. Participant P9 explained that for them, an environment that was as natural as possible seemed to be the most interesting and suitable for their personality. Several responded similarly, e.g. P3 when describing Scenario B in VR: “although there was a lot of interesting things in this place, the scenery was depressing”. P1 did not feel a connection to a place like presented in scenario B, but mentioned that seeing many wind turbines in the same place would be an experience itself. They could visit the place, but would not like to spend long periods of time there. More detailed differences of the perceived restoration can be seen in Fig. 6. For example, in coherence, there is a noticeable dispersion in the 2D version of scenario B. Some of the participants mentioned that confusion, distraction, or chaos is not always a negative thing. That might partly explain the split in the responses.



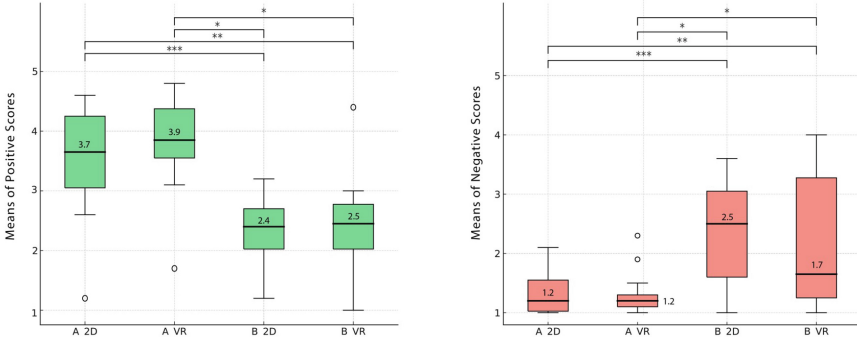
**Fig. 5.** Means of perceived restoration divided by four categories in both scenarios with VR and 2D ( $n = 14$ )

## 4.2 Green Energy and Sustainability

Scenario B who’s landscape was dominated by wind turbines, aroused mixed feelings in the viewers. Sustainability issues came to mind for many and both pros and cons were shared in the answers. More on the positive side of opinions the progress in the use of wind power was seen as a good thing. Also, huge wind turbines were mentioned to arouse interest or curiosity, e.g. “The landscape and the view are new. I have never been near wind turbines built so close together. A confusing, but at the same time curious state of being” (P10). On the other hand, the built environment was also stated to be more boring compared to nature.

**Table 1.** Significant Pairs

Dataset	Scenario 1	Scenario 2	Statistic	p-value	Significant (p < 0.0083)
Positive	A 2D	A VR	13.0	0.040921	False
Positive	A 2D	B 2D	0.0	0.002200	True
Positive	A 2D	B VR	5.0	0.001221	True
Positive	A VR	B 2D	0.0	0.000122	True
Positive	A VR	B VR	0.0	0.000122	True
Positive	B 2D	B VR	38.0	0.936938	False
Negative	A 2D	A VR	10.0	0.495868	False
Negative	A 2D	B 2D	0.0	0.002209	True
Negative	A 2D	B VR	0.0	0.003346	True
Negative	A VR	B 2D	1.0	0.002852	True
Negative	A VR	B VR	1.0	0.002852	True
Negative	B 2D	B VR	16.0	0.439659	False



**Fig. 4.** Boxplot of Positive and Negative Affect Schedule in scenarios A and B with VR and 2D (n = 14); Significant Pairs \*(p = 0.0001), \*\*(p = 0.001), \*\*\*(p = 0.002)

**Table 2.** Statistics (PANAS)

	Dataset	A 2D	A VR	B 2D	B VR
Mean	Positive	3.528571	3.857143	2.321429	2.378571
Median	Positive	3.650000	3.850000	2.400000	2.450000
Standard Deviation	Positive	0.940072	0.805476	0.609134	0.842321
Mean	Negative	1.321429	1.307143	2.342857	2.192857
Median	Negative	1.200000	1.200000	2.500000	1.650000
Standard Deviation	Negative	0.355568	0.370995	0.922967	1.125552



**Fig. 3.** Left a participant evaluation of the environment from a picture. Right a participant watching the scenery in VR.

their familiarity with VR, one had tried VR once, 4/14 a few times (1–5), and 9/14 had used VR more than five times.

## 4 Results

### 4.1 Comparing Scenarios Without (A) and with (B) Windmills

Statistical tests were run separately for positive and negative datasets (PANAS). First, a non-parametric Friedman test was conducted to examine whether there were significant differences in participant responses across the different scenarios and viewing methods: A 2D, A VR, B 2D, and B VR. The results of the Friedman test indicated a significant difference between the conditions,  $\chi^2(3) = 29.82$ ,  $p < .001$ . Similarly, the results of the negative dataset indicated a significant difference between the conditions,  $\chi^2(3) = 24.93$ ,  $p < .001$ . To further investigate these differences, pairwise comparisons were conducted using Wilcoxon Signed-Rank Tests with a Bonferroni correction applied to account for multiple comparisons ( $\alpha = 0.0083$ ). The results are presented in Table 1.

On average, participants associated more positive words towards nature with minimal human impact than nature with extensive human impact. Figure 4 illustrates how scenario A without windmills scored higher (3.53; 3.86) on the positive affect schedule than scenario B with windmills (2.32; 2.38) Table 2. Scenario A was also perceived as less negative (1.32; 1.31) than scenario B (2.34; 2.19). The difference between positive and negative affect was seen in the case of scenario A where positive affect surpasses 3.5 and negative affect stays under 1.5. On the contrary, scenario B was perceived almost as positive as negative. Generally, VR versions of the scenarios were perceived slightly more positively and less negatively than the 2D version in both scales. However, in the case of VR version of scenario B there is great dispersion between answers.



Fig. 1. AI generated virtual environment, left Scenario A, right Scenario B

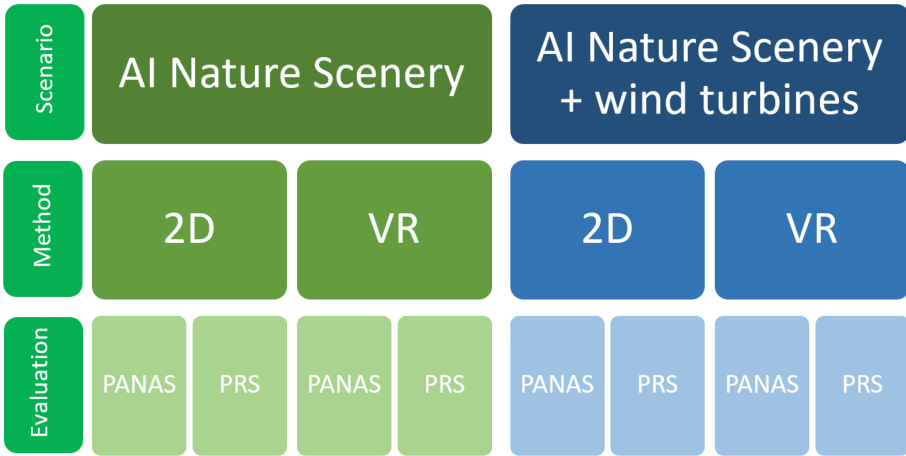


Fig. 2. Study structure

interaction with the software, with a researcher changing the images remotely using the pc when the participant indicated that they were ready to move to the next image. The wireless connection induced no significant latency that would affect the user experience. Half of the subjects were shown VR first (Fig. 2). After spending 5–10 min exploring the VR environments (scenario A and scenario B), the participants evaluated their experience using PANAS and PRS, and filled out the questionnaire on a computer. The other half of the subjects started by viewing 2D printed versions of the same scenarios and completing a paper version of the same questionnaires. The time participants spent in the study varied from 30 min to two hours, with an average of 45 min. The scenarios were shown to all the participants in the same order, scenario A first and scenario B second.

### 3.2 Participants

In total, the study included 14 participants (seven men, six women, and one non-binary person), with the ages between 18–24 (2/14), 25–34 (9/14), 35–44 (2/14), and 45–54 (1/14), with affiliation in different disciplines of design. Considering

### 3 Research Method

#### 3.1 Study Set-up

**The Virtual Nature Environment.** The study used two different scenarios of nature environments generated by AI (Fig. 1). SkyboxAI was used to create the 360 image. The used prompt was: Northern Europe Forest landscape, European spruce, European red pine, birch, blue river with waterfall, subshrub, lush plants. The idea was to generate a virtual nature environment that resembled the environment that the subjects were used to seeing. After three rounds of generation, the result was satisfying enough. However, AI wanted to include mountains and vegetation that might distance the virtual nature from the natural setting of the study area. A 360 picture was then edited in Adobe Photoshop. Using Photoshop's generative fill tool the environment was modified. The used prompt was: Wind turbines park. This time the AI tool had some issues completing the task. The wind turbines were distorted and too small. After seven rounds of generation, the best option was chosen from the 21 edited images. The image quality deteriorated during the editing process, although the 360 perspective remained intact. Also, the picture itself was quite provocative, changing the landscape significantly from the unedited version (lack of trees and mountains). Both mentioned factors were expected to affect the VR experience.

**Questionnaire with PANAS and PRS.** In the study, two different scales were used to evaluate the experience. Positive and negative Affect Schedule (PANAS) [42] and Perceived Restoration Scale (PRS) [14] Both scales were offered in English and with the subjects' native language. As the scales are commonly used in studies exploring the restoration of virtual and real environments, they were deemed as suitable for comparing a virtual nature experience through VR and from the 2D printed image. In PANAS there were 10 positive words and 10 negative words that respondents estimated in what extent the virtual nature made them feel that emotion, on a five-point Likert scale where 1=Not at all/very slightly; 5= Extremely. The total score is calculated by finding the sum of the 10 positive items, and then the 10 negative items. Scores range from 10–50 for both sets of items. For the total positive score, a higher score indicates more of a positive affect. For the total negative score, a lower score indicates less of a negative affect. [42]

Four elements Being away, Fascination, Coherence, and Compatibility of Perceived Restoration were measured on a seven-point Likert scale with sixteen statements. (1=Not at all; 7=Completely). After each section, the respondent was asked to explain in a few words the reasons that made them feel the way they had described above.

The scenarios were shown to the subjects in two forms, as a 360° image in VR via a head-mounted display and a 2D image printed on A4-sized paper, Fig. 3. Subjects could hold the picture in their hand during the evaluation. The VR headset was the Meta Quest 3, ensuring sufficient graphical fidelity and field of view, which was connected wirelessly to a computer. The participants had no

ments presented using VR video can alleviate stress to some extent. Although an artificial environment is not as effective in relieving stress physically as a natural environment, it can still relieve stress psychologically. The use of static 360 videos from peaceful natural locations with local cultural context and minimal visible or audible human influence is an effective approach when designing virtual natural environments for stress reduction [17].

### 2.3 Restoration and Measuring the Affect

One approach to studying the various positive effects of nature is restoration. Restoration means recovery from stress, and renewal of a person's physical, psychological, and social resources, which have decreased due to adaptation to stress factors triggered by the environment [13, 19]. According to Ulrich et al.'s [39] Stress Reduction Theory (SRT), nature, green spaces, and images of nature support restoration. Attention Restoration Theory (ART) suggests that spending time in nature or viewing it can improve mental fatigue and concentration [18]. According to ART, restorative experiences consist of four elements: being away, fascination, coherence, and compatibility. These four categories are used as a base of the Perceived Restorative Scale (PRS), which is a tool developed by Hartig et al. [14] to measure the restorative quality of environments. One of the main goals is to give planners a measurement tool that could be used to assess the effects of current and future spaces. The scale uses ART properties as measurement attributes. The object is evaluated according to 16 statements with a seven-point scale.

Another scale commonly used to evaluate feelings is the Positive and Negative Affect Schedule (PANAS) developed by Watson et al. [42]. This mood scale consists of 20 words, half of which measure positive affect and the other half negative affect. Each item is rated on a five-point scale. Because the PANAS is a self-report questionnaire, it can be inaccurate in assessing people's mood. The answer is always subjective and depends on the reviewer's interpretation of the word and their ability to estimate their feelings. Also, not every word applies to every type of situation, so it must be taken into account when analyzing the results.

When estimating the restorative qualities of the environment, several measurements are often combined to gain a broader understanding of the factors affecting the entire experience. The research of Mattila et al. [27] examined the effects of restorative experiences in an immersive virtual reality forest environment by using the Subjective Vitality Scale (SVS), Positive and negative Affect Schedule (PANAS), Restoration Outcome Scale (ROS) and Perceived Restoration Scale (PRS). The results showed that the VR environment and a real forest were generally as restorative, but the virtual environment was perceived as more fascinating and coherent. Often virtual reality is considered to be a better method than traditional 2D media in delivering audiovisual stimuli. However, researchers have differing findings about the restorative effectiveness of VR when compared with nature photos or actual physical nature [24].

There is no single definition of VR. The field is dynamically evolving due to new research, developing technologies, and expanding application possibilities. Classical definitions of VR include, e.g. “real-time interactive graphics” [7]. However, many scholars consider VR as a broader concept [2]. Milgram and Kishino’s virtuality continuum raises discussion on the distinction between real and virtual [29]. Kardong-Edgren et al. suggests immersion and presence as a way to define VR [20]. Since higher immersion results in higher presence, it is an important factor in the VR experience. Among the various VR definitions, three features are common: immersion, presence, and interaction with the virtual environment [7]. VR technology encompasses non-immersive, semi-immersive, and fully immersive spaces depending on the degree of presence and used equipment [9, 11]. Non-immersive VR utilizes screens to deliver the VR experience. In this case, the virtual world does not obscure the real world and users do not feel complete presence in the scene [11, 33]. Also, only some of a user’s senses are stimulated. Semi-immersive experiences, e.g. a flight simulator combining both physical and virtual elements [7, 11]. Fully immersive VR creates an illusion of “being there” by replacing the physical surroundings and making the person feel like they are present in the virtual world [9, 31]. Typically, fully immersive VR is presented through head-mounted displays (HMD) and interaction is enabled via hand-held controllers. thus the definition of VR can slightly vary between different sources and use situations. Therefore, in this study, VR relies on a broader definition of virtual reality including non-interactive viewing of 360 pictures through an HMD.

## 2.2 Nature Connectedness

Humans have an evolution-based need to feel the connection with nature [21, 44]. That concept is called biophilia-love of life, which Wilson [21, 44] defines as the desire to connect with living beings and the natural environment. This natural tendency of humans urges them to seek connections with nature and other forms of life, including a longing to be close to nature and a preference for things that resemble it. Richardson and Butler [32] emphasize the problem of how biophilic properties commonly remain superficial in practical design projects as the most important result of biophilic design, i.e. a deeper and stronger connection with nature, is often overlooked.

Nature connectedness is defined as an individual’s perceived relationship with the natural world [26]. Often it is discussed in a context of pro-environmental behavior [37], a variety of health and wellbeing outcomes, and overall happiness [4]. The basis for nature connectedness is created in childhood, but even short experiences in nature can increase connection with nature in adults [35]. A study by Calogiuri et al. [3] demonstrates that using 360 videos and computer-generated scenarios to create immersive virtual nature, can be equally effective in increasing nature connectedness, as long as the same level of immersion and interaction is offered.

Studies have presented that nature connectedness has a significant effect on health, wellbeing, and happiness [4]. Wang et al. [40] report that forest environ-

such as windmills become a visible part of the landscape in many places. Despite their environmental benefits, these technologies harm the local natural ecosystems by replacing nature with a constructed environment, and affecting animals such as reindeer [36] and birds [22]. The green energy constructions also alter the visual and experiential qualities of natural environments. This raises the question about green transition technologies' impact on nature engagement and experience. Thematically, our research is part of a larger discussion of just transition for the green transition, highlighting the complex nature of sustainable technologies, green transition, and the effects on local and larger societal frameworks [16, 23].

More research is needed on participatory design for pluriversal perspectives on sustainability. VR can be a useful tool in the domain, and its suitability for such studies needs to be evaluated. As the contribution of our work, we report the pros and cons of using VR compared to paper-based visualizations of nature landscapes without and with technology constructions. By exploring the methodologies to investigate green transition, we pave the way to better user research methods and research tools for citizen participation in green transition decision-making.

## 2 Related Work




### 2.1 HCI and Nature Experiences

In the area of Human-Computer Interaction, the combination of nature experiences and technology has been investigated from different viewpoints. Webber et al. [43] have presented typologies in use technology for nature engagement, and bring up dimensions of directness and distance in nature engagements. These dimensions of directness include direct, abstracted, and simulated nature experiences. It has been highlighted that technology has the potential to enhance engagement with nature, promoting wellbeing and awareness of natural environments as well as its fragility [12]. However, designing technology for nature encounters requires addressing the balance between support and intrusiveness, especially in mobile contexts, where devices can both enhance and disrupt the user's interaction with nature and outdoors [28].

