Book Chapter

A Virtual Tutoring System with Gamification, LLM-Guided NPCs, and Online Tutor Support†

Ariadni Barmpari, Iasonas Pavlopoulos, Eleni Voyiatzaki and Ioannis Hatzilygeroudis*

Computer Engineering and Informatics Department, University of Patras, 26504 Patras, Greece

†The paper is an extension of the following conference paper: Barmpari, A.; Voyiatzaki, E.; Hatzilygeroudis, I. An Educational Virtual World System with Gamification Features and LLM Guided NPCs. In *International Conference on Intelligent Tutoring Systems*; Graf, S., Markos, A., Eds.; ITS 2025, LNCS 15723; Springer Nature: Cham, Switzerland, 2026; pp. 213–223.

***Corresponding Author:** Ioannis Hatzilygeroudis, Computer Engineering and Informatics Department, University of Patras, 26504 Patras, Greece

Published **July 29, 2026**

This Book Chapter is a republication of an article published by Ioannis Hatzilygeroudis, et al. at Applied Sciences in January 2026. (Barmpari, A.; Pavlopoulos, I.; Voyiatzaki, E.; Hatzilygeroudis, I. A Virtual Tutoring System with Gamification, LLM-Guided NPCs, and Online Tutor Support. Appl. Sci. 2026, 16, 899. https://doi.org/10.3390/app16020899)

How to cite this book chapter: Ariadni Barmpari, Iasonas Pavlopoulos, Eleni Voyiatzaki and Ioannis Hatzilygeroudis. A Virtual Tutoring System with Gamification, LLM-Guided NPCs, and Online Tutor Support. In: Prime Archives in Applied Sciences: 2nd Edition. Hyderabad, India: Vide Leaf. 2026.

© The Author(s) 2026. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Author Contributions: Conceptualization, E.V. and I.H.; methodology, A.B., I.P., E.V. and I.H.; software, A.B. and I.P.; validation, A.B., I.P. and E.V.; formal analysis, I.H. and E.V.; visualization, I.H., A.B. and I.P.; writing—original draft preparation, I.H.; writing—review and editing, I.H. and E.V.; supervision, E.V. and I.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not require formal ethical review or approval because it involved the evaluation of educational software in a standard educational context with adult participants. No vulnerable populations were involved, participation was voluntary, data were collected anonymously, and no sensitive personal information was obtained.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study; participation was voluntary.

Data Availability Statement: The data presented in this study are available on request from Dr. Eleni Voyiatzaki (**evoyiatzaki@ceid.upatras.gr**) due to the fact that they are not yet publicly available (need presentation refinement).

Conflicts of Interest: The authors declare no conflicts of interest.

Abstract

Most of the existing Virtual World (VW)-based curriculum-related educational systems use conventional non-player characters (NPCs) to interact with users, represented as avatars, to guide and help them to accomplish learning activities. Also, a few of them use some kind of gamification and keep data for user interactions and activities, and even fewer allow for real-time tutor intervention. In this paper, we present the design, implementation, and evaluation of an educational system based on VW technology, which employs gamification features; two types of NPCs, one conventional and another LLM-based; and a database that stores, apart from educational information, information about the interactions users have with NPCs. Furthermore, we designed and implemented a learning management unit for online-tutor tracing and for supporting the learning progress of users. The evaluation of the system, via experimental use and questionnaires, shows that both types of NPCs were useful for different reasons, although there was a preference for the LLM-based NPC. LLM-based NPCs made dialogues more interesting and were perceived as more friendly and helpful, but conventional ones provided more targeted help. However, both were less interesting than the two gamification features: a scoring system and quizzes. Additionally, the effectiveness of the tutoring system was confirmed in terms of learning outcomes and overall experience, although in a subjective manner. Finally, online-tutor support was recognized as a very positive capability.

