



 Latest updates: <https://dl.acm.org/doi/10.1145/3742414.3795099>

WORK IN PROGRESS

Exploring the Entertainment Potential of Personalised LLM-Based Serious Games with Humanoid Robots

GIULIO CANAPA, Institute of Information Science and Technologies, Pisa, Pisa, PI, Italy

BENEDETTA CATRICALÀ, Institute of Information Science and Technologies, Pisa, Pisa, PI, Italy

MARCO MANCA, Institute of Information Science and Technologies, Pisa, Pisa, PI, Italy

FABIO PATERNÒ, Institute of Information Science and Technologies, Pisa, Pisa, PI, Italy

CARMEN SANTORO, Institute of Information Science and Technologies, Pisa, Pisa, PI, Italy

ELEONORA ZEDDA, Institute of Information Science and Technologies, Pisa, Pisa, PI, Italy

Open Access Support provided by:

Institute of Information Science and Technologies, Pisa



PDF Download
3742414.3795099.pdf
23 March 2026
Total Citations: 0
Total Downloads: 0

Published: 22 March 2026

Citation in BibTeX format

IUI '26: 31st International Conference on
Intelligent User Interfaces
March 23 - 26, 2026
Paphos, Cyprus

Conference Sponsors:
SIGCHI
SIGAI

Exploring the Entertainment Potential of Personalised LLM-Based Serious Games with Humanoid Robots

Giulio Canapa
HIIS Laboratory
CNR-ISTI
Pisa, Italy
giulio.canapa@isti.cnr.it

Benedetta Catricalà
HIIS Laboratory
CNR-ISTI
Pisa, Italy
benedetta.catricala@isti.cnr.it

Marco Manca
HIIS Laboratory
CNR-ISTI
Pisa, Italy
marco.manca@isti.cnr.it

Fabio Paternò
HIIS Laboratory
CNR-ISTI
Pisa, Italy
fabio.paterno@isti.cnr.it

Carmen Santoro
HIIS Laboratory
CNR-ISTI
Pisa, Italy
carmen.santoro@isti.cnr.it

Eleonora Zedda
HIIS Laboratory
CNR-ISTI
Pisa, Italy
eleonora.zedda@isti.cnr.it

Abstract

Serious games are increasingly being considered in several fields, offering opportunities that traditional methods cannot easily replicate. In this perspective, Large Language Models (LLMs) and social robots have begun to be considered as means to enhance engagement and user experience in serious games: the former by generating content tailored to users' specific needs, and the latter by exploiting their embodiment and reinforcing feedback through speech and animations. We investigated the entertainment potential of two personalised serious games that combine an LLM and a humanoid robot in a science outreach setting. The games, deployed through a Pepper robot, used GPT-4o support in order to personalise a collaboratively created story and a quiz game based on users' information and choices. Analysis of user ratings and qualitative feedback from 33 users indicates that games combining LLM-based personalisation and interaction with a humanoid robot were well-received, offering an entertaining and enjoyable experience.

CCS Concepts

• **Human-centered computing** → Human computer interaction (HCI); HCI design and evaluation methods.

Keywords

Serious Games, Large Language Models, Human-Robot Interaction, Personalisation

ACM Reference Format:

Giulio Canapa, Benedetta Catricalà, Marco Manca, Fabio Paternò, Carmen Santoro, and Eleonora Zedda. 2026. Exploring the Entertainment Potential of Personalised LLM-Based Serious Games with Humanoid Robots. In *Companion Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI Companion '26)*, March 23–26, 2026, Paphos, Cyprus. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3742414.3795099>



This work is licensed under a Creative Commons Attribution 4.0 International License. *IUI Companion '26, Paphos, Cyprus*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1985-1/2026/03
<https://doi.org/10.1145/3742414.3795099>

1 INTRODUCTION

Over the last two years, LLMs have begun to be considered [13, 15] in serious games development for different purposes. Their capability to generate content and tailor it to users' specific needs has been applied in applications such as storytelling [7], personalisation [1], and game difficulty balancing [16]. These applications may improve engagement and UX by proposing interactions that adapt content to make learning or rehabilitation more engaging, stimulating and suitable for users than traditional methods. At the same time, social robots have also been considered in serious games to make game activities more engaging and pleasurable than conventional devices, thanks to features such as embodiment, speech, animation and multimodal interaction [5, 6, 8, 10] However, while these technologies have been involved with positive effects on engagement and UX, their combined use in serious games remains underexplored. To better understand the potential of such a combination, we present a solution that delivers LLM-based serious games through a humanoid robot, along with the user feedback received. The application was proposed in a science outreach setting (the European Researchers' Night in 2025), during which many research results were proposed to the visitors. The entertainment potential of combining AI-driven personalisation with human-robot interaction in serious games is discussed through an analysis of users' ratings and feedback received via semi-structured interviews.