VR and AR technologies can contribute to linking individual users with environmental issues by providing immersive experiences for learning, increasing users' self-efficacy, and influencing behavior change [8]. However, while technology enables simulated nature experiences, it often lacks the multisensory qualities of authentic nature. For connecting with nature, extended reality (XR) technologies have been used for immersive experiences with nature [10], embodying animals in immersive virtual environments [1], and for gamified experiences for environmental awareness [34]. A systematic literature review of 80 papers focusing on VR and AR with environmental sustainability revealed that most commonly the technologies were used for education and learning purposes, followed by promoting ecological behavior [8] such as for behavioral change in plastic consumption [5].



# Exploring VR as a Method to Study User Perceptions of Challenges in Green Transition Technology in Nature

Petra Nurmela<sup>(✉)</sup> , Juri Etto , and Jonna Häkkinen 

University of Lapland, Rovaniemi, Finland  
{petra.nurmela,juri.etto,Jonna.Hakkila}@ulapland.fi

**Abstract.** In this paper, we explore Virtual Reality (VR) as a tool for user studies assessing nature perceptions. We collect user perceptions of green transition technology in nature, and compare a natural landscape and one with windmills, both as paper and VR representations, in a user study ( $n = 14$ ). Our salient findings indicate that while VR provides more immersive and engaging experiences, paper-based visualizations allow more freedom for the user's own imagination and reflections on personal experiences on green technologies in nature. Moreover, in VR, the participants tend to get more distracted by the visual details, for instance the implausibilities of the virtual world. The landscape without windmills was perceived as more positive (PANAS) and more restorative (PRS).

**Keywords:** nature · HCI · virtual reality · user studies · user experience · green transition · sustainability

## 1 Introduction

Nature connection has been reported to improve both physical and mental health and wellbeing [45, 46], and also emerging technologies have been utilized for fostering this connection [12, 38]. This paper explores the use of VR in assessing user perceptions of nature landscapes and compares how people perceive a natural landscape with a landscape containing built technology. Our research contribution is two-fold. Firstly, we investigate the method of using VR for assessing human perceptions of natural landscapes and compare the approach to paper-based visual presentations of nature landscapes. Secondly, we explore people's perceptions of integrating a green transition technology, windmills, into the landscape visualizations. The motivation of our research lies in evaluating and developing research methods to assess technology-mediated nature experiences.

Our work is partially motivated by the global trend of green transition, which includes setting up new large-scale technological solutions [41]. As we increasingly seek sustainable solutions to address environmental concerns, technologies

18. Wickens, C.D.: Virtual reality and education. In: 1992 IEEE International Conference on Systems, Man & Cybernetics, pp. 842–847. IEEE (1992). <https://doi.org/10.1109/ICSMC.1992.271688>
19. Witmer, B.G., Jerome, C.J., Singer, M.J.: The factor structure of the presence questionnaire. *Pres. Teleoper. Virtual Environ.* **14**(3), 298–312 (2005)
20. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. *Presence* **7**(3), 225–240 (1998)
21. Zechner, O., Kleygrewe, L., Jaspert, E., Schrom-Feiertag, H., Hutter, R.V., Tscheligi, M.: Enhancing operational police training in high stress situations with virtual reality: experiences, tools and guidelines. *Multimodal Technol. Interact.* **7**(2), 14 (2023)

## References

1. Allcoat, D., Von Mühlennen, A.: Learning in virtual reality: effects on performance, emotion, and engagement. *Res. Learn. Technol.* (2018)
2. Apostolakis, K.C., Dimitriou, N., Margetis, G., Ntoa, S., Tzovaras, D., Stephanidis, C.: Darlene—improving situational awareness of European law enforcement agents through a combination of augmented reality and artificial intelligence solutions. *Open Res. Europe* **1** (2021)
3. Barkoukis, V., et al.: Virtual reality against doping: the case of project VIRAL. In: Auer, M.E., Tsiatsos, T. (eds.) *IMCL 2021. LNNS*, vol. 411, pp. 487–496. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-96296-8\\_44](https://doi.org/10.1007/978-3-030-96296-8_44)
4. Caserman, P., et al.: Virtual reality simulator for police training with ai-supported cover detection. In: *Joint International Conference on Serious Games*, pp. 181–193. Springer, Heidelberg (2023). [https://doi.org/10.1007/978-3-031-44751-8\\_13](https://doi.org/10.1007/978-3-031-44751-8_13)
5. Conges, A., et al.: Situational awareness and decision-making in a crisis situation: a crisis management cell in virtual reality. *Int. J. Disast. Risk Reduct.* **97**, 104002 (2023)
6. Kinateder, M., Warren, W.H.: Social influence on evacuation behavior in real and virtual environments. *Front. Rob. AI* **3**, 43 (2016)
7. Kleygrewe, L., Hutter, R.V., Koedijk, M., Oudejans, R.R.: Virtual reality training for police officers: a comparison of training responses in vr and real-life training. *Police Pract. Res.* **25**(1), 18–37 (2024)
8. Lønne, T.F., Karlsen, H.R., Langvik, E., Saksvik-Lehouillier, I.: The effect of immersion on sense of presence and affect when experiencing an educational scenario in virtual reality: a randomized controlled study. *Heliyon* **9**(6) (2023)
9. Narciso, D., Melo, M., Rodrigues, S., Paulo Cunha, J., Vasconcelos-Raposo, J., Bessa, M.: A systematic review on the use of immersive virtual reality to train professionals. *Multimedia Tools Appl.* **80**, 13195–13214 (2021)
10. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. *J. Manag. Inf. Syst.* **24**(3), 45–77 (2007)
11. Ponder, M., et al.: Immersive vr decision training: telling interactive stories featuring advanced virtual human simulation technologies. In: *Proceedings of the Workshop on Virtual Environments 2003*, pp. 97–106 (2003)
12. Poulidou, P., Ourda, D., Barkoukis, V., Palamas, G.: Empowering young athletes: elevating anti-doping education with virtual reality. *EAI Endorsed Trans. e-Learn.* **9** (2023)
13. Reis, R.S., Hino, A., Añez, C.: Perceived stress scale. *J. Health Psychol.* **15**(1), 107–114 (2010)
14. Rosenkvist, A., et al.: Hearing with eyes in virtual reality. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1349–1350. IEEE (2019)
15. Scorgie, D., Feng, Z., Paes, D., Parisi, F., Yiu, T., Lovreglio, R.: Virtual reality for safety training: a systematic literature review and meta-analysis. *Saf. Sci.* **171**, 106372 (2024)
16. Unity Technologies: Unity Game Engine [Computer Software], Version 2022.3.22f1 (2022). <https://unity.com/>
17. Voigt, L., Frenkel, M.O.: How officers perform and grow under stress: police training in virtual reality. In: *Police Conflict Management*, vol. II: Training and Education, pp. 187–211. Springer, Heidelberg (2023)

long-term impact of VR training to determine whether these initial gains in performance and engagement are sustained over time.

Another potential area for future work is the incorporation of advanced features such as eye-tracking technology. Eye-tracking could offer more precise data on where participants are focusing their attention, allowing for more detailed analysis of performance. This would also enable trainers to identify areas where participants might be struggling or distracted, providing opportunities for targeted interventions.

Lastly, while this study focused on apartment searches, VR training can be expanded to include other areas of law enforcement, such as conflict de-escalation, crowd management, or tactical operations. Expanding the use of VR beyond basic operational training would help prepare police officers for the diverse scenarios they may encounter in the field.

## 6 Conclusion

This study examined the effectiveness of a Virtual Reality (VR) simulator as a supplementary tool for developing situational awareness skills in police training, particularly situational awareness during apartment search operations. The findings demonstrate that while VR enhances user confidence and allows for quick adaptation to virtual environments, the usability of control devices and visual display quality play critical roles in determining task focus and performance outcomes.

Control device interference was found to significantly impact participants' concentration and overall task performance, highlighting the need for further optimization of control mechanisms. Additionally, visual interference was strongly correlated with control issues, indicating that improvements in both areas are essential for enhancing the VR training experience.

Although stress was shown to slow down task completion times, it did not significantly affect performance accuracy, suggesting that participants were able to maintain precision under pressure. This indicates that VR can effectively simulate high-pressure environments without negatively impacting task quality, making it a valuable tool for preparing police cadets for real-world challenges.

However, the novelty of VR technology and the small sample size used in this study suggest that further research is necessary. Future studies should focus on larger participant groups, extended real-world application, and potential improvements to control devices and visual displays. Incorporating advanced technologies such as eye-tracking could provide more precise data on participant focus, enabling more targeted training interventions.

Overall, the VR simulator presents a promising and scalable solution for enhancing soft skill training, particularly situational awareness, in police academies. By addressing the identified limitations, VR can play an integral role in preparing law enforcement officers for complex and high-stakes scenarios through immersive and efficient training environments.

The usability of the VR interface, which participants rated positively, highlights the importance of user-friendly design. Participants were able to adjust to the controls relatively quickly, but the moderate correlation between control device interference and adjustment difficulty ( $r = 0.4$ ) suggests that even minor improvements in the intuitiveness of the interface could further enhance the user experience and training outcomes.

### 5.3 Effect of Stress on Performance

The study also revealed that while stress had a moderate negative correlation with task completion times ( $r = -0.45$ ), it did not significantly affect task performance quality. This finding aligns with previous studies by Caserman et al. [4], which indicated that VR-induced stress might slow down decision-making but does not necessarily impair the ability to perform tasks accurately.

This is particularly relevant in police training, where cadets must be able to function effectively under high-pressure situations. The ability of VR to simulate stress-inducing environments without negatively impacting task accuracy demonstrates its value as a training tool. Future iterations of VR simulators should aim to manage stress levels to improve task efficiency without compromising performance quality.

### 5.4 Training Efficiency and Practical Implications

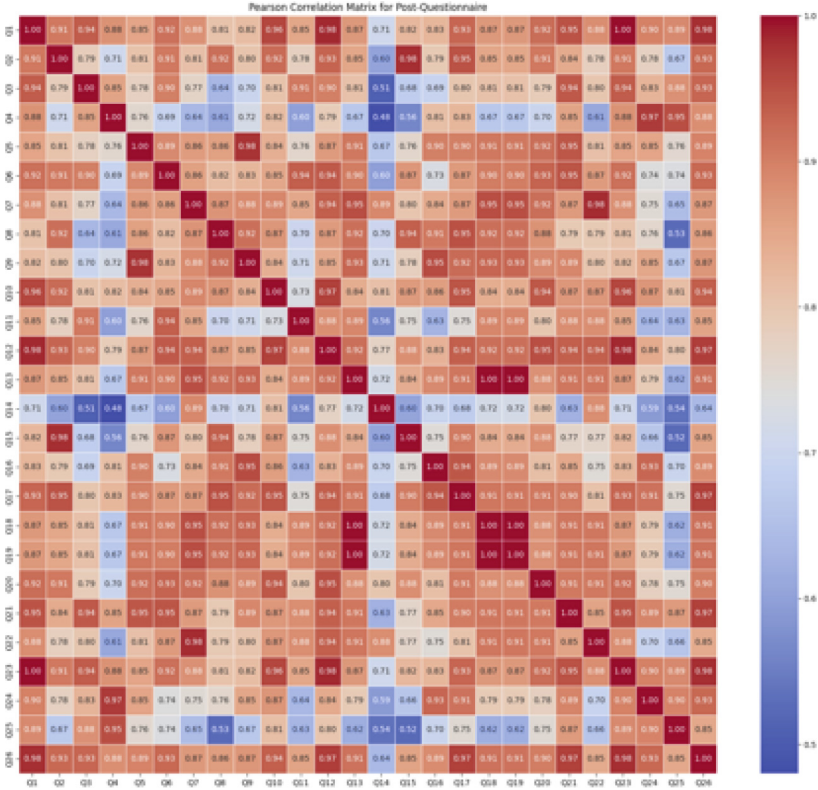
The relatively short average task completion time (180s) indicates that VR can efficiently replicate critical elements of traditional training, such as apartment search operations, without requiring extensive time and resource investments. VR's adaptability and scalability allow for the rapid creation of varied training environments, making it a valuable tool for police academies that may face logistical challenges when providing frequent training opportunities.

The strong correlation between control device interference and visual interference emphasizes the need for improving both control mechanisms and visual display quality to ensure high levels of training fidelity. If addressed, these improvements could make VR an even more effective and scalable alternative to traditional physical setups.

### 5.5 Challenges and Future Directions

While the results are promising, there are several challenges to be addressed in integrating VR as a mainstream training tool for police academies. Control device interference was found to significantly affect participants' focus and task performance. Therefore, improving the intuitiveness of the control devices and minimizing visual interference should be a priority in future developments of VR training programs.

Additionally, the novelty of VR technology may have contributed to initial enthusiasm, which could influence the results. Future studies should explore the

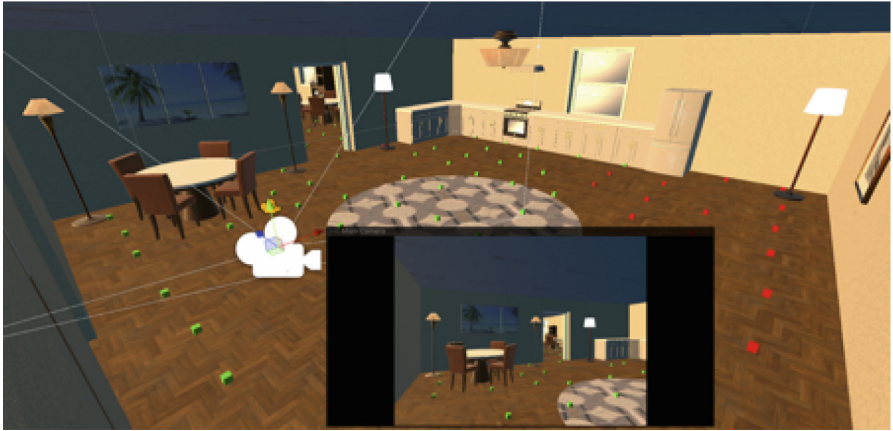


**Fig. 9.** Pearson correlation matrix for post-questionnaire results, illustrating relationships between control device interference, task performance, sense of immersion, and other factors. Strong correlations are in red, and weaker correlations are in blue. (Color figure online)

Moreover, the strong correlation between control device interference and visual interference ( $r = 0.7$ ) suggests that these two factors may compound the overall difficulty of performing tasks in VR. This indicates that improving control mechanisms must go hand in hand with ensuring high-quality visual displays to minimize distractions and improve performance.

### 5.2 Role of Immersion and Usability

The sense of immersion plays a critical role in the effectiveness of VR training. As noted by Witmer et al. [19], immersion is key to engaging users deeply in simulated scenarios. In this study, participants who reported higher immersion levels also experienced better task performance. This reflects the importance of not only providing intuitive control devices but also ensuring that the overall VR experience is seamless, with minimal technical distractions.



**Fig. 8.** Trainee's perspective during the training session, showing secured points in the virtual environment. This live scoring system allows the instructor to monitor the trainee's progress in real-time.

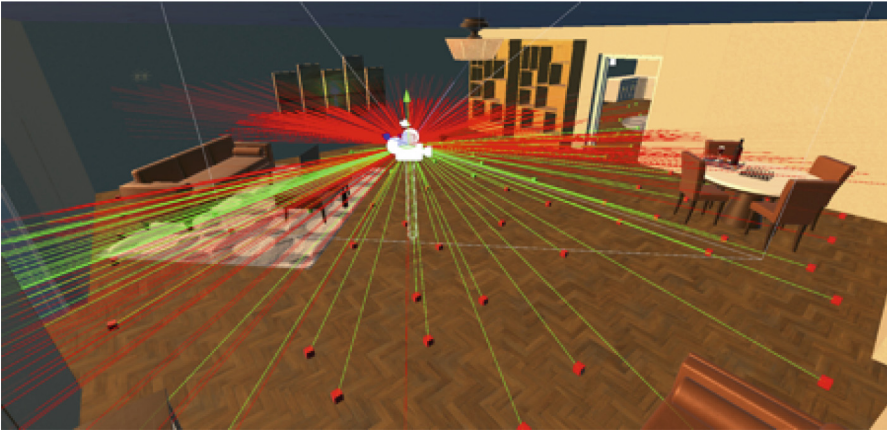
These findings suggest that improving control mechanisms and minimizing visual distractions could enhance task performance in VR training environments. Addressing stress management could further improve task efficiency without sacrificing performance quality.