Keywords

Intelligent Tutoring System; Virtual World; Gamification; LLM-Based NPC; Online Tutor Support; LLM-Based Support; Functional Aspects Evaluation; Affective Aspects Evaluation

1. Introduction

Virtual Reality (VR) is a simulated experience created through interaction with a computer-generated 3D environment that simulates reality, creating a virtual world. There has been

extensive research into using VR in education and training [1]. In this paper, we focus on non-immersive VR-based educational applications, i.e., those that do not require the use of special devices that create a sense of immersion [2]. More specifically, we are referring to Virtual World (VW) technology used for educational purposes: "A virtual world is a computer-simulated environment, which may be populated by many simultaneous users who can create a personal avatar and independently explore the virtual world, participate in its activities, and communicate with others" [3]. Avatars are humanoid graphical characters that mediate user experiences, allowing for interaction with various objects in the virtual world (e.g., displays, panels, machines, etc.), with other artificial entities, like Non-Player Characters (NPCs) and with other avatars, representing other users, to construct a shared understanding of the world [4,5]. In this context, we use the terms "VW", "virtual environment", "3D world", and "3D VW" as semantically equivalent.

To make educational software more attractive to student-users, game-based features are added to a VW system. This methodology is called gamification. Gamification takes educationally interesting features from video games, such as awards, badges, etc., and embeds them into purely educational systems. Gamification constitutes an active methodology that can enhance students' motivation and engagement [6].

In this paper, we refer to educational curriculum-oriented VW-based teaching/learning systems that use some type of gamification. Such systems are more education-oriented than game-oriented, in contrast to serious games [7]. They are also more easily developed by VW development tools. Second Life, OpenSim, and Unity are the most used VW developing tools for such educational software [8]. A VW implementing the above development tools may contain auditoriums, meeting rooms, libraries, media rooms, displays for graphical presentations, laboratories, 3D objects, simulations of machine or physical processes, and online tests. Also, it can contain interactive elements, like interactive 3D objects, notecards, and NPCs. Notecards are texts related to avatars, especially NPCs. They can

contain links to external web pages or to other notecards and images. They are activated under specific conditions.

Non-Player (or non-playable) Characters (NPCs) are computer-controlled avatars that play a crucial role in enhancing interaction and learning in VWs [5]. These artificial personalities can function as guides, leading students through activities, by providing information, and helping them to pass to the next stage(s) of a gamified process. In most (if not all) existing VW systems, NPCs are pre-programmed with pre-designed responses to pre-designed questions [9]. This may create disappointment for users.

On the other hand, for the tutor, it is essential to understand how students interact with the objects and personalities in the world and how the learning process unfolds through virtual worlds. Monitoring students' behavior during educational activities, such as their movements, decisions, and interactions with NPCs, can provide valuable insights into their learning needs. This possibility allows tutors to adjust their approach, identify the challenges students face, and provide targeted support. This creates a more effective learning environment and experience. However, this is not possible in most existing VW systems, because avatars' behaviors are not properly recorded, and there is no interaction mechanism between the tutor and the learning activities of the students in real time.

In this paper, we present the design, implementation, and evaluation of a VW-based educational system that fulfills the above shortcomings. First, a new type of NPCs is introduced, called LLM-NPCs, which exploits the capabilities of ChatGPT - 4 mini to provide flexibility and some type of intelligence to NPCs. Second, the system is supported by a local database, where learning material (e.g., presentations, quizzes) is stored, and avatar (user) activities are recorded. So, the tutor can look at and edit existing material, as well as store new material. Additionally, a learning management component is integrated into the system that provides statistics, performance summaries, facilitates communication between the tutor and the students' session data, and accomplishes interaction with the students by

providing support in real-time. Finally, we present a questionnaire-based evaluation of the new introduced capabilities of the system, across various qualities (related to NPCs, the tutor, and the system), based on learners' experience with the system.
This work is an extension of the work in [10]. The extension concerns the following:

- Update of related works;
- New section on research objectives and methodology;
- Extended system design and implementation;
- Extra evaluation experiment;
- Validation of experimental results via inferential statistical tests;
- New discussion section on limitations of results.

The structure of the paper is as follows. Section 2 presents related work, while Section 3 presents research questions and methodology. Section 4 deals with system design and implementation. Section 5 deals with the learning process and its management, whereas Section 6 presents the evaluation of the system. Section 7 includes a discussion on the statistical results, and finally, Section 8 concludes the paper.