2 RELATED WORK

Due to their potential for generating content adapted to users' preferences and specifics, LLMs are increasingly being considered in several personalised serious games aimed at education, empowering both students and teachers with adaptive tools tailored to students' needs. For example, Bonetti et al [1] considered GPT-3.5 in a framework for adapting the content of quizzes according to a user model based on the student's performance in previous interaction. In this work, teachers can adapt the content of the serious game to align it with the educational goal. For the same purpose, Camilleri and Camilleri [3] applied the LLM in a serious game intended to assist both teachers and students: they used GPT-3.5 Turbo for generating quizzes, balancing difficulty based on previous results, and providing explanations for incorrect answers. Adaptive difficulty is another task for which LLMs have been

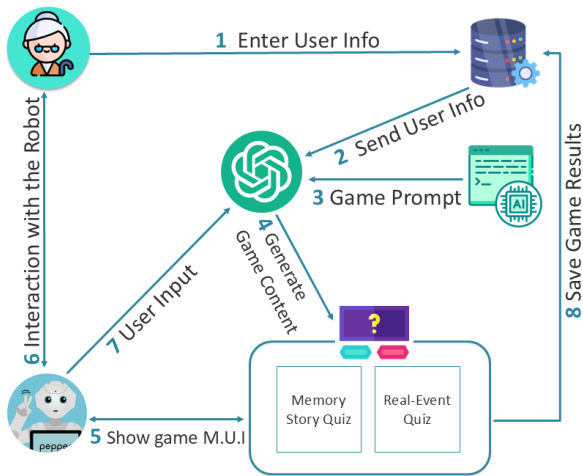


Figure 1: a) System architecture diagram. b) Game session at the science outreach.

considered in serious games. Volden et al. [14] exploited an LLM to generate rules that balance the difficulty of a serious game for children based on narratives and quizzes. Creating narratives represents another emerging approach in LLM-based games to make educational activities more entertaining for learners. For example, GPT was used in a narrative game for adolescent health education [12] to generate dialogues, with guidelines describing the game’s narrative and mechanics. Other contributions further expanded the application of narration by allowing users to co-create stories with the LLM. For example, two visual novels [7, 9] have been presented that allow users to influence the narrative by choosing between different dialogue options. Consequently, we decided to expand the collaborative narration approach by utilising an LLM to generate stories and quizzes, whose content is adapted to users’ choices and personal preferences.

3 THE GAME PLATFORM

We designed and developed two personalised serious games for a Pepper humanoid robot, whose content is generated using GPT-4o. The application starts with a brief interview to ask the users their age, their favourite narrative genre, and, depending on whether the user is a minor or an adult, a hobby (e.g.: sports, tv series, or other personal interests) or a personal memory (e.g.: memories about their travels, family, job, or other significant information). The content of the two serious games is adapted to the requested information.

In the Memory Story Quiz game, the LLM is involved in storytelling and quiz generation. At the start of the game, the LLM starts creating a short story inspired by the narrative genre and the shared information. The story is then co-created with the user, who can influence the outcome by selecting one of three possible choices proposed by the LLM in five subsequent decision points. After selecting an option for each decision point, the story concludes, and the player is asked to answer three quiz questions based on the textual information composing the story.

In the Real Event game, the LLM generates three general culture questions based on the information the participant shares. The user’s age is considered in balancing the difficulty of the questions (general-purpose questions for minors, and more specific ones for adults).

Each serious game is handled by GPT-4o via a prompt that employs three prompting techniques [11].

- Role prompting: the LLM is asked to assume the role of a playful robotic assistant.
- Decomposition: the main task is divided into several steps, each one describing a phase of the game.
- Few-shot prompting: for each phase of the game, several examples in JSON format are provided to guide the LLM in generating coherent responses.

Both games are delivered via a Pepper humanoid robot, which enables multimodal interaction via its touchscreen and speech recognition, and provides reinforcing feedback through its speech and body animations. Such feedback was mainly provided when users entered an answer, depending on its correctness, to convey an emotional state (for example, happiness when the answer was right, sadness when it was wrong).

The system architecture is represented in Figure 1. The interaction begins with the user entering their information, which is saved in a database (step 1). This information is sent to GPT-4o (step 2), along with the game prompts (step 3), to generate content for the two serious games (step 4). User interaction is performed through the Pepper robot (step 5), which delivers game content for voice or tablet-based interaction, depending on the user’s preferences (step 6). During the session, user inputs are forwarded to GPT-4o to advance the games (Step 7). When the session ends, results are sent to the backend to store game time and the number of correct and wrong answers for each game (step 8).