## 5 Discussion

The results of this study suggest that Virtual Reality (VR) can be a viable tool for supplementing training in tasks such as apartment searches. By leveraging immersive environments, VR offers potential for improving both perceived and actual task performance. This section discusses the implications of these findings, aligns them with existing literature, and explores potential challenges and opportunities for integrating VR into police training programs.

### 5.1 Impact on Training Outcomes

The study results highlighted the importance of control device design and usability in influencing task performance in VR training. Participants who experienced less interference from control devices reported better focus, leading to improved task performance. The moderate negative correlation between control interference and task focus ( $r = -0.5$ ) aligns with previous research by Narciso et al. [9], which emphasized the role of ease of use in enhancing the effectiveness of VR training. This finding suggests that optimizing control devices could significantly enhance training outcomes by enabling participants to better concentrate on task objectives rather than the mechanisms of interaction.



**Fig. 7.** Visualization of the operator’s field of view and sensor rays used for tracking areas observed during the training session. These rays help the simulator detect which parts of the virtual apartment the operator has examined.

As shown in Fig. 9, the Pearson correlation matrix highlights significant relationships between various aspects of the VR experience. In particular, control device interference (*Q02*) shows strong correlations with task focus (*Q06*) and visual interference (*Q07*), suggesting that difficulties with control devices and visual display quality often co-occur and can negatively affect task performance.

## 4.2 Impact of Perceived Stress on Performance

A moderate negative correlation ( $r = -0.45$ ) was found between perceived stress and task completion time, indicating that participants who experienced higher levels of stress tended to take longer to complete their tasks. However, the correlation between stress levels and performance quality ( $r = 0.21$ ) was not significant, showing that while stress affected the speed of task completion, it did not reduce the overall quality of task execution.

## 4.3 Key Findings

The results of the study point to several key conclusions:

- Control Device Interference: Control device interference significantly impacted task performance and focus, with greater interference correlating with decreased concentration and a reduced sense of control.
- Visual Interference: Control device interference often coincided with visual display issues, compounding distractions and impairing task execution.
- Stress and Task Completion: Higher stress levels were associated with slower task completion times but did not affect task quality, indicating that participants maintained accuracy despite feeling stressed.



**Fig. 6.** Detailed view of a room with furniture in the VR simulator. This image highlights the intricate details included in the virtual environment to improve immersion and support cadets' situational awareness during training.

ronment. Question *Q02* (“How much did the control devices interfere with the performance of the assigned tasks?”) revealed significant correlations with various aspects of the VR experience, indicating that control mechanisms played a critical role in shaping task performance and user experience.

A negative correlation ( $r = -0.5$ ) between *Q02* and *Q06* (“How well could you concentrate on the assigned tasks rather than on the mechanisms used to perform said tasks?”) indicates that greater interference from the control devices was associated with reduced task focus. Our findings align with prior studies showing VR's impact on focus and stress handling. For example, in anti-doping training, VR was found to support focus and situational awareness under pressure, a critical aspect of high-stakes decision-making [3]. This parallels the need for police officers to remain focused and effective in demanding training scenarios. Participants who experienced more difficulty with the control devices had greater trouble concentrating on the task objectives.

A strong positive correlation ( $r = 0.7$ ) between *Q02* and *Q07* (“How much did the visual display quality interfere or distract you from performing the assigned tasks?”) suggests that participants who reported interference from the control devices also experienced difficulties with the visual display quality. This highlights the potential interconnectedness between control and visual issues, which could amplify the overall interference with task execution.

A moderate positive correlation ( $r = 0.5$ ) between *Q02* and *Q01* (“How much influence did you feel you had over interactions with the virtual apartment?”) indicates that as interference from the control devices increased, participants reported feeling less control or influence over their interactions within the virtual environment. This further emphasizes the need for user-friendly controls to enhance participants' sense of agency.



**Fig. 5.** Room with basic furniture in the VR simulation. The inclusion of furniture increases the complexity of the virtual apartment, enhancing the realism of the training and challenging cadets to identify potential risks and obstacles.

by participants were reflected in their actual performance during the simulation. Furthermore, improvements in self-reported confidence post-training provided additional evidence supporting the effectiveness of the VR simulator as a supplementary training tool.

### 3.5 Limitations

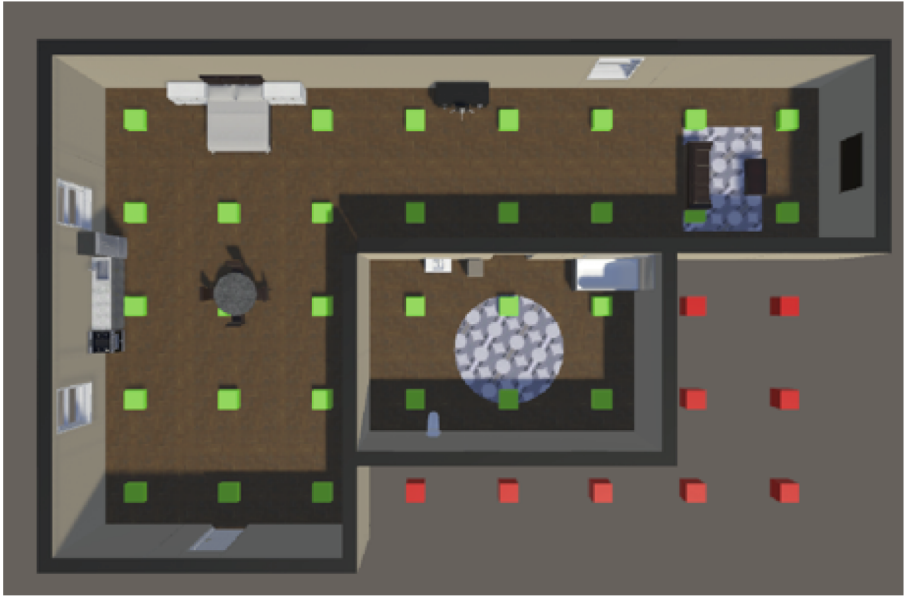
Despite the promising findings, there were several limitations to this study. The small sample size of participants limits the generalizability of the results to a broader population. Additionally, there is the potential for novelty bias, as participants may have been more enthusiastic about using the VR system due to its novelty rather than focusing solely on the training content. Although every effort was made to mitigate this by clearly communicating the training objectives to participants, the influence of novelty on their responses remains a potential limitation.

The technical capabilities of the VR equipment also imposed certain restrictions. For instance, the simulator lacked advanced features such as eye-tracking, meaning that only head-tracking was used to determine where participants were looking. This limitation may have impacted the precision of the performance metrics, although the system still provided valuable insights into how well participants navigated and cleared risk zones.

## 4 Results

### 4.1 Control Device Interference and Task Performance

The analysis of post-questionnaire responses provided valuable insights into how control device interference affected participant performance within the VR envi-



**Fig. 4.** 3D model of a room with checkpoints. This image shows the integration of checkpoints within the training environment, guiding cadets through key areas they need to secure during virtual apartment searches.

VR simulations. The primary aim was to develop the prototype and evaluate it by determining the correlations between the participants' self-reported confidence, perceived stress, and their objective performance scores and time metrics.

The analysis involved comparing pre- and post-questionnaire results to evaluate how the VR training impacted participants' confidence and sense of presence. For instance, the self-reported confidence levels from the pre-questionnaire were compared against those in the post-questionnaire, allowing for a measure of improvement after completing the VR training. Box plots and other visualizations were used to illustrate the spread and central tendency of responses, helping to identify trends in how participants perceived their training experience.

Additionally, correlation matrices were used to identify relationships between variables, such as the association between perceived stress and performance scores or time taken to complete the simulation. High correlation scores between these variables helped determine the extent to which stress influenced performance in the VR environment. Correlation matrices were particularly useful in revealing patterns, such as whether participants with higher initial confidence saw greater performance gains from VR training, or whether stress levels significantly affected task completion time.

To ensure that the findings were robust, the performance metrics from the simulator were cross-referenced with the subjective data. This cross-analysis enabled the study to explore whether improvements in confidence as reported



**Fig. 3.** Overhead floor plan of a virtual apartment. This image shows the layout used to guide cadets in navigating virtual environments, providing a clear view of risk zones and checkpoints within the simulation.

The pre-questionnaire aimed to gather information about the participants' prior experience with VR, their perceived stress levels using the Perceived Stress Scale (PSS) [13], and their confidence in conducting apartment searches. Additionally, participants were asked to estimate how effective they believed VR would be as a tool for improving their search abilities. The outcome of this question compared with the actual performance score and helped to determine the

After completing the VR training, participants filled out a post-questionnaire based on the Presence Questionnaire by Witmer et al. [20], which assessed their sense of presence and engagement in the virtual environment. Further questions were included to evaluate the usability of the control devices and how much the training experience improved their confidence in conducting apartment searches.

Objective data was collected directly from the VR simulator during the training exercises. Two key metrics were recorded: (1) the time taken to complete the apartment search and (2) the performance score, which was calculated based on the trainee's ability to correctly secure all risk zones in the virtual environment. Each risk zone was equipped with sensors, and the system automatically marked them as secure when they were observed by the trainee during the simulation.

### 3.4 Data Analysis

Data analysis focused on both the subjective responses from the pre- and post-questionnaires as well as the objective performance metrics collected during the



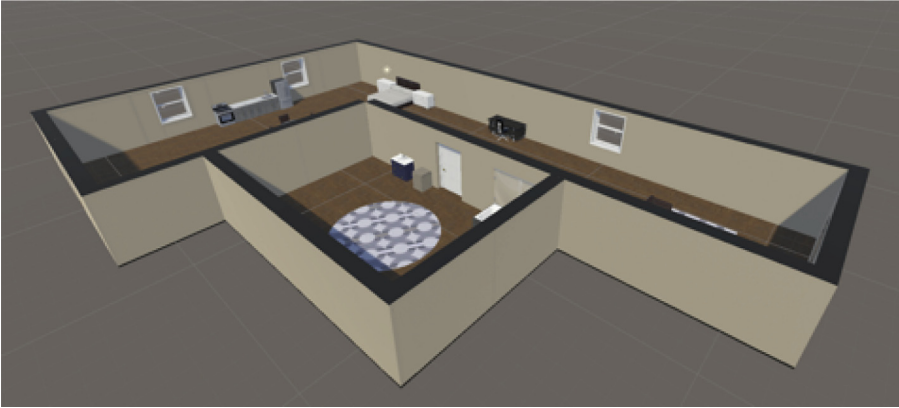
**Fig. 2.** Another view of a 3D room model used in the VR simulation. This provides a different perspective of the training environment, emphasizing the variability in apartment layouts for different training scenarios.

session. If a sensor is obscured by furniture, it will automatically be considered secure at the start of the session unless there is enough space between the sensor and the furniture to allow for proper observation. Each secured area contributes to the operator's score, which is assessed by the simulator (Fig. 8).

The automated assessment system was designed to accurately detect which zones in the apartment have been properly checked without compromising the simulator's performance. This allows instructors to provide live feedback based on the operator's score as they secure each risk zone.

### 3.3 Data Collection

To evaluate the effectiveness of the VR simulator as a training tool, both subjective and objective data were collected during the study. The subjective data consisted of pre- and post-questionnaires administered to the participants, while the objective data were automatically collected by the simulator during the training sessions.



**Fig. 1.** 3D model of a room used in the VR simulator. This model represents the basic environment in which police cadets train for apartment searches, helping them develop situational awareness and decision-making skills.

helps cadets develop situational awareness and learn to focus on key objectives even when presented with cluttered or visually stimulating surroundings.

Although the visual fidelity of the simulator balanced simplicity and realism, the detailed models (see Figs. 5 and 6) ensured that the training environments were engaging without being overwhelming. Previous research by Wickens [18] showed that excessive realism could detract from learning by introducing unnecessary visual information. Therefore, the simulator maintained an optimal level of detail to help cadets concentrate on identifying risks, such as the checkpoints seen in Fig. 4, and improve their ability to secure each zone effectively.

Overall, the VR simulator’s design aimed to maximize immersion while focusing on core training objectives, such as decision-making under pressure and the development of situational awareness, which is a critical skill for law enforcement officers.

### 3.2 Simulator Monitoring and Performance Scoring

The simulator was responsible for monitoring the various inputs from the operator and using these inputs to score performance through data collection. Inputs were collected via the VR headset and the two VR controllers held by the user.

The simulator determines which objects are within the operator’s field of view using the headset’s position and rotation sensors. Since precise eye-tracking equipment was unavailable, the simulator assumes that everything in the operator’s field of view is being observed. This data was essential in facilitating the functionality that allows the operator to secure a risk zone (Figs. 7).

Whether an area of the apartment is secured or unsecured is determined by sensors placed in key risk zones. These sensors are considered secured once they have been observed at least once, and they stay secure for the remainder of the

### 3.1 Artifact Development

The development of the VR training simulator followed an iterative process, divided into two phases: the creation of a level editor and the development of the VR simulator itself. The level editor enabled easy assembly of diverse apartment layouts, while the VR simulator executed apartment search scenarios in real-time. This iterative approach allowed for continuous refinement based on feedback from the police academy instructors, ensuring the final artifact met the specific training needs of law enforcement.

**Level Editor.** The development of the VR simulator began with the creation of a level editor, which allowed for the rapid design and customization of virtual apartment layouts used during training. This editor was designed to enable instructors to easily configure rooms by adjusting the layout and positioning of various elements such as walls, doors, and windows. By using the editor, trainers could quickly generate diverse environments that simulated real-world apartment settings.

The 3D models of these rooms (see Figs. 1 and 2) demonstrate the basic structure and variability of the apartment search environments used in the VR simulation. These environments provide the foundation for training police cadets in situational awareness, an essential soft skill for identifying potential risks and hazards in unfamiliar spaces.

In addition to the 3D room models, an overhead floor plan (Fig. 3) was generated for each virtual apartment. This floor plan outlines key areas, such as risk zones and checkpoints, that trainees need to clear during the apartment search operation. The floor plan ensures that each simulation is well-structured, enabling cadets to navigate the environment efficiently and focus on securing critical areas.

The flexibility provided by the level editor allows trainers to modify environments as needed, ensuring that trainees encounter a wide range of scenarios. This capability is critical for helping police officers develop the quick decision-making and situational awareness necessary in real-life operations.

**VR Simulator.** The VR simulator was designed to run on the Oculus Quest 3, providing an immersive training experience for police cadets. The system tracked the cadets' head and hand movements, enabling them to interact with and navigate through virtual apartments while searching for risk zones. As demonstrated in previous studies, room-scale VR simulations can create a high level of immersion, which is critical for training exercises that require situational awareness and decision-making under stress [4].

The virtual environments used in the simulator featured various levels of complexity, starting from basic layouts and advancing to more intricate spaces furnished with obstacles. Figures 5 and 6 illustrate these environments, where cadets were challenged to identify and secure risk zones amid furniture and other potential distractions. This progressive increase in environmental complexity

training emergency responders by providing them with realistic, stress-inducing environments that mirror real-world incidents. This suggests that VR's capability to create dynamic, responsive training environments is especially beneficial for police officers, who must remain calm and effective in unpredictable, high-risk situations.

Moreover, the use of VR for decision-making in crisis situations has shown significant potential, Conges et al. [5] emphasized the importance of training police officers and crisis managers in simulated environments where decision-making under stress is a key factor. This aligns with research by Witmer et al. [19], which showed that presence and immersion in VR directly impact how effectively trainees engage with the training scenarios, ensuring better retention of learned skills.

Despite the growing body of research on VR's effectiveness in various professional fields, there remains a gap in the literature regarding its specific application to soft skill training for police officers, particularly in apartment search operations [7]. This study seeks to address that gap by investigating how VR can be used to enhance situational awareness, decision-making, and risk management in such contexts. By utilizing a design science methodology, we aim to evaluate the extent to which VR can complement traditional training methods while providing additional flexibility and scalability [17].

This research builds on existing studies and aims to contribute new insights into the role of immersive technology in law enforcement training. By focusing on situational awareness skills, this study adds a critical dimension to the broader conversation about how VR can be leveraged to enhance the training of professionals who regularly operate in high-stress, life-threatening environments.

### 3 Methodology

This study employed a design science research methodology, which emphasizes the iterative development of artifacts through collaboration with stakeholders, ensuring the practical applicability of the results [10]. The primary goal of this research was to prototype a Virtual Reality (VR) simulator aimed at supplementing traditional training for police cadets. Specifically, the study focused on apartment search operations, where trainees must identify and secure risk zones under pressure, simulating real-world scenarios that require situational awareness and decision-making skills.