2. Related Work

2.1. Educational Virtual Worlds and NPCs

In this paper, we focus on 3D Virtual Worlds that are used for educational purposes tightly related to some kind of curriculum. Such systems, both early and even those recently, were not as interactive as required, remaining at representations or simulations of real-world phenomena, processes of devices, and their functionalities. In those cases, students were entering the worlds as avatars to explore various items [11,12,13,14]. Later, interactive objects came onto the scene, as well as the use of notecards and NPCs. As time progressed, the degree of interaction between the user-avatar and the NPCs increased and improved. On the other hand, gamification features were gradually added. Finally, most of them were implemented in OpenSim and the rest in Unity.

In [15], the authors present an innovative 3D virtual reality educational environment that aims to assist students in learning and tutors in explaining various processes of a physics course. In the 3D virtual reality environment, laboratories facilitated students to carry out virtual experiments, explore procedures, and gain a deeper cognition and understanding of how procedures are conducted, and how physics processes work. In addition, pedagogical virtual agents, such as one that looked like Albert Einstein, were designed and implemented as NPCs to guide students in the virtual environment and assist them during the training activities. No gamification was used.

In [16], a 3D virtual world, called VR4STEM (Virtual Reality for STEM Entrepreneurship Training), was presented. VR4STEM was proposed to assist young people in gaining entrepreneurship skills. The virtual environment of VR4STEM was composed of several 3D islands. Each of them presented a specific subject in the STEM (Science, Technology, Engineering, and Mathematics) and ICT (Information and Communications Technology) industry domains, e.g., "World of Lasers" and "World of Unmanned Aerial Vehicles". Notecards and NPCs were used to give information to users.

A 3D VW was used as the testbed for testing the "sense of presence", considered as attention and involvement of the student-user in [17]. The subject of the 3D VW was "Financial Management". The educational scene consisted of a building appropriately structured to simulate an accounting company. Student-users were represented as avatars, but there were also several NPCs distributed in the company's offices that guided and helped them. The NPCs could "express themselves bodily and textually". The world embedded some features of gamification (e.g., the need to complete a phase before going to the next, score-based quizzes, etc.).

The authors of [18] presented a 3D virtual world for learning aspects of environmental engineering and renewable energy sources. The 3D VW included various constructions, buildings, machines, and power plants that resembled the respective constructions of the real world and mimicked and visualized

their functionality. Interaction with objects was possible via scripts. Various learning activities were offered to the students. The students could manipulate and interact with constructions in the context of gamification scenarios. NPCs, acting as pedagogical agents, were incorporated into the virtual world to accompany learners and support them during the training activities. The agents were avatars, controlled exclusively by scripts that could guide students and provide context about their objectives during the activities. When needed, students could request assistance from the agents, who could provide confirmations as to whether an answer was correct or not, address the answer and related topics, specify particular error(s), and guide the student about what to do towards the correct answer.

In [19], a 3D VW implemented in OpenSim was reported. Three different pedagogical agents, represented as NPC avatars, were used to make learning more attractive and to improve engagement. Jella Delta had a humanoid form, resembling the role of the instructor or educator, and was a conversational agent (chatbot) with knowledge-intensive and domain-specific question answering capabilities. Its role was to facilitate the learning process and to support students by providing useful and meaningful answers to queries related to the virtual world. Queen Kong was also a chatbot, though of a nonhuman type, as an example of the contradictory content that virtual worlds could accommodate. Its role is to disorientate students by providing incorrect or 'nonsense' answers to their queries in a 'ludicrous' way. Gizmo Gear had a robot-like form, operating as a vendor. This agent became interactive upon a student's call, and its role was to provide informational notecards (digital text-based notes), assign or suggest tasks, and offer freebies (premade 3D objects and scripts).

The development of a 3D VW to be used as an educational tool for teaching a broad spectrum of energy concepts was presented in [20], focusing on energy saving, through a wind farm management environment. Users were transported into an engaging virtual setting, where they encountered real-life energy threat scenarios. These scenarios were ingeniously presented as

interactive games, allowing users to navigate through, engage with, and learn about various aspects, risks, and threats associated with energy use. The game included interactive elements, such as discussions with NPC experts, a quiz on wind energy, and tasks that require critical thinking and decision-making based on environmental impact and technical considerations.