4 USER STUDY

The system was deployed in a science outreach setting to evaluate the entertainment potential of personalised LLM-based serious

games (see Figure 1). Visitors who wanted to interact with the application were asked to play at least one game and to sign a consent form beforehand, which informed them about the study, what their participation involved, and how data were collected, used, and protected. If the participant was a minor, a parent or a legal guardian was required to sign the form. During the event, the game system was played by 33 participants (16 female, 17 male) aged 6-76 ($M = 18.09$, $SD = 17.25$), with the majority being elementary school students aged 6-10 (57.6%). Participants were asked to choose one game to play. After their session, users were asked to complete a semi-structured interview with three main questions: Q1 (“How much did you like playing with the robot?”) to evaluate the entertainment value of the game system on a 5-point scale, Q2 (“Which aspects did you like the most and the least of the AI-generated serious games?”) to gather positive and negative feedback on the serious games, and Q3 (“Do you have other comments or suggestions?”) to collect recommendations for further development.

5 RESULTS

The average session time was approximately 10 minutes. The analysis of quantitative data, including game time and users’ errors, revealed no significant correlations among errors, game completion time, or participants’ age. The two-way ANOVA on game completion time ($F = 225.48$, p -value: $6.35e-15$) indicated that the difference between the two games was significant, with Memory Story Quiz requiring more interactions to create stories and quizzes than Real Event. The user’s assessment (Q1) yielded a high mean score and a low standard deviation ($M = 4.24$, $SD = 0.70$), indicating that users found the robot application fun and entertaining. The answers to Q2 and Q3 were analysed using a thematic analysis [2] to identify strengths, weaknesses, and areas for improvement in the current application. In particular, the thematic analysis revealed the following recurrent themes in users’ responses.

Personalisation through LLMs: the most recurrent theme is personalisation, with fifteen participants appreciating how the content of stories and quizzes is coherent and adapted to the information shared at the start of the game. Users also appreciated the creativity and collaborative nature of the story generation, thanks to the ability to influence the narrative through their choices.

Perception of the social robot: five participants commented on Pepper’s movement and animations, with four of them appreciating how the robot moved its body and hands. One participant, however, commented negatively on its animations, judging the movements in some cases to be too frequent and excessive. Two users also commented positively on the robot’s behaviour, describing it as funny and pleasant.

Interaction: six users reported interactions with the system, with four commenting that the robot exhibited highly interactive behaviour. Two users reported that they would have preferred voice interaction over touchscreen (vocal interaction was deactivated due to the noisy environment).

Criticism: Six users raised concerns about specific aspects of the system, with three commenting that the content generated by a single question was not coherent or relevant to the information they had entered (an issue not reported by the other participants). Two users commented on the answer speed, considering it too slow.

Lastly, one user mentioned that during their session, the correct answer was always generated in the same position, making the experience feel more repetitive and predictable.

6 CONCLUSIONS AND FUTURE WORK

The results of the semi-structured interviews allowed us to understand the strengths, weaknesses, and areas for improvement of the proposed application in science outreach, as well as identify the requirements for future work involving LLMs and HRI simultaneously. Personalisation was identified as the main factor required for LLM-based generation in a serious game. Users appreciated how the LLM adapted the game content to their memories and personal interests, and how they could modify the narrative outcome with their input. This feedback suggests that designing the game prompt by indicating the relevant subtasks and providing few-shot examples permitted the LLM to generate personalised, yet coherent and relevant content. However, a small sample of users (3 out of 33 participants) reported that the LLM was not always able to generate relevant responses, indicating that identifying more specific requirements during the interview phase and in the game prompt could further improve the coherence of the generated content. The behaviour and movements of the social robot, as well as its interaction with participants, were also appreciated, indicating that the robot was perceived as a positive agent in proposing game activities. However, movements and animations could be further refined to convey a more natural behaviour.

This study has limitations regarding the duration of application use and user heterogeneity. Since the system was primarily used by elementary school students, and was evaluated positively by most users, one possible future direction is to expand upon LLM-based personalisation for cognitive stimulation [4] and childhood education activities. Deploying LLM-based story and quiz tools in education could make the learning experience more engaging, enjoyable, and tailored to the specific needs and characteristics of upper elementary school children.

Acknowledgments

The study was partially funded by the SERENI CNR research project (<https://hiis.isti.cnr.it/sereni/index.html>).