The study was conducted at Malmö University Police Academy in Sweden, where police instructors served as stakeholders and provided critical feedback throughout the artifact development process. The VR simulator was designed and implemented using the Unity game engine [16], and the hardware employed included the Oculus Quest 3 VR headset, which offered a portable and immersive environment for training.

can stimulate situational awareness by creating a sensory-rich experience. Such multisensory simulations play a crucial role in developing high-pressure decision-making skills essential for police officers in training. Allcoat et al. [1] found that VR contributes significantly to learning outcomes by enhancing engagement, immersion, and interactivity. Blacker et al. [21] demonstrated the effectiveness of VR in improving decision-making under stress, particularly during high-pressure military operations, further highlighting the potential for VR in law enforcement training.

Despite these clear advantages, the integration of VR into official training curricula has been slow due to logistical, technological, and financial constraints [15]. Traditional police training methods often involve the creation of physical environments, such as mock apartment complexes or training facilities, which are both resource-intensive and difficult to adapt to different scenarios. As police training continues to evolve to address increasingly complex threats, VR offers a flexible, cost-effective alternative capable of generating diverse training scenarios on demand.

The educational benefits of VR are well-documented beyond law enforcement. Wickens [18] argued that while VR environments should focus on key learning objectives, it is essential to avoid unnecessary visual complexity that could overwhelm trainees. Peffers et al. [10] further highlighted the importance of structured design in developing training systems to ensure that they align with pedagogical goals. This balance between immersion and simplicity has informed the design of many contemporary VR training systems, including those used for police training.

Skills training, such as situational awareness and decision-making, is a central aspect of police education. For example, in apartment search operations, officers must quickly identify and secure risk zones, an essential task in law enforcement operations [21]. However, traditional training often lacks the scalability and adaptability needed to prepare officers for the full range of environments they may encounter in the field. VR can address this gap by offering a platform where training scenarios can be endlessly modified and adapted to different contexts, thus providing a richer, more comprehensive training experience [8].

Previous research has highlighted the role of immersion in VR-based learning environments. Ponder et al. [11] emphasized that immersion and user interaction are critical factors for the successful transfer of skills from virtual to real-world scenarios. Their work on interactive storytelling in VR found that more immersive environments resulted in better user engagement and skill retention. Narciso et al. [9], in their systematic review of VR applications in professional training, concluded that VR is particularly effective in environments requiring complex decision-making, such as those faced by law enforcement officers.

Furthermore, advances in VR technology have made it possible to integrate real-time performance tracking and automated feedback into training sessions, providing trainers with valuable data on trainee performance. This real-time assessment allows for more precise evaluations of skill development, particularly in high-pressure scenarios. Kinateder et al. [6] found that VR was effective in

training, where cadets are required to develop skills such as situational awareness, decision-making, and risk management in complex and unpredictable scenarios.

Traditional police training often involves physically replicating environments where trainees must practice techniques such as securing risk zones in apartment searches. These physical setups can be costly, time-consuming, and may not always provide the variety needed to adequately prepare cadets for the diverse situations they will encounter in the field. Additionally, ensuring psychological resilience in high-pressure situations requires repeated exposure to stress-inducing scenarios, something that can be difficult to achieve consistently through traditional means.

VR offers an alternative by providing immersive, interactive environments that can replicate these complex training scenarios with high fidelity. In a VR simulation, cadets can engage in tasks such as clearing rooms, identifying risk zones, and making split-second decisions in a controlled yet realistic virtual setting. These environments are highly customizable, allowing for varied scenarios to be presented without the need for extensive physical resources.

This study explores the use of VR as a supplementary tool for training in police academies, specifically within the context of apartment search operations. The primary goal of this research was to prototype a Virtual Reality (VR) simulator aimed at supplementing traditional training for police cadets. More specifically, to which degree can VR replicate the immersive experience of real-world scenarios, maintain training quality, and contribute to the psychological preparedness of police students?

The evaluation of the VR system focused on key factors such as user immersion, interactivity, and performance outcomes. The results of this study aim to contribute to the broader understanding of how VR can be effectively utilized as a supplement to traditional training in fields where situational awareness is critical to success under pressure.

## 2 Background

As Virtual Reality (VR) technology continues to evolve, its applications in professional training, particularly in law enforcement, have gained increasing attention. VR offers the ability to simulate complex, high-stakes environments that are difficult or expensive to replicate in the physical world, making it a powerful tool for training in areas such as situational awareness, decision-making, and risk management. These skills are critical for police officers in high-pressure field operations where the ability to make split-second decisions can have life-or-death consequences [2].

VR has shown great potential in the realm of police training, with multiple studies demonstrating its effectiveness. Caserman et al. [4] explored VR's role in tactical police training using AI-supported simulations for cover detection, showing how immersive environments can enhance critical thinking and tactical skills. VR enhances sensory immersion, a key factor for skill development in complex environments. In [14] Rosenkvist et al. (2019) demonstrated how VR



# Utilizing Virtual Reality to Enhance Situational Awareness in Swedish Police Training

Hampus Hansen, Måns Holmgren, Carl Johan Gribel, Joakim Ingrell ,  
and George Palamas 

Malmö University, Malmö, Sweden  
georgios.palamas@mau.se

**Abstract.** This study examines police students experience of using Virtual Reality (VR) as a tool for training situational awareness during apartment search operations. The study included both a qualitative and quantitative approach to evaluate a prototype aimed to supplement traditional training. The findings highlight that control device and visual display interference significantly affect task performance and concentration. While stress increased task completion times, it did not reduce accuracy, demonstrating the value of VR in simulating high-pressure environments. VR offers a flexible, scalable, and cost-effective alternative to traditional training methods. However, optimizing control mechanisms and visual fidelity remains essential. Future research should explore advanced features like eye-tracking to further enhance the training experience. These findings contribute to the growing body of research on VR's role in police and vocational training.

**Keywords:** Virtual Reality · Police Training · Situational Awareness · Decision-Making · Immersion · Crisis Management · Real-Time Performance Tracking · AI-Supported Training · Law Enforcement · VR Simulation · Training Technology · High-Stakes Environments

## 1 Introduction

Virtual Reality (VR) technology has seen significant advancements in recent years, leading to its increasing adoption across a range of sectors, including education, healthcare, and law enforcement. In particular, VR has shown potential as a valuable tool for training in high-stakes environments, where traditional training methods may be resource-intensive and limited in scope. Beyond traditional law enforcement applications, VR has been increasingly recognized for its educational value in high-stakes environments. For instance, VR-based anti-doping education for young athletes has shown that immersive environments can effectively promote ethical decision-making and situational skills [12]. This underscores VR's versatility as an educational tool, validating its application in diverse training scenarios such as police work. One such environment is police

3. Bravou, V., Oikonomidou, D., Drigas, A.S.: Applications of virtual reality for autism inclusion. a review. *Retos: nuevas tendencias en educación física, deporte y recreación* (45), 779–785 (2022)
4. Carnett, A., Neely, L., Gardiner, S., Kirkpatrick, M., Quarles, J., Christopher, K.: Systematic review of virtual reality in behavioral interventions for individuals with autism. *Adv. Neurodev. Disord.* **7**(3), 426–442 (2023)
5. Fleury, S., Chaniaud, N.: Multi-user centered design: acceptance, user experience, user research and user testing. *Theor. Issues Ergon. Sci.* **25**(2), 209–224 (2024)
6. Kim, G.J., et al.: A swot analysis of the field of virtual reality rehabilitation and therapy. *Presence* **14**(2), 119–146 (2005)
7. Københavns Kommune: Klimainvestering ii - mindre kørsel af elever i specialtilbud og modtagerklasser (2022). <https://www.kk.dk/sites/default/files/agenda/fl431fc0-ce1e-4a37-aa71-0a47d879f013/e4a05e77-9029-4479-8051-5b23d8fe10b7-bilag-7.pdf>. Accessed 22 Apr 2024
8. Miller, I.T., Wiederhold, B.K., Miller, C.S., Wiederhold, M.D.: Virtual reality air travel training with children on the autism spectrum: a preliminary report. *Cyberpsychol. Behav. Soc. Netw.* **23**(1), 10–15 (2020)
9. Nikki, M., Skjoldborg, P.K.B., de López, K.M.J.: The efficacy of head-mounted-display virtual reality intervention to improve life skills of individuals with autism spectrum disorders: a systematic review. *Neuropsychiat. Dis. Treat.* **18**, 2295–2310 (2022). <https://doi.org/10.2147/NDT.S331990>
10. Parsons, S., Cobb, S.: *State-of-the-Art of Virtual Reality Technologies for Children on the Autism Spectrum*. Routledge, Abingdon (2016)
11. Pfeiffer, B., Davidson, A.P., Brusilovskiy, E., Feeley, C., Kinnealey, M., Salzer, M.: Effectiveness of a peer-mediated travel training intervention for adults with autism spectrum disorders. *J. Transp. Health* **35**, 101781 (2024)
12. Pfeiffer, B., Sell, A., Bevans, K.B.: Initial evaluation of a public transportation training program for individuals with intellectual and developmental disabilities. *J. Transp. Health* **16**, 100813 (2020)
13. Preece, J., Rogers, Y., Sharp, H.: *Interaction Design: Beyond Human-Computer Interaction*, 6th edn. Wiley, Hoboken (2023). <https://www.williamdam.dk/interaction-design-beyond-human-computer-interaction-sixth-edition...2887867>
14. Reger, G.M., Parsons, T.D., Gahm, G.A., Rizzo, A.: Virtual reality assessment of cognitive functions: a promising tool to improve ecological validity. *Brain Inj.* **7**, 24–26 (2010)
15. Rowland, D.: Redefining autism. *J. Neurol. Psychiat. Brain Res.* **2020**(01) (2020)
16. Simões, M., Bernardes, M., Barros, F., Castelo-Branco, M.: Virtual travel training for autism spectrum disorder: proof-of-concept interventional study. *JMIR Ser. Games* **6**(1), e8428 (2018)
17. Sudirjo, F., Dewa, D.M.R.T., Kesuma, L.I., Suryaningsih, L., Utami, E.Y.: Application of the user centered design method to evaluate the relationship between user experience, user interface and customer satisfaction on banking mobile application. *Jurnal Informasi Dan Teknologi*, 7–13 (2024)

of three sessions per participant, highlighting a significant deviation from our goals.

Student 10 received six sessions in a row, while Student 11 only received one VR session, both showing some improvements during post-observations. Despite the teachers' enthusiasm and dedication, the logistical difficulties and constraints faced by the teachers and pedagogues at Frejaskolen, such as time management and scheduling, posed significant barriers.

Future iterations should include some changes to the VR train station simulation. This includes scenarios like delayed or canceled trains and last-minute track changes. Additionally, the simulation does not accurately reflect the usual crowd at Anonymous station. Anonymous station is typically quite busy, not represented by the number of NPCs walking around in the current version of the application. The number of passengers on the trains is also lacking, as trains would be almost full during peak hours, and children would have to wait their turn to get on or off the train. Since the target group might have sensory issues, including this aspect in the simulation could prepare the participants for the sensory overload they might experience at a crowded station or on a packed train. This was, however, not possible in this version due to the limited computational capabilities of the Meta Quest 2.

Additionally, in real life, children would sometimes encounter a ticket inspector, which is not part of the current simulation. Including this element could provide a valuable learning experience, teaching children how to interact with the inspector and presenting scenarios where others might not have a ticket. These situations could lead to raised voices or other behaviors, creating unfamiliar situations for the children to navigate.

Our study demonstrates VR's potential to bridge gaps in practical skill acquisition for children with ASD, while highlighting the need for robust and adaptable VR systems. Future work could explore the integration of AI to dynamically adjust scenarios based on real-time user feedback and behavior, enhancing the personalized learning experience. Additionally, creating an application on a different device, such as a tablet, would allow the teachers to follow the student's performance in real-time as well as enable them to make changes to the simulations, such as the number of NPC passengers, noise level or even introduce events such as train delays, cancellations and changes of tracks.

This short paper aims to explore VR's opportunities and illustrate its potential. It hopes that future research and development can make significant strides in utilizing VR to support and enhance the independence and quality of life for individuals with ASD and other special needs.

## References

1. Attardo, D.: Kommunerne bruger 1,4 milliarder: Helt vildt (2023). <https://ekstrabladet.dk/nyheder/politik/kommunerne-bruger-14-milliarder-helt-vildt/10003890>. Accessed 22 Apr 2024
2. Autistic Self Advocacy Network: About autism (2015). <https://autisticadvocacy.org/about-asan/about-autism/>. Accessed 14 May 2024

and introduce them to the steps required for independently utilizing this form of public transportation. In this section, we will discuss the results of the two participants who went through all phases of the planned evaluation: pre-observation, VR training, and post-observation: student 10 and student 11.

The data collected from pre- and post-VR training observations demonstrated improvements in the participants' skills in key areas, such as recognizing transportation signs, understanding schedules, and managing check-in systems. These findings suggest that the VR intervention can potentially enhance independent public transportation skills.

During the pre-observational phase, participants exhibited varying levels of confidence and familiarity with the public transportation system. Notably, some participants relied heavily on habitual behaviors or required peer support. For instance, Student 10 focused exclusively on reaching the platform to catch the train, demonstrating an unawareness of the need to take a different train. Conversely, in the post-observational phase, the same student displayed marked improvements in confidence and attentiveness, consistently checking relevant screens, successfully navigating to the correct platform, and entering the correct train. This transformation, observed after six VR training sessions with Student 10, underscores the effectiveness of VR training in altering habitual behaviors and enhancing situational awareness.

Similarly, Student 11 exhibited challenges during the pre-observational phase, including a complete lack of awareness regarding when to disembark from the train. Pedagogues had to intervene as the student had traveled one stop too far. During post-observation, however, Student 11 demonstrated considerable improvement by independently getting off at the correct station. The progress after just one training session highlights the potential for further advancements with additional training.

Interestingly, the students spent more time in the real train station after VR training compared to the pre-training observations. Student 10's time increased from 2 min to 5:44 min, and Student 11's time increased from 3 min to 6:42 min. This increase in time suggests that the students were perhaps more attentive to their surroundings during the second attempt, likely because they had trained in the VR simulation. The VR training may have made them more aware of their environment and more diligent in following the steps necessary to navigate the station independently. However, it is also possible that unknown factors influenced their performance, and further investigation would be required to understand these changes entirely.

The special education school participating in this study has VR and gaming on their schedule for their students. Therefore, the teachers received complete control of the VR training sessions, and the authors had no interaction with the students during the whole process. This, however, removed the ability to control the training phase, resulting in training not being performed as planned due to scheduling issues at the school. Out of the 11 participants initially participating in the pre-observation phase, only five proceeded to the training phase, with varying numbers of sessions. Only one participant met the expected minimum

- **Student 6:** Conducted one session, spent 4 min and 4 s, asked for help, entered the correct train on the first attempt but missed one task, and noted a lack of realism.
- **Student 10:** Conducted six sessions over 9 days. In the first session, the student spent 10 min and 57 s, did not ask for help, and entered the correct train on the second attempt. In subsequent sessions, the student consistently entered the correct train on the first attempt, noted task completion discrepancies, and discovered glitches.
- **Student 11:** Conducted one session, spent 34 min and 53 s, asked for help initially, entered the correct train on the second attempt, completed most tasks, had issues with the virtual school transportation card, and seemed to enjoy the experience.

In summary, 12 sessions were conducted by 5 participants. Most boarded the correct train on the first session, with Students 10 and 11 requiring a second attempt. Participants generally completed the tasks, though Student 6 missed one. Technical issues included glitches and incorrect displays, with varied participant behaviors observed.

### 3.3 Post-observations

Due to scheduling issues, only two participants, Student 10 and Student 11, were observed in this phase. The following details summarize their actions and behaviors during the observation.

- **Student 10:** Checked the departure screens at the entrance, outside 7-Eleven, and at Platform 5. Successfully checked in and boarded the correct train. No conversation was recorded during the observation. The observer noted that Student 10 was very aware of their surroundings, checked all screens, and felt more confident and attentive after the training. The student spent 5:44 min from entering the station to walking onto the right train compared to 2 min in the pre-training observation phase.
- **Student 11:** Checked the departure screens at the entrance, outside 7-Eleven, and at Platform 5. Successfully checked in and boarded the correct train. During the observation, Student 11 asked, “Which side of the travel card is the correct side?” The observer noted some confusion between virtual and real experiences, but the student used the tools learned in the simulation to navigate successfully. The student spent 6:42 s in the station during the post-VR training observation compared to 3 min during the pre-VR training observation.

## 4 Discussion

This paper explored whether a VR intervention can be used to familiarize autistic children attending a special education school with their local train station

- **Participant 6:** Recognized the stop at Carlsberg station but needed assistance to get off.
- **Participant 10:** Initially went to the wrong platform and was guided back by the accompanying teacher
- **Participant 11:** Zoned out, missed the correct stop and did not get off, according to the accompanying teacher.