The authors of [21] presented the development of a game-based VW, called "The Wonder land of IT", aiming at promoting a BSc degree curriculum, namely "Bachelor of Science in Information Technology". The scene of the game included a real university environment, featuring authentic buildings, infrastructure, and computer labs. NPCs played a crucial role in the game by communicating with players, providing stage-related information, notifying them of missions to be completed, and posing questions to progress through the stages. It encompassed various elements, such as the presentation of Information Technology concepts and the completion of missions at each stage. These missions involved puzzles, questions, and tasks to assess students' progression. Examples of missions included navigating to specific destinations, overcoming obstacles, searching for objects and secret codes, answering questions based on presented information, interacting with NPCs, and collecting coins to obtain a larger score.

The work in [22] refers to proposed improvements to a 3D VW (called ENTREALITY) that concerned learning social entrepreneurship concepts in an expressive, entertaining, comprehensive, and gamified way, which was implemented in OpenSim. An important proposed feature was the incorporation of mechanisms for monitoring user activity in the 3D world, analyzing and generating useful feedback in a personalized way through NPCs. To this end, learning analytics, like time spent, participation in the activities, score achieved, interactions with NPCs, etc., were recorded. Based on them, NPCs would provide personalized feedback.

2.2. NPCs and LLMs

To the best of our knowledge, there have yet to be non-immersive VW-based educational systems of the above type that use NPCs driven by an LLM. One work that could be considered close to the above is [23]. This work presented a framework for personalized teaching, regarding student, tutor, and institution levels, based on LLM capabilities. Regarding student level, a student-oriented learning assistant guided by an LLM was used. The assistant collected and analyzed student data and, based on the data, designed suitable learning paths and also gave feedback to students.

Also, in [24], a VW designed for brainstorming was presented. The main educational objects of the VW were two PCAs (Pedagogical Conversational Agents), called Rosie and Gigi. Rosie welcomed the user into the virtual world, provided the user with an introduction to the environment, and explained the brainstorming setting. Gigi could act as a brainstorming supporter and moderator in the virtual world. It was implemented with the "GPT Turbo Model", an AI-based language model. It helped learners to develop ideas on their own, step by step. Gigi was designed with the prompt for assessing learners' ideas and pointing out potential opportunities and challenges.

Ref. [25] introduced a framework that allowed instructors to collaborate with large language models to dynamically design realistic scenarios for students to communicate in the context of social skills training. An NPC, powered by an LLM, generated dynamic, novel dialogues on the fly by mapping specific paths within a scenario graph. Feedback on a student's response was provided through an LLM using an independent prompt, allowing the student to adjust their response based on the feedback received.

The authors of [26] proposed an architecture for multimodal immersive educational VR systems that integrated Speech-to-Text (STT), Text-to-Speech (TTS), and LLM components. The verbally expressed user query was converted into text by the STT component, which was then provided as the input for an

LLM (LLama 3.2/DeepSeek). The text-based output answer of the LLM was converted into speech via the TTS component and guided the verbal articulation of the answer via an NPC. Unity software was used for prototype implementation. Their findings showed improved engagement. Furthermore, Ref. [27] presented an extension to the previous study as far as the use of the LLM is concerned. More specifically, the same architecture was employed, but the cache-optimized Llama 2 LLM was used for improved performance, while Unity was also used for implementation. This significantly optimized interaction latency, which was a shortcoming of the above architecture. For evaluation purposes, the authors added an extra level in their framework, outside the VR system, called the Motivational Evaluation Layer, which applied Keller's ARCS model as a post-test instrument to determine whether the system experience sustained learner motivation, using a structured questionnaire. Learner motivation was assessed using Keller's ARCS model, yielding good scores across all dimensions.