References

- [1] Federico Bonetti, Antonio Bucchiarone, and Vanessa Wanick. 2025. Using LLMs to Adapt Serious Games with Educators in the Loop. In: Schönbohm, A., *et al.* Games and Learning Alliance, GALA 2024, Lecture Notes in Computer Science, vol 15348, pp. 68–77, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-031-78269-5_7
- [2] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. In: *Qualitative Research in Psychology*, Vol. 3, pp. 77–101, <http://dx.doi.org/10.1191/1478088706qp063oa>
- [3] Reno Yuri Camilleri and Vanessa Camilleri. 2024. Beyond the Maze: How AI Personalizes Learning and Drives Engagement in Educational Games. In: Arai, K. (eds) Proceedings of the Future Technologies Conference (FTC) 2024, Vol., 2, FTC 2024. Lecture Notes in Networks and Systems, vol 1155, pp. 301–320, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-031-73122-8_19
- [4] Giulio Canapa, Fabio Paternò, and Carmen Santoro. 2025. Interactive Serious Games for Cognitive Training of Older Adults: A Systematic Review. *IEEE Transactions on Computational Social Systems*, vol. 12, no. 5, pp. 2445–2468, <https://doi.org/10.1109/TCSS.2025.3532307>
- [5] Benedetta Catricalà, Marco Manca, Fabio Paternò, Alessandro Sale, Carmen Santoro, and Eleonora Zedda. 2025. Exploiting personal memories in humanoid robot serious games for mild cognitive impaired older adults. *Behaviour & Information Technology*, pp. 1–26, <https://doi.org/10.1080/0144929X.2025.2484647>

- [6] Norina Gasteiger, Ho Seok Ahn, Christopher Lee, Jongyoon Lim, Bruce A. MacDonald, Geon Ha Kim, and Elizabeth Broadbent. 2022. Participatory Design, Development, and Testing of Assistive Health Robots with Older Adults: An International Four-year Project. *J. Hum.-Robot Interact.* 11, 4, Article 45 (December 2022), 19 pages. <https://doi.org/10.1145/3533726>
- [7] Mustafa Can Gursesli *et al.* 2025. The Role of Large Language Model-Generated Stories in the Narrative Experience of Serious Visual Novel Games. *International Journal of Human-Computer Interaction*, pp. 1–20. <https://doi.org/10.1080/10447318.2025.2528993>
- [8] Marco Manca, Fabio Paternò, Carmen Santoro, Eleonora Zedda, Chiara Braschi, Roberta Franco, and Alessandro Sale. 2021. The impact of serious games with humanoid robots on mild cognitive impairment older adults. In: *International Journal of Human-Computer Studies*, Vol. 145, No. 102509, pp. 1–11, <https://doi.org/10.1016/j.ijhcs.2020.102509>
- [9] Nicola Marsden., Annelie Rothe-Wulf., Yvonne Tang, Claudia Herling and Tim Reichert. 2025. Navigating Bias: Using LLMs to Analyze Discrimination in Entrepreneurial Game Dialogues. In: Fang, X. (eds) *HCI in Games. HCII 2025. Lecture Notes in Computer Science*, vol 15816. Springer, Cham. https://doi.org/10.1007/978-3-031-92578-8_14
- [10] Emilyann Nault, Lynne Baillie, and Frank Broz. 2024. Socially Assistive Robots and Sensory Feedback for Engaging Older Adults in Cognitive Activities. *J. Hum.-Robot Interact.* 14, 1, Article 13 (March 2025), 26 pages. <https://doi.org/10.1145/3698241>
- [11] Sander Schulhoff *et al.* 2024. The Prompt Report: A Systematic Survey of Prompting Techniques. In: *arXiv preprint arXiv:2406.06608*, pp. 1-80, <https://doi.org/10.48550/arXiv.2406.06608>
- [12] Ian Steenstra, Prasanth Murali, Rebecca B. Perkins, Natalie Joseph, Michael K Paasche-Orlow, and Timothy Bickmore. 2024. Engaging and Entertaining Adolescents in Health Education Using LLM-Generated Fantasy Narrative Games and Virtual Agents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 126, pp. 1–8. <https://doi.org/10.1145/3613905.3650983>
- [13] Penny Sweetser. 2024. Large Language Models and Video Games: A Preliminary Scoping Review. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 45, pp. 1–8. <https://doi.org/10.1145/3640794.3665582>
- [14] Thomas Volden, Djordje Grbic, and Paolo Burelli. 2023. Procedurally generating rules to adapt difficulty for narrative puzzle games. In: *2023 IEEE Conference on Games (CoG)*, pp. 1–4, Boston, MA, USA, <https://doi.org/10.1109/CoG57401.2023.10333251>
- [15] Daijin Yang, Erica Kleinman, and Casper Hartevelde. 2024. GPT for Games: An Updated Scoping Review (2020-2024), in *IEEE Transactions on Games*, <https://doi.org/10.1109/TG.2025.3563780>
- [16] Haiwei Zuo, King Hann Lim, Sivaraman Eswaran, and Basil Andy Lease. 2024. Development of Serious Game Theory Framework in Virtual Reality for Alzheimer's Patients. In: *2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, pp. 190–194, Miri Sarawak, Malaysia, <https://doi.org/10.1109/GECOST60902.2024.10474644>