All participants successfully boarded the correct train. One participant (11) did not get off at the right station.

### 3.2 VR Training Sessions

Due to scheduling issues at the school, not all students who participated in the observation phase received VR training. Five students tried the VR application following the observation phase. Each participant was tasked with navigating a virtual train station, entering the correct train, and getting off at the current station. A total of four subtasks were tracked by the VR application: Looking at the information screen, going to the right track, using the travel card to purchase a ticket, and walking onto the correct train. The aim was to assess their ability to complete these tasks independently and to observe their behaviors and any difficulties encountered (Fig. 4).

Participant	Sessions	Asked for help	Correct train 1st attempt	Task Completion	Issues
<i>Student 1</i>	2	No	Yes	No	
<i>Student 8</i>	2(1st)	Yes	Yes	No	
<i>Student 6</i>	1	Yes	Yes	No	
<i>Student 10</i>	6	No	No(1st), Yes	No	
<i>Student 11</i>	1	Yes	No	Yes	

Fig. 4. Training phase results

#### Session Summaries.

- **Student 1:** Conducted two sessions 14 days apart. In the first session, the student spent 25 min and 52 s, did not ask for help, entered the correct train on the first attempt, and noted a wall clipping glitch. In the second session, the student spent 17 min and 42 s, did not ask for help, entered the correct train on the first attempt, and noted a glitch with external elements intruding into the train.
- **Student 8:** Conducted two sessions 7 days apart. In the first session, the student spent 4 min and 3 s, asked for help due to difficulty reading the phone, entered the correct train on the first attempt, and noted a technical issue. In the second session, the student spent 13 min and 14 s, did not ask for help, and entered the correct train on the first attempt but encountered an incorrect station name display, causing a premature exit.

study. Out of the 11 participants who participated in the pre-observations, only five received VR training due to issues such as scheduling conflicts, student and teacher illnesses, and other logistical constraints. Ultimately, only two participants could participate in the post-VR training observations at the real train station. While these challenges altered our original study design, they highlight the complexities and realities of conducting research in natural settings, where flexibility is often necessary to accommodate unpredictable variables.

### 3 Results

#### 3.1 Pre-observations

The pre-observational study was conducted with 11 participants aged between 11 and 16 years who were tasked with navigating the Anonymous train station to assess their ability to utilize public transport independently. For clarity and confidentiality, each participant was assigned an ID number during pre-observation. These ID numbers correspond to their student numbers throughout the study (e.g., Student 1 was ID 1, Student 2 was ID 2, etc.).

The main goal was to evaluate each participant's ability to interact with the environment and other individuals and to check departure screens, locate the right platform, and board the correct train.

**Screen Checks.** Of the eleven participants, six (54.5%) were observed checking the departure screens. Among these, three individuals checked the screen outside the 7-Eleven store, while the remaining three participants checked it at the entrance. The average duration for these checks was approximately 5 s, indicating a quick interaction with the digital information displays.

**Platform Identification and Boarding.** All participants successfully identified and moved towards the correct train platform and boarded the correct train. A notable exception was one participant, Student 10, who initially walked down to platform 5, contrary to the task requirements. This was explained by the student's habitual use of platform 5 for their daily travel, illustrating a reliance on routine over task-specific instructions.

**Time.** The time taken by participants from entering Anonymous station (start observation) to boarding the train (end observation) varied. The minimum time recorded was between one minute, and the maximum time was three minutes seconds.

- **Participant 1:** Struggled with navigation and asked strangers for directions.
- **Participant 2:** Demonstrated independent awareness and readiness.
- **Participant 3:** Displayed familiarity with the train system, checking screens and boarding with minimal hesitation.

The school card is used to purchase train tickets, and the phone tracks objectives. The main objective is to take a specific train and get off at the correct station, with several sub-objectives like checking departure screens, navigating to the correct track, and checking in before boarding.

Upon boarding the correct train, players transition to a new scene inside the train. The game spawns them in the corresponding train cart and requires them to stay attentive to red dots on the screen indicating journey progress. Additionally, the player could look out via the windows to see signs indicating which station has been reached. Successfully getting off at the correct station leads to a congratulatory room displaying the number of completed subtasks, with options to restart or quit the game. Missing the stop or getting off at a wrong stop results in a room encouraging a retry.

## 2.2 Evaluation

A before-and-after training evaluation employing a quasi-experimental design measured the VR simulation’s impact on children with ASD’s ability to navigate public transportation independently. This evaluation took place in real-world train stations to ensure authentic behavior observation, focusing on interactions with informational screens, schedule verification, platform safety procedures, and overall navigation.

The first part of the evaluation involved initial observations at Annonymus train station with teachers accompanying the children to the Anonymous train station. Two authors, unfamiliar to the children, were strategically positioned at the station to observe from a distance without influencing behavior. They recorded interactions with the environment, such as checking departure screens and the time to board the train. The children only met and interacted with their teachers throughout the study, with whom they were already familiar. The teachers received signed parental approval for the children participating in the research and approval from the anonymous school ethical counsel.

After the observation session, a debriefing meeting was held at the Annonymos school to combine the teachers’ insights with the observers’ data, ensuring an accurate understanding of each child’s behavior. This collaborative approach is a core component of UCD, enhancing the reliability of collected data [5].

A total of 11 observations were conducted on students aged 11 to 16. Following the initial observations, teachers were introduced to setting up the VR intervention, including establishing a guardian system, starting the application, setting NPC and ambient sound levels, and restarting the app. They were provided with a comprehensive setup manual and an online questionnaire to complete after each session, including items on subtask completion and observation notes. The training was planned to be conducted once a week for three weeks.

Post-training observations followed the same procedure as the pre-training observations, ensuring consistency and comparability of the data collected.

Conducting a study in a natural, real-world environment without interference from the researchers inherently brings unpredictable challenges [13]. In our case, despite initial plans, several unforeseen events affected the execution of the

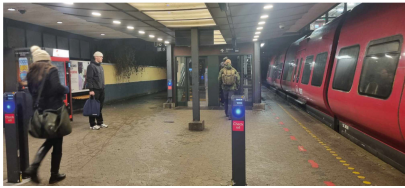


(a) Real life

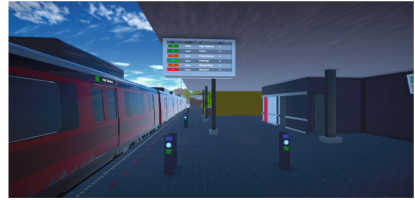


(b) Virtual

**Fig. 1.** Comparison of real-life and the virtual juice shop by the entrance of the train station and a departure information screen



(a) Real life



(b) Virtual

**Fig. 2.** Comparison of real-life and virtual train station elements

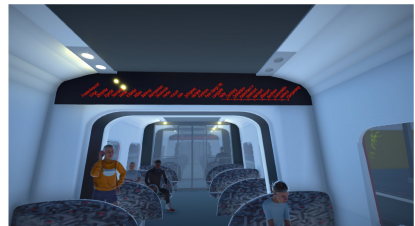
Interactive elements within the game, such as interacting with train doors and ticket machines using VR controllers, were implemented to simulate real-life interactions. Each session starts at a control panel, allowing teachers to switch on and off elements such as the presence of other passengers and ambient sounds to control sensory stimuli.

The Non-Player Characters (NPCs) are designed to wander around the station, wait, or stand together in conversations, creating a realistic environment. Ambient sounds include traffic noise, background diegetic music from the shops, and announcements similar to those at real train stations in Anonymous.

Players control a human avatar with inverse kinematics-managed arms, equipped with a school card in the right hand and a phone in the left hand.



(a) Real life



(b) Virtual

**Fig. 3.** Comparison of real-life and virtual journey progress indicators

with a smartphone [8]. Over three weekly sessions, participants showed improvement in navigating check-in, security, and boarding processes in a real-world airport.

The benefits of VR in teaching transportation skills to individuals with ASD are multifaceted. It can enhance independence and societal integration and provide a cost-effective solution to the escalating expenses associated with their transportation needs [6]. Given the significant investment in transportation for students with ASD and the innovative capabilities of VR technology, this approach warrants thorough exploration as a sustainable and effective educational tool. Additionally, VR offers ecological validity in training interventions as the tasks and experiences in VR-based simulations can be designed to be similar to those that are required in the real world [14].

In contrast to previous studies, this short paper presents a VR intervention simulating a virtual train station, mirroring a real-world station's appearance and functionality. Our VR environment is designed to replicate essential tasks such as checking departure times, purchasing tickets, finding the correct track, and boarding and exiting the train-offering an immersive experience that parallels the actual procedures required in a real station. This localized, detailed simulation specifically focuses on train travel, representing a novel contribution to the field by providing a practical, real-world scenario for autistic children to practice independent transportation skills

## 2 Methods

The development of the VR intervention for public transportation training in autistic children was grounded in User-Centered Design (UCD) principles. UCD prioritizes the needs, preferences, and constraints of end-users throughout the design and development phases to ensure the final product is both functional and user-friendly [5, 17].

In collaboration with Annonymus school, a specialized institution for children with special needs, our development process involved iterative feedback loops with teachers and pedagogues at the school. Several design workshops were held at the school to gather input and feedback. Based on these workshops, it was decided that the VR simulation should realistically mimic the Anonymous S-train station frequently used by the students for their daily commutes.

### 2.1 The VR Intervention

The VR simulation was developed using Unity3D to run on a Meta Quest 2 and was designed to closely replicate the environment of the Anonymous S-train station. To ensure a lifelike experience, 3D models of real station elements-such as ticket check-in stands, departure screens, and various interiors and shops-were meticulously crafted to look like their real-life counterparts (see Figs. 1a, 1b, 2a, 2b, 3a, 3b).

ing approximately 1.4 billion (anonymous currency) annually [1]. This expenditure has been increasing; for instance, Anonymous’s budget for student transportation grew from 47.9 million (anonymous currency) in 2018 to 78.2 million (anonymous currency) in 2022, largely due to the rising costs of individual taxi services and an increase in the number of diagnosed children [7]. Despite efforts to manage these costs through centralizing services and encouraging collective transportation solutions, significant financial and logistical challenges remain.

Personalized support and interventions are crucial to promote independence and improve the quality of life for individuals with ASD. Pfeiffer et al. [11] investigated the effect of peer-mediated travel training to help autistic individuals achieve independent travel skills and community mobility. The results showed that peer-mediated training significantly improved the travel skills of the participants. Another study by Pfeiffer et al. [12] showed the effectiveness of the Kennedy Center’s travel training guide and curriculum (TCG), where individualized training with educated travel coaches helped autistic individuals gain the necessary skills for independent travel. These methods are, however, resource intensive, requiring trained peers or educators. Another disadvantage of using trained peers or coaches with autistic individuals is the social barriers to interacting with another person, which is one of the most common difficulties of autistic individuals [11, 12].

Virtual Reality (VR) can provide a more cost-effective intervention, offering a socially safe environment where autistic individuals can practice essential skills without the pressure or discomfort of direct social interactions.

Previous studies have demonstrated the potential of VR in teaching DLS to autistic children and adolescents [3, 4]. VR allows individuals with ASD to engage in immersive simulations of real-world scenarios in a controlled and safe environment [10]. This technology facilitates repeated practice and mastery of DLS, enabling personalized learning experiences that accommodate the unique needs and learning paces of individuals with ASD. VR simulations offer error-free learning and self-directed exploration opportunities, which are crucial for building confidence and competence in navigating public transportation systems [9].

A recent review by Carnett et al. highlights the efficacy of VR in behavioral interventions aimed at enhancing independence among autistic individuals [4]. This review supports VR’s use in teaching specific skills like driving and interview techniques, advocating for its inclusion as an evidence-based practice. Of the 23 studies reviewed, two address transportation difficulties using VR.

One study by Simoes et al. describes a VR intervention for teaching public bus usage to individuals with ASD [16]. This game provided a three-dimensional city environment where participants practiced tasks associated with bus travel, such as finding the appropriate bus, purchasing a ticket, finding a seat, and getting off at the correct stop. Results indicated significant improvements in participants’ knowledge and comfort with bus travel procedures.

Another study involved practicing air travel by exposing participants to 360-degree videos of airports and airplanes using Google Cardboard in combination



# Virtual Station: Virtual Reality as a Bridge Towards Real Independence in Public Transportation for Autistic Children and Adolescents

Ali Adjorlu<sup>(✉)</sup>, Christian Sonne Rosendahl, Maria Barbro Lanther,  
and Pawe Karman Ahmad

Aalborg University, Copenhagen, Denmark  
adj@create.aau.dk

**Abstract.** Public transportation can be a significant challenge for autistic adolescents, who often face difficulties with navigating complex environments, managing schedules, and interacting with unfamiliar systems. In this short paper, we explore the potential of Virtual Reality (VR) as an intervention to address these challenges. Specifically, we developed a VR simulation of a local S-train station in collaboration with a specialized school, allowing autistic students to practice essential transportation tasks such as checking schedules, purchasing tickets, and navigating platforms in a safe and controlled environment.

Our study employed a quasi-experimental design with pre- and post-training observations to evaluate the effectiveness of the VR intervention. Results demonstrated improvements in students' abilities to understand schedules, recognize transportation signs, and manage check-in systems. Post-training observations also revealed increased time spent navigating real train stations, suggesting greater attentiveness and confidence. Despite logistical challenges and the need for more immersive elements, such as handling crowded environments and interacting with ticket inspectors, our findings highlight the promise of VR in preparing autistic adolescents for independent public transportation.

**Keywords:** Everyday living skills · Virtual Reality · Autism Spectrum Disorder

## 1 Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by a spectrum of symptoms affecting social interaction, communication, and sensory processing. These challenges can significantly impede daily living skills (DLS), such as using public transportation [2, 15].

In Anonymous country, the financial burden of accommodating the transportation needs of students with ASD is substantial, with municipalities spend-

18. Neustaedter, C., Sengers, P.: Autobiographical design: what you can learn from designing for yourself. *Interactions* **19**(6), 28–33 (2012)
19. O’Kane, A.A., Rogers, Y., Blandford, A.E.: Gaining empathy for non-routine mobile device use through autoethnography. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 987–990 (2014)
20. Osterwalder, A., Pigneur, Y.: *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*, vol. 1. John Wiley & Sons, Hoboken (2010)
21. Rapp, A.: Autoethnography in human-computer interaction: Theory and practice. In: *New Directions in Third Wave Human-Computer Interaction*, vol. 2-Methodologies, pp. 25–42. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-73374-6\\_3](https://doi.org/10.1007/978-3-319-73374-6_3)
22. Reis, E.: *The Lean Startup*. Crown Business, New York (2011)
23. Rodríguez-Abad, C., et al.: A systematic review of augmented reality in health sciences: a guide to decision-making in higher education. *Int. J. Environ. Res. Public Health* **18**(8), 4262 (2021)
24. Sangeethaa, S., Jothimani, S.: Blockchain in the metaverse. In: *Cultural Marketing and Metaverse for Consumer Engagement*, pp. 51–70. IGI Global, Hershey (2023)
25. Shanbhag, N., Pardede, E.: The blitz canvas: a business model innovation framework for software startups. *Systems* **10**(3), 58 (2022). <https://doi.org/10.3390/systems10030058>
26. Singh, M.: *Pokemon go changes everything and nothing for arvr* (2016). <https://techcrunch.com/2016/08/12/pokemon-go-changes-everything-and-nothing-for-arvr/>
27. Souchet, A.D., Lourdeaux, D., Pagani, A., Rebenitsch, L.: A narrative review of immersive virtual reality’s ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload. *Virt. Real.* **27**, 19–50 (2023)
28. Thiel, P., Masters, B.: *Zero to One: Notes on Startups, or How to Build the Future*, 1st edn. Currency, New York (2014)
29. Van Der Land, S., Schouten, A.P., Feldberg, F., van Den Hooff, B., Huysman, M.: Lost in space? cognitive fit and cognitive load in 3d virtual environments. *Comput. Hum. Behav.* **29**(3), 1054–1064 (2013)