On the other hand, there have been recent serious games and video games that used agents that were LLM-driven [7]. They constituted complex systems with goals, rules, and game mechanics that merged gameplay with learning, giving priority to the gaming aspect. They provided learning activities in a more playful, interactive, and engaging manner while increasing students' motivation and involvement. For example, in [28], an LLM was used to interact with players in a text-adventure game to explore how this interaction could give rise to emergent behaviors, empowering players to participate in the evolution of game narratives. Players could freely interact with non-player characters generated by GPT-4. In [29], the authors explored ways to communicate with LLM-based NPCs within a VR video game setting. Users could freely interact with ChatGPT-4 LLM-based NPCs to solve a murder mystery.

Furthermore, Ref. [30] presented a sophisticated architecture of integrating GPT in embodied agents for use in video games. The GPT-NPC agent can be split into four core modules: Perception, Situation, Conversation, and Speech Synthesis, and four complementary modules: Memory, Thoughts, Emotions, and

Needs. Player agents communicate with speech, which is then internally converted into text. Agents were aware of the surrounding virtual world, which was represented through strings called World Events. The Situation module was GPT-powered. Its role was to aggregate perceived world events into a summary to provide a comprehensive overview of what was happening. GPT was also involved in the Thoughts module through its API, which took in the internal agent state and generated new thoughts. The system proposed in the present paper is not meant to be a serious game or a video game.

3. Materials and Methods

3.1. Research Objectives and Questions

Our primary aim is to design, implement, and evaluate a virtual world tutoring system featuring both LLM-guided and traditional script-based NPCs, as well as incorporate a gamification aspect, assessing its impact on user perception and learning outcomes.

The main challenge in system development is the design of its architecture, which should ensure real-time dual-direction communication between a Unity-based application and an LLM, as well as communication with the system database and a tutor-monitored learning management unit. Another challenge is the design and implementation of a tutor intervention mechanism in real time during the learning process, without disturbing it.

Our educational research objectives concern the functional and affective aspects related to the use of two types of NPCs. Therefore, *functional aspects* of NPCs, like friendliness and usage, should be investigated. Also, we want to investigate whether the introduction of LLM-based NPCs gives significant value to the system as far as *affective aspects*, such as increases in interest and helpfulness, compared to traditional script-based NPCs. Furthermore, *tutor intervention usage* and *helpfulness* during the learning process should be evaluated. Finally, *learning outcomes* and the *overall experience* at the system level should be investigated. Figure 1 illustrates a taxonomy of the aspects of the system that are targeted for evaluation.

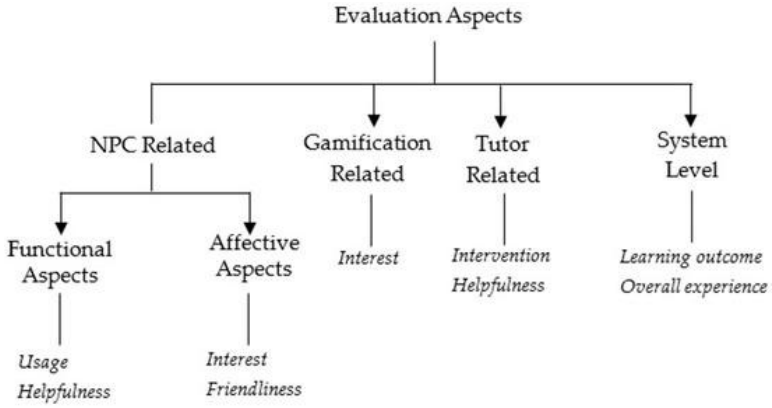

Figure 1: Categorization of the evaluation aspects of the system.

Given the above, our main research questions are as follows:

RQ1: What system architecture and communication mechanisms enable the seamless integration of LLM-driven dialogue with a Unity-based virtual-tutoring environment and a tutor-guided learning-management unit in real time?
RQ2: Is there a significant difference in the perceived usage and helpfulness between LLM-guided and script-based NPCs?
RQ3: Is there a significant difference in the perceived interest increase and friendliness between LLM-guided and script-based NPCs?
RQ4: How do users rank various tutoring system elements (NPCs, gamification elements, etc.) in terms of motivating their interest?
RQ5: Does a system that combines LLM-guided and script-based NPCs, as well as gamification, lead to significant learning outcomes and a positive overall experience?
RQ6: How do users perceive the tutor interventions during the learning process?