## References

1. Alford, H.: 3 industries that will be disrupted by pokémon go and ar. Medium (2016). <https://harryalford.medium.com/3-industries-that-will-be-disrupted-by-pok%C3%A9mon-go-and-ar-402f2715306f>
2. Bambušek, D., Materna, Z., Kapinus, M., Beran, V., Smrž, P.: Handheld augmented reality: overcoming reachability limitations by enabling temporal switching to virtual reality. In: Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 698–702. IEEE (2022)
3. Buchner, J., Buntins, K., Kerres, M.: The impact of augmented reality on cognitive load and performance: a systematic review. *J. Comput. Assist. Learn.* **38**(2), 285–303 (2022)
4. Chen, D.: The metaverse is here...but is the hardware ready? Spiceworks (2022). <https://www.spiceworks.com/tech/hardware/guest-article/the-metaverse-is-here-but-is-the-hardware-ready/>
5. Dionisio, J.D.N., Burns, W.G., III., Gilbert, R.: 3d virtual worlds and the metaverse: current status and future possibilities. *ACM Comput. Surv. (CSUR)* **45**(3), 1–38 (2013)
6. Glaser, B.G.: *Theoretical Sensitivity*. Sociology Press, Mill Valley, CA (1978)
7. Glaser, B.G., Strauss, A.L.: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine, Chicago (1967)
8. Gomes, A., et al.: Extended by design: a toolkit for creation of xr experiences. In: Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 57–62. IEEE (2020)
9. Graser, S., Nielsen, L.H., Böhm, S.: Factors influencing the user experience of mobile augmented reality apps: An analysis of user feedback based on app store user reviews. In: Godulla, A., Böhm, S. (eds.) *Digital Disruption and Media Transformation*, pp. 109–128. Springer, Heidelberg (2023). [https://doi.org/10.1007/978-3-031-39940-4\\_9](https://doi.org/10.1007/978-3-031-39940-4_9)
10. Hegde, S.: Defining your customer engagement strategy: Which game are you playing? (2023). <https://amplitude.com/blog/customer-engagement-strategy#defining-your-customer-engagement-strategy-which-game-are-you-playing>
11. Heshmat, Y., Neustaedter, C., DeBrincat, B.: The autobiographical design and long term usage of an always-on video recording system for the home. In: Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 675–687 (2017)
12. Matney, L.: Investments in vr/ar have already hit \$1.1 billion in 2016 (2016). <https://techcrunch.com/2016/03/07/investments-in-vrar-have-already-hit-1-1-billion-in-2016/>
13. Maurya, A.: *Running Lean*, 3rd edn. O’Reilly Media, Inc., Newton (2022)
14. Metinko, C.: Vr/ar investments increase just as metaverse talk heats up—but that may not be the only reason (2022). <https://news.crunchbase.com/startups/metaverse-augmented-reality-virtual-reality-investment/>
15. Montes: What is spatial computing and how is it revolutionizing our world? (2022). <https://www.verses.ai/blogs/what-is-spatial-computing-and-how-is-it-revolutionizing-our-world>
16. Munro, A.J.: Autoethnography as a research method in design research at universities. In: Proceedings of the 20/20 Design Vision Conference, p. 156 (2011)
17. Naess, M.: The metaverse in retail: a game-changer that’s not ready yet (2022). <https://venturebeat.com/datadecisionmakers/the-metaverse-in-retail-a-game-changer-thats-not-ready-yet/>

as blockchains, artificial intelligence, and the Internet of Things [24] continue to shape the metaverse landscape, future research could explore the integration of these advancements into the MIC, enabling a more comprehensive and future-proofed ideation process. Conducting empirical studies to validate the efficacy of the MIC in real-world scenarios could provide quantitative evidence of its impact on the success rates of AR/VR startups.

## 7 Concluding Remarks

This research made several notable contributions to the understudied domain of augmented reality (AR) and virtual reality (VR) entrepreneurship. First, it identified prevalent failure factors based on an in-depth investigation of unsuccessful AR/VR startups, elucidating common pitfalls that can lead ventures in this space to fail.

Understanding these factors can provide valuable lessons for aspiring entrepreneurs seeking to increase their chances of success. Second, grounded in design research methods such as autoethnography, autobiographical design, and insights from industry expertise, this study presented the novel MIC as a strategic ideation toolkit tailored to addressing unique considerations of extended reality ventures.

This specialized canvas guides founders through prompts that encourage the consideration of critical factors such as the user experience, the technological feasibility, the problem-solution fit, and the business viability starting from the earliest stages. Third, the proposed MIC framework prompts a human-centered, problem-driven approach to conceive products and services that leverage the differentiating capabilities of immersive technologies such as AR and VR.

Its tailored blocks push entrepreneurs to ideate solutions that provide value propositions exclusive to XR, going beyond simply replicating non-XR experiences. Fourth, the research methodology synthesized mixed qualitative techniques such as focus groups, content analyses, and external evaluations by startup consultants to produce actionable and valid insights. The triangulation of perspectives enhanced the reliability and applicability of the findings.

Overall, this work combined an insightful analysis of the real-world challenges faced by failed AR/VR startups with an innovative design output—the MIC—aimed at enhancing the viability of future ventures in this space through a tailored, reflective ideation process. The MIC provides a model for human-computer interactions and design researchers seeking to bridge conceptual contributions with practical entrepreneurial impact to drive the adoption of novel technologies.

ogy, suggesting that founders may need the assistance of an XR specialist to accurately complete this field. This highlights the technical knowledge required to address usability concerns and design effective user experiences within the AR/VR domain.

Furthermore, two participants emphasized that the lack of scalability within a short timeframe was a prominent failure factor for two of the startups. One participant pro-posed the addition of a timeline block to the MIC to address this concern. By incorporating a timeline, startups would be prompted to consider and plan for the scalability of their ventures over time, recognizing the potential threats and opportunities that may arise in the future.

Overall, the engagement of the startup consultants in completing the MIC high-lighted key areas of focus, such as the XR value propositions, usability considerations, and scalability.

## 6 Limitations and Future Work

One limitation of this study was the sole reliance on qualitative data from startup consultants to evaluate the MIC. While the feedback of these consultants provided valuable insights, incorporating quantitative metrics to objectively measure key factors such as the XR value propositions, usability considerations, and scalability could further strengthen the evaluation. Future research should explore the development and validation of quantitative instruments to assess these critical elements within the context of XR business modeling. By combining qualitative expert feedback with quantitative measurements, a future study could triangulate the findings and provide a more comprehensive understanding of the canvas’s efficacy in enhancing the viability of metaverse startups.

Another limitation of this study was the lack of an analysis on the success factors for thriving AR/VR startups. While our investigation of failed ventures provided valuable insights into common pitfalls, incorporating an examination of successful cases could further enrich the findings and provide a more comprehensive understanding of the critical factors influencing a startup’s success in the AR/VR domain. Future research could extend this work by analyzing prosperous AR/VR startups and integrating the identified success factors into the MIC framework.

The analysis of market dynamics and user adoption patterns in the AR/VR industry may have been oversimplified in this study. A more in-depth exploration of these factors could yield additional insights and nuances that were not captured in the current scope of this research. The data collection for this study was conducted in the previous year, and may not have fully captured the rapidly evolving nature of the AR/VR industry. As new technologies and market trends emerge, the identified failure factors and the applicability of the MIC may require further validation and refinement.

Expanding the research scope to include AR/VR startups from diverse cultural and geographical contexts could enhance the understanding of region-specific factors that influence startup success or failure. As new technologies such

**Table 1.** Analysis of five startups

Category	Start Year	Stop Year	Summary
Startup #1 Shopping	Q3 2022	Q2 2022	This startup was an AR android application for buying home decor. The revenue model was based on selling commissions. The app included a Web back end, and shop owners were able to list their products. The shop included more than 1000+ products and achieved 100,000+ installations
Startup #2 Shopping	Q2 2019	Q2 2021	This startup was a VR app with 10,000+ installations that simulated shopping malls. The app's revenue model relied on selling subscriptions to shop owners. The users needed to have Google Cardboard VR to use the app
Startup #3 Social Network	Q2 2018	Q3 2021	This startup was an AR-only location-based social network that allowed people to meet other people around them and evaluate. It crossed 5K users. The app's revenue model relied on selling ad-spaces for sending AR gifts and AR. The app was available for phones, tablets and never models, and it was installed on more than 100,000 devices
Startup #4 Gaming	Q1 2018	Q1 2022	This startup developed a location-based AR mobile game available on iOS and Android, generating 200,000+ installation. Revenue was generated through in-app purchases. Despite a successful launch, the game faced a swift decline in user retention. The gameplay required physical movement, incorporating the player's actual locomotion as a core mechanic
Startup #5 Productivity	Q3 2017	2021	This app was released for iPhone and iPad, and most features required the constant use of AR mode. The app aimed to facilitate the organization of to-do lists by augmented reality by animating 3D objects. The users were able to use the app for free and it crossed 20,000 installations

modification in expanding the possibilities for innovations within the XR space. When the designers were queried about the usability blocks, there was a general consensus among the participants that these blocks are crucial for effectively communicating ideas prior to the development phase. However, they also noted that filling this section may require a solid understanding of AR/VR technol-

The “existing scalability limitations” and “future threats” counter the myth and difficulty of short-term planning by pushing founders to forecast the obstacles to long-term growth. “Future opportunities” guides the consideration of how evolving technology and trends can enable the XR solution to be scaled over time.

#### 4.7 Measurable Metrics

To account for the critical role of usability in the profitability of XR products and services, we incorporated predefined metrics into the MIC. These metrics included the expected usable session (minutes), representing the anticipated duration of each user session; the expected daily engagement for each user (sessions), capturing the projected frequency of user engagement; and the expected revenue per user, providing in-sights into the anticipated revenue generated from individual users.

## 5 Testing the Model with Startup Experts

This study included an analysis of the five startups presented in Table 1: two developed a shopping application, another specialized in gaming, one ventured into social networking, and the final one concentrated on productivity solutions.

The consultants had no prior knowledge of the specific reasons behind the failures of these startups. They were provided with the summary, main usage scenario, and start and stop dates of these ventures.

A common theme emerged among all the participants that emphasized the significance of XR value propositions. It was collectively observed that three out of the five failed startups did not offer value propositions that leveraged the unique capabilities of XR systems, missing out on the potential benefits that the technology could provide. This highlights the importance of creating value propositions that are exclusive to XR experiences in order to fully capitalize on the potential of the technology.

The shopping app, VR social network, AR social network, and productivity app startups faced challenges that justified the value of their solutions against the cognitive and physical demands placed on the users. All the participants emphasized the need to justify motion-based interactions. One consultant suggested that XR startups should avoid brief, buzzword-focused pitches centered on being “AR/VR,” and instead clearly articulate their unique value proposition. For example, the shopping app only implemented an AR display feature. As its competitors had integrated similar in-room display capabilities, this startup lacked a defensible advantage. One participant stated that pointing out future threats prompted the founders to think beyond the short-term trends and consider sustainable differentiation.

One of the participants expressed appreciation for the modification made to the existing alternative block, recognizing its ability to stimulate ideation and inspire new ideas during the design process. They acknowledged the value of this

- **VR UX opportunity:** How does this solution help users in a fully immersive environment?
- **Social opportunity:** How does this solution create value for users via social interactions?
- **Virtual economy opportunity:** How does using the virtual world economy paradigm create value for customers?
- **Interoperability features:** How does the solution adopt interoperability or achieve the interoperability needs of users?
- **Minimum viable experience scenario:** This is a written form of a user storyboard that explains an ideal usage story in a short manner, which may include all or some pieces of the usability elements.

The **motion-based interaction load** addresses the mistake of deferring usability considerations by prompting the upfront justification of motion interfaces versus physical/cognitive burdens. The “AR UX opportunity” and “VR UX opportunity” guide the ideation of AR and VR use cases distinctly. The “social opportunity” and “virtual economy opportunity” promote the consideration of the novel social and economic models afforded by XR worlds. The “minimum viable experience scenario” narrative description helps convey and refine the envisioned user experience holistically.

#### 4.5 Viability Elements

The viability blocks of the MIC encompass essential elements such as channels, costs, and revenue streams, which were derived from the Lean Canvas framework. These blocks allow startups to strategically plan and analyze their distribution channels, estimate the costs associated with their metaverse ventures, and identify potential revenue streams. These elements inherited from the Lean Canvas address the difficulties around premature scaling by starting with a minimally viable offering before expanding.

#### 4.6 Future Scalability

The future scalability block within the MIC includes various crucial aspects, including the existing scalability limitations, future threats, and future opportunities.

By considering these elements, founders can assess the potential for scaling their metaverse ventures over time. This block enables startups to identify and understand any existing limitations or challenges that may hinder their scalability, allowing them to proactively strategize and overcome such obstacles. Additionally, by considering future threats and opportunities, startups can anticipate the market dynamics, emerging technologies, and changing user preferences that may impact the scalability of their metaverse solutions. This forward-thinking approach empowers startups to position themselves for future growth, adapt to evolving trends, and seize new opportunities within the rapidly evolving metaverse landscape.

tasks or areas of interest. Solutions in this category address this need by offering value in the form of information, entertainment, or self-expression. Netflix, social media, and video games are examples of solutions that address focus drive problems [10]. The myths around only considering digital/XR alternatives and the thought that bringing an app to XR is a value proposition are addressed by separating “existing XR alternatives” from “existing non-XR alternatives” and “existing non-digital alternatives.” This prompts founders to consider a broader set of solutions beyond just XR. The difficulty in defining the core problem is mitigated by the “user engagement goal” checkboxes that guide problem framing in terms of the focus drive, productivity, or transaction needs.

### 4.3 Solution

The concept of a unique value proposition has been adapted to the context of XR, resulting in the XR unique value proposition. This shift is essential for startup designers to recognize and emphasize the value propositions that users can exclusively experience through XR solutions, distinguishing them from non-XR alternatives.

In the XR landscape, it is crucial to identify and leverage the immersive and trans-formative qualities that XR technologies offer. Startups need to go beyond replicating existing non-XR solutions and instead focus on harnessing the unique capabilities of XR to provide novel and engaging experiences that cannot be achieved through traditional means.

The “XR unique value proposition” block directly counters the myth that simply bringing something to XR is valuable by pushing founders to articulate the value enabled specifically through XR capabilities.

### 4.4 Usability

We aimed to minimize the usability ideation in a text-based manner by creating the following blocks. Filling all the fields is not mandatory for each startup, but it can give a quick overview about the position of a startup in the metaverse landscape. Motion-based interaction load:

The motion-based interaction load encompasses the physical and cognitive demands imposed on users as they interact with XR experiences using gestures and body movements. The value proposition of motion-based interactions lies in their ability to provide a more immersive and intuitive user experience, enhancing engagement and enabling new possibilities in fields such as training simulations and gaming. However, it is crucial to strike a balance between the benefits and the associated load, considering factors such as user comfort and the learning curve. By justifying their value proposition and addressing challenges, startups can leverage motion-based interactions to create compelling XR experiences that captivate users and drive the adoption of this transformative technology.

- **AR UX opportunity:** How does this solution help users in a semi-immersive environment?

use face computers such as AR/VR HMDs for multiple hours. These considerations directly im-pact the sustainability of the business model.

In response to these concerns, we developed a one-page business plan template specifically tailored for XR startups, known as the Metaverse Innovation Canvas, as depicted in Fig. 1. The MIC comprises five distinct types of blocks, with each serving a unique purpose: problem (red blocks), solution (green blocks), usability (blue blocks), viability (yellow blocks), and future scalability (purple blocks).



Fig. 1. Metaverse Innovation Canvas

## 4.2 Problem

The problem section includes blocks that describe a problem and the people who have the problem. In order to help designers and founders define their problems more effectively, we added user engagement goal checkboxes to the problem section, which included focus drive, productivity, and transaction. We adopted this from an article about the Three Games of Customer Engagement Strategy [10].

Focus drive represents users’ inherent need to direct their attention and concentration towards specific subjects or activities. It reflects individuals’ natural motivation or inclination to concentrate their mental resources on particular

#### 4.1 Myths, Difficulties and Mistakes

A common mistake during idea development is only considering digital and XR alternatives, while neglecting non-digital options. For example, when conceptualizing an XR shopping app, the founders may focus solely on existing digital solutions such as online shopping websites. However, they should also consider brick-and-mortar stores as a non-XR alternative that users may currently utilize. The tendency is to view the landscape strictly within the XR application space, but founders need to look beyond this and analyze all potential digital and non-digital alternative solutions to the problem they are trying to solve.

Bringing an application to XR is a value proposition. This is wrong; it might provide an early blessing in an untapped market, but XR brings extra costs and additional physical and even cognitive loads [29] for users, so these extra costs need to be justified by providing value propositions for end users. Pokémon Go motivated [26] its users to move outside and follow location-based interaction practices because the users were collecting Pokémon cards before that. Niantic did not need to justify this extra load for the users, since years of marketing by Nintendo had already accomplished this for them.

Problem definition is the toughest job for every entrepreneur and product designer, and the problem is further complicated by the fact that most people do not have clear understanding of what can be considered a problem. For example, the case of Angry Birds may not immediately seem relevant to the problem-solution definition. However, the main issue that Angry Birds addresses is people's need to maintain their attention on something, and this game accomplishes this in an entertaining and user-friendly manner.

The primary aim of the Lean Canvas is to help founders validate their idea as soon as possible, and for AR/VR startup founders, it is ideal to refine their ideas into viable, desirable, and feasible solutions. However, many AR/VR solutions struggle to achieve feasibility or scalability in a short time due to the lack of advancements in enabling technologies. Therefore, the perfect business plan for AR/VR startups should consider the impacts of predictable future market events.

The entrepreneurs have demonstrated two common mistakes in designing startups: Large customer segment, and lack of attention to usability considerations:

Entrepreneurs tend to expand their customer segments and write more items in this segment; it may sound more interesting for some investors because it can increase the size of their market and the valuation of the company, but it can be a fatal mistake in the XR industry. This is because it increases the development costs exponentially and will postpone idea validation for a long period.

It would be wise to limit the early adopters and find a niche market for the minimum viable products. Thiel and Masters [28] argued that only monopoly profits can help a business to survive, since a monopoly provides the profits to continue the innovation.

Postponing the usability considerations until the minimum marketable product is reached. UX issues in XR environments are far more complex than screen-based applications. Handheld AR can cause hand fatigue [3], and users cannot

our proposed MIC provides specific fields to prompt the consideration of these critical factors.

The experiences of 14 of the startups imposed high physical and cognitive loads on their users through excessive motion-based interactions. The benefits did not outweigh the burdens of these demanding natural user interface requirements.

An AR productivity startup failed after the users reported physical strain from the extended motion-tracking required for 3D modelling. An App Store review stated: “It’s innovative but using hand gestures for more than 10 min is very tiring. I can’t see people using this daily.” The startup could not justify the interaction load versus mouse/keyboard inputs.

### 3.5 Failure to Address a Clear Problem

Successful startups often solve tangible problems or fulfill specific needs. However, some AR/VR startups failed to address a clear problem or failed to effectively communicate the problem they were solving. This lack of focus on a defined problem reduced the perceived value of their offerings and hindered their ability to attract and retain customers.

A total of 11 of the analyzed startups did not have a well-defined central problem that their AR/VR solution effectively solved for users. A lack of focus on addressing real user needs undermined their perceived value.

A VR startup for remote communication had no clear user workflow or pain-point that their solution addressed better than video calls. As one reviewer stated: “It’s novel tech but I don’t understand when I would actually use this instead of a Zoom call.”

By recognizing these key failure factors, future AR/VR startups can learn from the mistakes of their predecessors and take proactive measures to mitigate these challenges. Addressing issues related to long-term planning, hardware limitations, usability, the value proposition, motion-based interactions, and problem-solving can significantly increase the chances of success in the competitive AR/VR landscape.

## 4 Development of the Metaverse Innovation Canvas

The Lean Canvas, as a perfect one-page business plan template for brainstorming business ideas, was introduced in the Running Lean book [13] based on an extension of the Business Model Canvas [20], and it has helped a countless number of startups around the globe. However, since the Lean Canvas [22] is a general-purpose framework, entrepreneurs may not use it correctly. The following is a list of common myths, mistakes, and difficulties that AR/VR startup founders usually encounter when trying to deconstruct their business assumptions with the Lean Canvas.

immersive technologies, their work stopped short of outlining actionable strategies tailored to the practical context of XR startups and product development. Specifically, their review did not translate their insights on ergonomic pitfalls into a coherent framework for mitigating issues starting from the early stages of business modeling and design. Our research built upon their analysis by proposing the MIC as an entrepreneurial toolkit for embedding a consideration of critical usability and user experience (UX) factors directly into the ideation process. While [23] systematic review highlighted the usability and physical burden issues with augmented reality in health sciences education, their work did not provide an actionable framework for addressing these concerns during business ideation for AR/VR ventures. Our proposed MIC aims to bridge this gap by directly incorporating human-centered experience scenarios and prompting the justification of motion-based interactions from the start.

A total of 17 of the failed startups exhibited poor usability of their AR/VR offerings, with unintuitive or cumbersome user experiences that diminished the value proposition. Additionally, 16 of them failed to clearly articulate a compelling value proposition that justified the use of AR/VR over non-immersive alternatives.

An AR home decor app received widespread complaints about poor usability, with a 2.1/5 rating on the App Store. The users cited issues such as “keeps crashing” and “difficult to place objects naturally.” Without a compelling value-add over browsing decor websites, the app struggled with engagement.

### 3.4 Motion-Based Interactions and Physical Load

Some failed AR/VR startups struggled with motion-based interactions and physical load. While these startups offered innovative motion-based experiences, the physical demands and discomfort associated with prolonged use did not justify the value propositions for users. The disparity between the expected advantages and the physical burden on users eroded the feasibility and attractiveness of their offerings.

Although the Extended by Design Toolkit [8] acknowledges physical and perceptual strain as challenges in XR, it lacks constructs prompting developers to justify motion-based UX choices versus less-demanding interaction methods. Our canvas guides teams to weigh the benefits and burdens of natural user interfaces while considering the long-term feasibility. Although the review by Buchner et al. [3] described the physical burdens associated with augmented reality, their analysis lacked guidelines for mitigating these demands during design and development. Studies have shown that demanding augmented reality experiences increase the strain, reducing the comfort and performance over time [2]. Similarly, research has found significant fatigue from the extended use of augmented reality glasses compared to mobile devices, limiting their feasibility for daily tasks. Similar to the usability issues, the problems with motion-based interactions and physical load for AR/VR systems [9] are not being accounted for in current generalized startup planning models such as the Lean Canvas, whereas

contributed to a deeper understanding of the reasons behind the failures of these startups. The following factors emerged as significant contributors to the failure of AR/VR startups:

### 3.1 Short Timeframe and Lack of Future Planning

Many AR/VR startups faced difficulties due to a short-sighted approach and a lack of clear plans for future trends. The rapidly evolving nature of the AR/VR industry demands foresight and adaptability. Startups that failed often lacked the ability to anticipate market shifts, technological advancements, and changing consumer preferences, ultimately leading to their downfall.

This factor was observed in 18 out of the 29 failed startups analyzed. Many of these companies focused on short-term trends or gimmicks without accounting for longer-term market shifts or technological advancements.

### 3.2 Lack of Scalability and Hardware Limitations

The lack of scalability in the AR/VR industry can be attributed to the limited availability of AR/VR head-mounted displays (HMDs), preventing the creation of a mass market service. The shortage of AR/VR HMDs hampers the scalability of startups, as there is not yet a sufficient consumer base to support widespread adoption. The hardware limitations and the industry's challenge in providing affordable, accessible, and high-quality devices have further hindered the scalability of AR/VR startups.

A total of 22 of the 29 startups struggled with scalability issues stemming from the limited availability and high cost of AR/VR headsets and other enabling hardware during their active years. Their solutions could not achieve mass consumer adoption due to these hardware constraints.

Quantifying this constraint, only 8.1 million AR/VR headsets were shipped globally in 2018 (source: IDC Worldwide Quarterly AR/VR Device Tracker). With limited device penetration and high costs, such as USD 599 for the Oculus Rift, 15 of the 29 failed startups (52%) directly cited hardware availability and pricing as a key factor hampering the scalability of their AR/VR solutions.

### 3.3 Lack of Usability and Value Proposition

Usability plays a crucial role in the success of AR/VR startups. The startups that neglected the importance of intuitive and user-friendly experiences faced significant challenges. Additionally, the lack of a compelling value proposition, considering non-AR/VR alternatives, contributed to the failure of some startups. When users could find similar solutions outside the AR/VR environment that were more accessible, more cost-effective, or easier to use, the value proposition of the AR/VR startups became diminished.

While [27] comprehensively reviewed the risks of visual fatigue, cybersickness, and other physical discomforts that undermine the usability and value of

in advising early-stage ventures and were identified by searching the websites of startup accelerators in Iran. Each consultant was presented with the Lean Canvas and background information for the five failed AR/VR startups identified in stage 1. They were then asked to fill out the MIC for each startup.

These consultants were asked to evaluate the usability and feasibility aspects of the MIC by reviewing the startup cases and providing feedback. Their insights from an expert perspective contributed to assessing the potential and usability of the proposed framework. Semi-structured interviews were conducted to gather feedback on the usefulness and usability of the canvas. The sessions were audio-recorded and transcribed.

It is important to note the limitations of this approach. We did not have access to proprietary internal information for Startups 2, 3, and 4, since we did not interview their founders directly. Our perspectives on these startups were based on publicly available information, which may have lacked key insights. There was also an element of subjectivity in how we interpreted and synthesized the data on the startups based on our own prior knowledge and experiences.

Furthermore, the participants in this study had access to limited information about the startups and reviewed this information in a compressed timeframe. During the MIC testing session, three startup consultants were brought in to analyze and complete the canvas for five different failed AR/VR startups.

## 2.4 Analysis Method

The analysis of the collected data in this study was guided by Glaserian grounded theory methodology, which is a systematic approach for generating a theory from empirical data [6, 7]. This methodology recommends the constant comparative method, where data collection and analysis occur simultaneously, and the researcher continuously compares incidents, codes, and categories to construct a coherent theoretical framework grounded in the data [7].

## 2.5 Ethical Considerations and Limitations

Ethical considerations were carefully addressed throughout the research process. Confidentiality and anonymity were maintained when referencing specific startup cases or individuals within the startup studio. Informed consent was obtained to protect individuals' identifiable information and prevent potential harm to their reputation during the research. After the analysis, the collected data were destroyed. The authors maintained reflexivity throughout the study by acknowledging their positional ties to the startup community and the potential influence it may have had on their observations and interpretations.

## 3 Key Failure Factors of AR/VR Startups

During the analysis of AR/VR startups, several key failure factors were identified, shedding light on the challenges that led to their downfall. These factors

Additionally, we utilized online startup directories such as Crunchbase, AngelList, and local startup groups to identify relevant AR/VR ventures operating within the specified timeframe of 2016–2022. We then contacted the founders or leadership teams of these startups, explaining the purpose of our research and requesting their participation by sharing relevant business documentation and data. For startups that had ceased operations, we attempted to contact former employees or stakeholders to obtain access to archival materials. Throughout this process, we ensured the anonymity and confidentiality of the participating individuals and companies.

## 2.2 Stage 2: Autoethnography and Autobiographical Design

The objective of this stage was to develop a startup business-modeling framework that aimed to mitigate the mistakes made by entrepreneurs when using the Business Model Canvas (BMC) and the Lean Canvas, which have been identified as contributing factors to the failure of AR/VR startups. By recognizing and addressing the key failure factors identified in Sect. 3, the MIC framework sought to provide entrepreneurs with a comprehensive and effective tool for strategically planning their AR/VR startups, thus increasing their chances of success in the competitive extended reality industry. Autoethnography has gained popularity in the field of human-computer interactions in recent years [11, 19, 21].

This method is also employed in universities for conducting design research, as it allows for the collection of “creative and innovative processes” experienced by designers [16]. Autobiographical design, a form of design research, utilizes autoethnography primarily for the creation of solutions that are intended to be used extensively by their creators [18].

The primary source of data for this research included the authors’ personal and extensive experience with the BMC [20] and the Lean Canvas [22]. We provided consulting and design services to startups and utilized strategic design toolkits such as BMC and its adaptations. Additionally, as the strategic designers and heads of a startup studio, we shared firsthand accounts and insights gained from extensive experience working with AR/VR startups. We conducted autobiographical design interviews to gather rich and personal narratives, focusing on the challenges, successes, and lessons learned in the context of AR/VR startup ventures.

In addition, the involvement of other authors in this research played a crucial role in facilitating ethnography and enabling data triangulation. The authors collaborated in conducting a comprehensive literature review and netnography of the startup community on Twitter, which served to triangulate the findings. This cross-validation approach enhanced the reliability and accuracy of identifying the factors contributing to failure in this study.

## 2.3 Stage 3: Evaluation

In the third stage, the MIC was evaluated through expert testing with three startup consultants. These participants were recruited based on their expertise

## 2 Methodology

This research consisted of three main stages. In the first stage, data from domestic and international AR/VR startups were collected and analyzed to identify the key factors contributing to their failure. Four focus group sessions were conducted to discuss the data, and the factors were coded based on the analysis.

In the second stage, the authors applied autoethnography [21] and autobiographical design [18] to create an innovation tool for startup founders. By leveraging their more than a decade of experience in entrepreneurship, the authors developed the MIC. This canvas serves as a framework for investors and entrepreneurs in the metaverse field. Finally, in the third stage, we recruited experts to evaluate five failed startups with the MIC and interviewed them about the MIC's potential for use in the AR/VR industry.

### 2.1 Stage 1: Analyzing the AR/VR Startups

The first stage of the research involved analyzing AR/VR startups. The first author's startup studio actively monitored the AR/VR industry and collected data on startups operating in various sectors, including content creation, gaming, social media, education, healthcare, tourism, and retail. The data spanned from 2016 to 2022, with 2016 being a significant milestone in AR/VR history due to the commercial success of Pokémon Go, which validated the XR market and led to the widespread adoption of computer-mediated reality technology [4].

The collected data included the Lean Canvas of the startups, which was developed based on the principles outlined in [20]. Additionally, the data encompassed the logs of usability testing sessions, led by the first author. To filter out failed startups, the authors examined the recent status of these ventures and identified 29 startups as failures. Participants were recruited through searching on LinkedIn. In order to reduce the bias, the participants were not recruited from the connections of the authors.

To evaluate the reasons behind the failure of these startups, the authors conducted focus group sessions. There were five group sessions in total, and four participants were recruited through searching on LinkedIn. In order to reduce the bias, the participants were not recruited from the first connections of the authors. These focus group sessions were conducted with the same participants. The sessions took between 60 and 90 min and were conducted in Q1 2023.

The analysis was based on the collected data, the current status of the startups obtained from the Crunchbase directory, and the application of stress-testing desirability [13]. This qualitative test focused on evaluating the problem rather than the solution, providing valuable insights into the marketability of the product or service. Finally, through a content analysis and codification, the key themes of AR/VR startup failures were identified.

The selection of AR/VR startups for the analysis was conducted through multiple channels. First, we leveraged the authors' extensive network within the startup ecosystem, reaching out to companies they had previously advised or collaborated with.

acknowledges challenges such as physical fatigue and increased cognitive load, it lacks specific constructs prompting developers to justify motion-based user experience choices versus less-demanding interaction methods. Specialized business model frameworks such as the Blitz Canvas [25] have been proposed for specific domains like software startups, but there remains a need for tailored tools addressing the unique considerations of extended reality ventures.

Successful implementations of AR/VR have the potential to revolutionize user experiences through immersive gaming, intuitive 3D designs, interactive remote collaboration, digitally-enhanced shopping environments, and many other applications. However, achieving sustainable, scalable ventures in this space requires addressing key issues such as hardware limitations, creating compelling XR-exclusive value propositions, justifying physical/cognitive interaction loads, and planning for interoperability and ecosystem forces [17].

This has motivated the need for tailored business modeling and design tools that can guide entrepreneurs in ideating viable, usable, and scalable AR/VR products and services by embedding these critical considerations starting in the initial stages.

This research proposed the MIC, a novel framework for developing XR products and services. The MIC addresses the limitations of existing tools such as the Lean Canvas in capturing unique considerations for XR ventures. The key advantages of the MIC include the following: 1) emphasizing problem solving, with blocks dedicated to framing the problem and articulating XR-exclusive value propositions; 2) incorporating usability metrics and scenario writing to communicate user experiences; 3) considering scalability issues specific to XR, such as hardware constraints; and 4) guiding ideation about the social interactions and virtual economies afforded by XR. Overall, the MIC provides a holistic, human-centered approach to ideation, enhancing the viability of XR startups by bridging the gaps between design and business modeling. Its tailored blocks prompt entrepreneurs to leverage the distinctive capabilities of immersive technologies. This research made several notable contributions to the understudied domain of augmented reality (AR) and virtual reality (VR) entrepreneurship. First, it identified prevalent failure factors based on an in-depth investigation of unsuccessful AR/VR startups, elucidating common pitfalls.

Second, grounded in design research and industry expertise, it presented the novel MIC as a strategic ideation toolkit tailored to addressing the unique considerations of extended reality ventures. Third, the proposed framework prompted a human-centered, problem-driven approach to conceiving products/services that leverage the differentiating capabilities of immersive technologies. Fourth, the methodology synthesized mixed qualitative techniques and an external evaluation to produce actionable, valid insights. Overall, this work combined an insightful analysis of real-world challenges with an innovative design output, enhancing the viability of future AR/VR startups through a tailored, reflective ideation process. It provided a model for human-computer interactions and design researchers seeking to bridge conceptual contributions with practical entrepreneurial impact.