3.2. Research Design and Implementation

To address RQ1, we adopted an engineering research approach combining system development with empirical validation. The study focuses on the architectural design, communication

pipeline, and integration mechanisms required to embed LLM-driven dialogue agents within a Unity-based virtual tutoring system, while supporting real-time interaction and experimental control. We adopted a client–server architecture based on a REST (Representational State Transfer) API, and we used the WebSocket protocol for real-time communication, due to its full-duplex communication capabilities over a TCP connection. A detailed description of our solution is provided in Section 4.

To address RQ2–RQ5, we conducted two experiments in different time periods. We employed suitable participants who used the virtual tutoring system under specific scenarios. After each experiment, the participants were given a specific structured questionnaire to answer questions related to RQ2–RQ5, thus expressing their experience with the system. Afterwards, we statistically analyzed the answers to produce valid results/conclusions. A detailed description of our experiments and results is provided in Section 6.

4. System Design and Implementation

4.1. System Objectives

Our main objective was to improve aspects of existing 3D VW systems, like those presented in Section 2.1, towards the directions presented in Section 3. We used, as our basis, the VW produced in the context of an Erasmus+ European project, called ENTREALITY (https://projectentreality.etcenter.eu/index.php/en/ (accessed on 25 February 2025)), where we had participated in the consortium. That system was implemented in OpenSim and offered gamified scenarios for learning entrepreneurship concepts; however, in that instance, no NPCs were used.

A first design step in improving that system was presented in [22], where NPCs were employed and designed to offer adaptive feedback to users. This required registering and storing learning analytics measures and use of them for providing personalized feedback. In the present work, we went a step further by adding intelligence to NPCs via LLMs. Also, to improve graphics and

simulation capabilities, we used Unity instead of OpenSim to implement it. Also, we introduced a database for storing useful data and a tutor interface for handling data and items in the VW. Finally, we introduced a learning management unit (LMU) to allow tutors to track learning progress in real-time and support learners by giving hints.

LLMs, like OpenAI's ChatGPT, are advanced AI models trained to understand and generate human language. These models use transformer architectures to process vast datasets, enabling complex language understanding and contextual interactions [31]. In our application, LLMs were integrated into NPCs and chat interfaces, with the expectation that they would enhance interactivity and provide personalized responses. We initially utilized the free OpenAI version of ChatGPT. By introducing these intelligent NPCs, we aimed to improve communication and interaction, thus enhancing the learning process by offering feedback, answering questions, and guiding students through various learning activities in a non-predefined way, hopefully leading to better educational outcomes.

4.2. Tutoring System Architecture and Implementation

The system architecture is depicted in Figure 2. It consisted of four modules: *Unity VW environment*, Backend, Frontend, and OpenAI-ChatGPT. The core component (i.e., the VW) was built using Unity. It included a set of entities that users interact with in the learning environment to produce the desired outcome. Such entities were the Users (or Learners), and the NPCs, namely the SCRIPT-NPCs and LLM-NPCs, acted according to their scripts. Notice that the terms 'user', 'student', 'learner', and 'player' are considered as synonymous in this paper and used interchangeably.

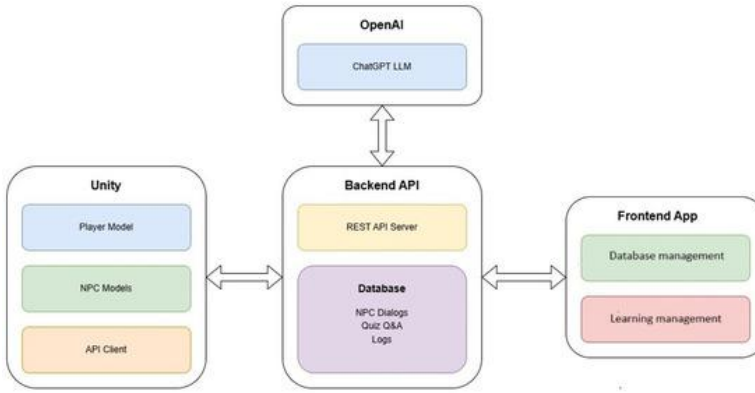

Figure 2: Virtual tutoring system overall architecture.