The emergence of AR/VR applications, services, and tools has had a profound impact on shaping the metaverse into a dynamic and immersive digital realm. AR applications overlay digital information onto the real world, enriching our physical environment with virtual elements, while VR applications create entirely simulated environments.

Over the past few decades, there has been a significant inflow of billions of dollars in investments into the AR/VR startup ecosystem [12, 14]. However, despite the substantial financial backing, a considerable number of these startups have faced insurmountable challenges and have ultimately failed to thrive. This trend piqued our curiosity and motivated us to embark on an in-depth exploration of the underlying reasons behind the failure of these companies. While it is commonly acknowledged within the industry [1, 17] that technological and hardware limitations have played a pivotal role in these setbacks, we recognized the importance of delving deeper into the diverse array of factors that may have contributed to these failures. Thus, our research aimed to unravel the complexities surrounding the demise of AR/VR startups, going beyond the prevailing notion of technological constraints.

Through an exhaustive investigation, we sought to broaden the understanding of the dynamics at play within the AR/VR startup landscape. Our analysis encompassed a comprehensive examination of market dynamics, user adoption patterns, funding challenges, regulatory hurdles, and competition dynamics. Through a meticulous examination of these multifaceted factors, we aimed to attain nuanced insights into the intricate challenges confronted by these startups.

By leveraging the insights obtained from our research, we moved beyond a theoretical understanding to develop a practical solution. This solution aimed to assist founders and investors in refining their business ideas within the distinctive context of extended reality (XR) environments. While the Lean Canvas serves as a useful starting point for business model development, our analysis identified critical gaps when applying it to augmented and virtual reality ventures, including an inadequate emphasis on validating the problem-solution fit and accounting for the user MIC, an innovative tool designed to facilitate the iterative refinement of business concepts. The MIC serves as a structured framework for founders and investors, enabling them to identify and address the specific needs and intricacies of XR environments. By leveraging this canvas, entrepreneurs can systematically explore various dimensions such as the user experience, the technological feasibility, the market viability, monetization strategies, and the scalability potential within the metaverse ecosystem.

Our approach of examining failed startups specifically aimed to avoid survivorship bias, a common pitfall in entrepreneurship research where only successful cases are studied, as highlighted in [28]. Previous research has highlighted potential pitfalls in AR/VR adoption, such as visual fatigue, cybersickness, muscular fatigue, acute stress, and mental overload [27]. However, these studies have primarily focused on examining ergonomic and physiological risks rather than providing actionable strategies tailored to the practical context of XR startups and product development. Similarly, while the Extended by Design Toolkit [8]



# Metaverse Innovation Canvas: A Tool for Extended Reality Product/Service Development

Amir Reza Asadi<sup>1</sup>, Mohamad Saraee<sup>2</sup>, and Azadeh Mohammadi<sup>2</sup>

<sup>1</sup> School of Information Technology, University of Cincinnati,  
Cincinnati, OH 45221, USA  
asadiaa@mail.uc.edu

<sup>2</sup> School of Science, Engineering and Environment, University of Salford,  
Manchester M5 4WT, UK  
{m.saraee, a.mohammadi1}@salford.ac.uk

**Abstract.** This study investigated the factors contributing to the failure of augmented reality (AR) and virtual reality (VR) startups in the emerging metaverse landscape. Through an in-depth analysis of 29 failed AR/VR startups from 2016 to 2022, key pitfalls were identified, such as a lack of scalability, poor usability, unclear value propositions, and the failure to address specific user problems. Grounded in these findings, we developed the Metaverse Innovation Canvas (MIC) a tailored business ideation framework for XR products and services. The canvas guides founders to define user problems, articulate unique XR value propositions, evaluate usability factors such as the motion-based interaction load, consider social/virtual economy opportunities, and plan for long-term scalability. Unlike generalized models, specialized blocks prompt the consideration of critical XR factors from the outset. The canvas was evaluated through expert testing with startup consultants on five failed venture cases. The results highlighted the tool's effectiveness in surfacing overlooked usability issues and technology constraints upfront, enhancing the viability of future metaverse startups.

**Keywords:** Lean Canvas · Mixed Reality · AR · VR · Innovation · Entrepreneurship

## 1 Introduction

The concept of the metaverse has been around for decades, but it has been the subject of renewed interest and excitement in recent years as technology has advanced and virtual reality has become more accessible. The metaverse can be understood as an integrated network of spatial realities and virtual worlds, where users can engage with each other and digital elements within a three-dimensional environment, often utilizing personalized avatars [5, 15].

# **Using Immersive Technologies for Learning, Accessibility, and Technological Innovation**

Exploring Phantasmagorical Game Design: Integrating Anomalous  
Behaviour of AI and Player Into Conversational AI-Native Games ..... 120  
*Tobias Wissigkeit, Daniela Lope-Nicholls, Christian Stindl,  
and Ido A. Iurgel*

Learning Formats for Theory Courses in Graphic Design Undergraduate  
Programmes ..... 136  
*Christof Breidenich and Anna Unterholzner*

**Author Index** ..... 163

# Contents

## Using Immersive Technologies for Learning, Accessibility, and Technological Innovation

Metaverse Innovation Canvas: A Tool for Extended Reality Product/Service Development .....	3
<i>Amir Reza Asadi, Mohamad Saraei, and Azadeh Mohammadi</i>	
Virtual Station: Virtual Reality as a Bridge Towards Real Independence in Public Transportation for Autistic Children and Adolescents .....	22
<i>Ali Adjorlu, Christian Sonne Rosendahl, Maria Barbro Lanther, and Pawe Karman Ahmad</i>	
Utilizing Virtual Reality to Enhance Situational Awareness in Swedish Police Training .....	33
<i>Hampus Hansen, Måns Holmgren, Carl Johan Gribel, Joakim Ingrell, and George Palamas</i>	
Exploring VR as a Method to Study User Perceptions of Challenges in Green Transition Technology in Nature .....	51
<i>Petra Nurmela, Juri Etto, and Jonna Häkkinen</i>	
Unlocking Cultural Heritage: The Gamified Digitisation Project of SMA-UniGe .....	69
<i>Lara La Tessa, Mauro Coccoli, Stefano Schiaparelli, and Daniele Zolezzi</i>	
<b>Engaging Learners Through Gamification, Playful Design, and Generative AI</b>	
A Case Study in Playful Facilitation for Designing Embodied Experiences .....	79
<i>Jose Font, Mikaela O’Byrne, Alberto Alvarez, and Raquel Robinson</i>	
An Auto Battler Game to Engage High School Student’s Understanding of Ionic Bonds in Chemistry .....	91
<i>Thomas Bjørner, Jacqueline Tully, Tobias Dalsgaard Larsen, Mikkel Andreas Frederik Lau Larsen, Mikkel Sloth Lauridsen, Tony Thai Do, and Daniel Skau Rossing</i>	
Generative AI in Foreign Language Education: Student Use and Perspectives .....	104
<i>Kristine Bundgaard and Anders Kalsgaard Møller</i>	

Henrik Kasch	University College of Southern Denmark, Denmark
Jeanette Sjöberg	Halmstad University, Sweden
Judit Hahn	University of Jyväskylä, Finland
Katrine Ramsgaard	Aalborg University, Finland
Kristine Bundgaard	Aalborg University, Denmark
Lotte Stausgaard Skrubbeltrang	University College of Northern Denmark, Denmark
Maja Højslet Schurer	Aalborg University, Denmark
Marie Charlotte Lyngbye	Aalborg University, Denmark
Susanne Dau	University College of Northern Denmark, Denmark
Sylvana Sofkova Hashemi	University of Gothenburg, Sweden
Thomas Bjørner	Aalborg University, Denmark
Thomas Kjærgaard	University College of Northern Denmark, Denmark
Tobias Alexander Bang Tretow-Fish	University College Absalon, Denmark

## **Local Chair**

Maria Vesterlund Halmstad University, Sweden

## **Publicity and Social Media Chair**

Eva Brooks Halmstad University, Sweden

## **Publications Chair**

Eva Brooks Aalborg University, Denmark

## **Web Chair**

João Martinho Moura Polytechnic Institute of Cávado e Ave, Portugal

## **Posters and PhD Track Chairs**

Emma Edstrand Halmstad University, Sweden  
Maja Højslet Schürer Aalborg University, Denmark

## **Workshop Chair**

Jeanette Sjöberg Halmstad University, Sweden

## **Technical Program Committee**

Anders Kalsgaard Møller Aalborg University, Denmark  
Anders Melbye Boelt Aalborg University, Denmark  
Anthony Lewis Brooks Aalborg University, Denmark  
Camilla Finsterbach Kaup University College of Northern Denmark,  
Denmark  
Emma Edstrand Halmstad University, Sweden  
Eva Brooks Aalborg University, Denmark and Halmstad  
University, Sweden

# Organization

## Steering Committee Chair

Eva Brooks  
Aalborg University, Denmark and Halmstad  
University, Sweden

## Steering Committee Members

Emma Edstrand  
Thomas Bjørner  
Halmstad University, Sweden  
Aalborg University, Denmark

## Organizing Committee

### General Chairs

Emma Edstrand  
Anders Kalsgaard Møller  
Thomas Bjørner  
Halmstad University, Sweden  
Aalborg University, Denmark  
Aalborg University, Denmark

### TPC Chairs

Anders Kalsgaard Møller  
Emma Edstrand  
Thomas Bjørner  
Camilla Finsterbach Kaup  
Aalborg University, Denmark  
Halmstad University, Sweden  
Aalborg University, Denmark  
University College of Northern Denmark,  
Denmark

## Sponsorship and Exhibit Chair

Maria Vesterlund  
Halmstad University, Sweden

impacts from emerging technologies such as AR, VR, and AI on the future of collective dreaming.

The paper presentations addressed new dimensions and key challenges and provided critical and innovative perspectives of employing design theories and tools to develop and implement future digital learning and innovation practices. This was reflected, among others, by the paper which received the Best Paper Award of EAI DLI 2024, titled *A Case Study in Playful Facilitation for Designing Embodied Experiences* by Jose Font, Alberto Alvarez, Raquel Robinson, and Mikaela O'Bryan, all from Malmö University in Sweden – congratulations on receiving the award!

The collaboration between the General Chairs, Emma Edstrand, Anders Kalsgaard Møller, and Thomas Bjørner, and the Technical Programme Committee Chairs, Anders Kalsgaard Møller, Emma Edstrand, Thomas Bjørner, and Camilla Finsterbach Kaup, was essential for the successful planning and performance of the conference. We sincerely appreciated the coordination with Veronika Kissova, Senior Conference Leader of the European Alliance of Innovation (EAI), as well as with the steering chair Eva Brooks. We are genuinely thankful for the organising committee team's support: João Martinho Moura (Web Chair), Eva Brooks (Publicity and Social Media Chair), Jeanette Sjöberg (Workshop Chair), Eva Brooks (Publication Chair), Maria Vesterlund (Sponsorship and Exhibit Chair), Emma Edstrand and Maja Højslet Scürer (Poster Chairs), and Maria Vesterlund (Local Chair).

We also acknowledge the outstanding work by the Technical Programme Committee Members. Last, but not least, we are grateful to all the authors who submitted their papers and workshop proposals to the DLI 2024 conference.

The conference was held online and organized by Halmstad University in Sweden.

In conclusion, we firmly believe that the DLI conference offers a valuable platform for researchers, designers, educators, and practitioners to engage in meaningful discussions about the interdisciplinary field of digital technology and its impact on design, learning, and innovation. Additionally, we anticipate that future DLI conferences will continue to serve as a vibrant arena for knowledge sharing, as reflected in the contributions included in this volume. We look forward to welcoming you all to DLI 2025 – see you there!

Eva Brooks  
Emma Edstrand  
Anders Kalsgaard Møller  
Thomas Bjørner

# Preface

We are delighted to introduce the proceedings of the ninth edition of the European Alliance for Innovation (EAI) International Conference on Design, Learning & Innovation (DLI 2024). This conference brought together researchers and practitioners around the world to share their latest research findings in relation to development, use, and learning with digital and emergent technology. The theme for DLI 2024 was: *New theories, new methods, new technologies*. Design, learning, and innovation are often embedded in various practices, discourses, and target groups. As media and technology in a wide range of ways are intertwined with design, learning, and innovation, these boundaries are also challenged – empirically as well as theoretically and methodologically. The aim of the EAI DLI conference is to provide several different perspectives for developing theoretical and methodological approaches to the study of design, learning, and innovation.


The technical program of DLI 2024 included 10 paper presentations and a workshop session. The paper presentations were organized into two track sessions:


- **Session 1:** Using Immersive Technologies for Learning, Accessibility, and Technological Innovation
- **Session 2:** Engaging Learners through Gamification, Playful Design, and Generative AI

The two tracks were chaired by Thomas Bjørner from Aalborg University in Denmark and Emma Edstrand from Halmstad University in Sweden. Their contribution to the discussion related to the sessions was productive and created fruitful and engaged dialogues. The workshop session was chaired by Anders Kalsgaard Møller from Aalborg University in Denmark and focused on: *Designing for Diversity, Equality, and Inclusion in Higher Education Institutions: Exploring issues of sustainability, sense of belonging, inclusive designs and co-creation strategies*. It was organized by Sylvana Sofkova Hashemi (Halmstad University), Nataliya Berbyuk Lindström (University of Gothenburg), and Eva Brooks (Aalborg University) and was an outcome of the EDUCERE Nordic Network. During the workshop, the participants were invited to collaborative activities where they shared teaching and learning experiences, design models, and research evidence reflecting upon future digital skills and design strategies ensuring diversity, equality, and inclusion in higher education.


Elizabet (Liz) Sanders of Ohio State University, USA, gave an inspirational keynote presentation titled *Design and the Social Sciences: A Transdisciplinary Transformation*, which was grounded in her extensive research commitments within the transdisciplinary fields of co-creation and co-design. She addressed the integration between design and the social sciences and introduced some of the frameworks that she developed as a design research practitioner and design educator. Further, she described the current movement from participatory design to “collective dreaming” and introduced potential

*Editors*

Eva Brooks   
Aalborg University  
Aalborg, Denmark

Emma Edstrand   
Halmstad University  
Halmstad, Sweden

Anders Kalsgaard Møller   
Aalborg University  
Aalborg, Denmark

Thomas Bjørner   
Aalborg University  
Aalborg, Denmark

ISSN 1867-8211

ISSN 1867-822X (electronic)

Lecture Notes of the Institute for Computer Sciences, Social Informatics  
and Telecommunications Engineering

ISBN 978-3-031-85662-4

ISBN 978-3-031-85663-1 (eBook)

<https://doi.org/10.1007/978-3-031-85663-1>

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Eva Brooks · Emma Edstrand ·  
Anders Kalsgaard Møller · Thomas Bjørner  
Editors

# Design, Learning, and Innovation

9th EAI International Conference, DLI 2024  
Virtual Event, November 7–8, 2024  
Proceedings

The LNICST series publishes ICST's conferences, symposia and workshops.

LNICST reports state-of-the-art results in areas related to the scope of the Institute.

The type of material published includes

- Proceedings (published in time for the respective event)
- Other edited monographs (such as project reports or invited volumes)

LNICST topics span the following areas:

- General Computer Science
- E-Economy
- E-Medicine
- Knowledge Management
- Multimedia
- Operations, Management and Policy
- Social Informatics
- Systems

# Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering

604

## Editorial Board Members

Ozgur Akan, *Middle East Technical University, Ankara, Türkiye*

Paolo Bellavista, *University of Bologna, Bologna, Italy*

Jiannong Cao, *Hong Kong Polytechnic University, Hong Kong, Hong Kong*

Geoffrey Coulson, *Lancaster University, Lancaster, UK*

Falko Dressler, *University of Erlangen, Erlangen, Germany*


Domenico Ferrari, *Università Cattolica Piacenza, Piacenza, Italy*

Mario Gerla, *UCLA, Los Angeles, USA*

Hisashi Kobayashi, *Princeton University, Princeton, USA*

Sergio Palazzo, *University of Catania, Catania, Italy*

Sartaj Sahni, *University of Florida, Gainesville, USA*

Xuemin Shen , *University of Waterloo, Waterloo, Canada*

Mircea Stan, *University of Virginia, Charlottesville, USA*

Xiaohua Jia, *City University of Hong Kong, Kowloon, Hong Kong*

Albert Y. Zomaya, *University of Sydney, Sydney, Australia*

Eva Brooks  
Emma Edstrand  
Anders Kalsgaard Møller  
Thomas Bjørner (Eds.)



604

LNICST

# Design, Learning, and Innovation

9th EAI International Conference, DLI 2024  
Virtual Event, November 7–8, 2024  
Proceedings