The *Backend* is the backbone of the entire system and consists of the Database and the REST API. The *Database* served two primary purposes: managing user accounts and logging learner activities. Through user account management, users could create personalized accounts, allowing the system to maintain detailed records for each learner. Simultaneously, the database captured and stored extensive data on learner activities during a learning session. This included their movements and navigation within the educational environment; quiz performance, such as answers given and scores achieved; interaction with educational materials, like slideshows and panels; as well as communication with NPCs, recording both the questions asked and the responses provided. By consolidating this data, the database offered educators a comprehensive overview of each learner's actions and learning journey.

Given that Unity could not communicate directly with a database, the creation of a *REST API* was necessary. A REST API (Representative State Transfer Application Programming Interface) is a set of rules that enables communication between software applications over the internet. It uses HTTP methods (GET, POST, PUT, DELETE) to manage and process data. In this case, the REST API facilitated communication between Unity and the Database. While Database was implemented in PostgreSQL, a relational database development open-source tool,

REST API was built using Nest.JS, a Typescript framework built on the Node.js engine.

The Backend was also responsible for communicating with the external service, *OpenAI-ChatGPT*, to provide answers to any of the learner's questions through the LLM-NPCs. The Backend stored the history of all messages the learner had sent and acted as a proxy to the OpenAI-ChatGPT LLM, which in turn answered the user's messages. The response was then relayed over to the Unity game via the Backend. This process continued until the user stopped sending messages.

To optimize and simplify the management of learning materials (e.g., presentations, quizzes) and the learning process (e.g., learning progress or difficulties of a user), a dedicated platform, the *Frontend*, was developed (with the React development tool), which was accessible exclusively to system administrators. It included two basic components: *database management* and *learning management*. Through database management, administrators could view all quizzes stored in the database, each accompanied by its unique identifier (ID) used to link it with the system. Additionally, it enabled administrators to easily delete or edit existing quizzes, as well as create new ones, streamlining content management and updates. Lastly, administrators had access to various statistical insights derived from the database logs, further enhancing their ability to oversee and analyze quiz-related activities. The learning management component was implemented as a learning management application and is presented in the next subsection.

4.3. Learning Management Unit Architecture and Implementation

To support the learners, a web-based *learning management unit* (LMU) was developed, enabling tutors to monitor and intervene in real time during the learning process executed within the Unity-based environment. The application addressed the critical challenge of transforming raw data generated by the educational system into meaningful pedagogical information, offering a comprehensive system for monitoring and decision

support. The implemented architecture established a real-time bidirectional communication between the Unity-based educational system and a web-based tutor dashboard, exploiting the WebSocket protocol for real-time communication (see Figure 3).

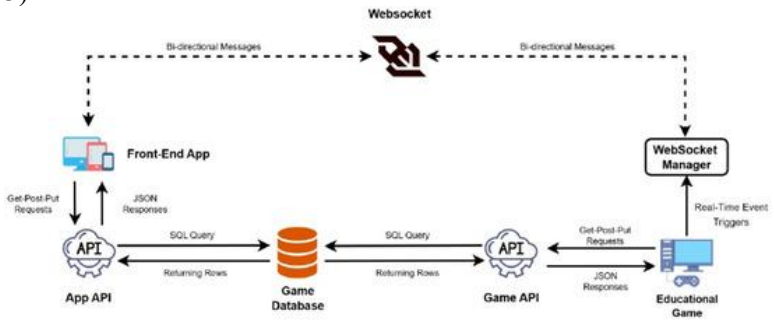

Figure 3: Learning management application architecture.

WebSocket protocol was selected as the primary real-time communication mechanism due to its full-duplex communication capabilities over a single TCP connection. Unlike traditional HTTP polling or long-polling techniques, WebSocket maintains a persistent connection after the initial handshake, eliminating the overhead of repeated connection establishments and reducing latency from milliseconds to microseconds. This persistent connection proved essential in educational gaming contexts where tutor intervention timing and student action monitoring required immediate responsiveness.

LMU handled bidirectional communication through distinct message types. Incoming messages from the tutor dashboard included screenshot requests for viewing learner screens and hint messages for providing real-time assistance. Outgoing messages encompassed initial connection authentication with user UUID (Universal Unique Identifier) and username, stage completion notifications with timing data, room navigation tracking, and screenshot responses in base64 format. Each message type was JSON-formatted and processed through appropriate handlers that triggered corresponding learning actions. Connection management features included automatic connection establishment upon user login, graceful disconnection handling

during application shutdown, and state-aware message sending that waited for connection readiness before transmission.

The WebSocket connection enabled a sophisticated monitoring and intervention system that activated when learners encountered difficulties during a learning session. The system specifically tracked user interactions with NPCs (both SCRIPT-NPCs and LLM-NPCs).

The monitoring system continuously tracked the time spent at each NPC interaction point through the system's internal timing mechanisms. When the duration exceeded predefined thresholds, suggesting that the learner may be struggling with the content, the WebSocket Manager transmitted a stage time notification to the tutor dashboard.

This notification, sent via the SendNotification method, included critical context: the learner's username for identification, the specific NPC type (SCRIPT or LLM), the quiz name linking to the current learning objective, and the elapsed time. This automated detection ensured tutors receive timely alerts about learners requiring assistance without needing to constantly monitor every individual learner's screen.

Through the application, instructors had access to a structured interface composed of key sections such as *Overview*, *Players*, and *Leaderboard*, enabling seamless navigation between modules for efficient monitoring of student progress and statistics.

On the *Overview* page, the tutor was presented with concise yet informative metrics in the form of dashboard cards, including the number of learners currently engaged in the system (calculated via the number of connected clients on the WebSocket server), the average task completion time, and overall task completion rates. These real-time indicators offered a high-level snapshot of learner engagement and performance trends.

The *Players* page allowed the tutor to monitor all active learners within the environment. At the top of the page, the tutor could filter users by group, thereby displaying only the users that belonged to a selected cohort. For each user, relevant metadata, such as username and the timestamp of their most recent in-system activity, were provided. In addition, a *Show Group Stats* feature was available, along with an icon on each user row, enabling the display of group-level and individual-level analytics, respectively.

The tutor was also provided with access to the *Performance Summary* feature, which leveraged the ChatGPT API to automatically generate a concise progress report. This report synthesized the learner's performance data in relation to predefined thresholds, offering actionable insights into individual achievement and areas requiring further attention (see Figure 4 for an example performance summary).

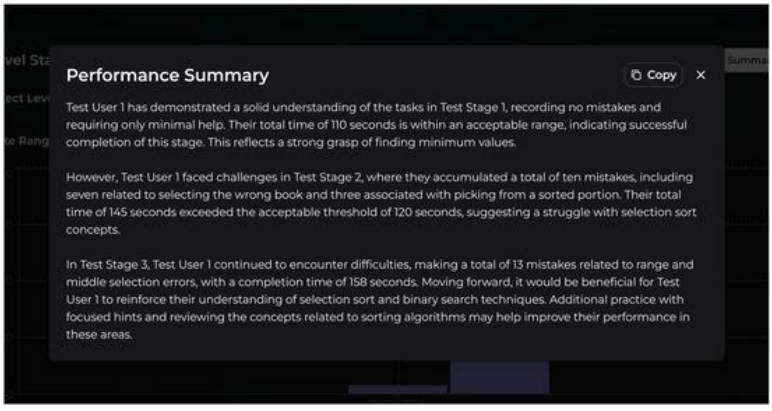

Figure 4: Performance summary example.

The *Leaderboard* page presented a structured table displaying users' performance results, enabling easy comparison and ranking based on selected criteria.

A bell icon was integrated into the interface, functioning as the entry point for the notification system to attract the tutor's attention. Clicking on the icon opened a sidebar containing all active notifications. Each notification was displayed in real time,

within a dedicated message card, whenever an issue arose during gameplay. The content of the notification included the following (Figure 5):

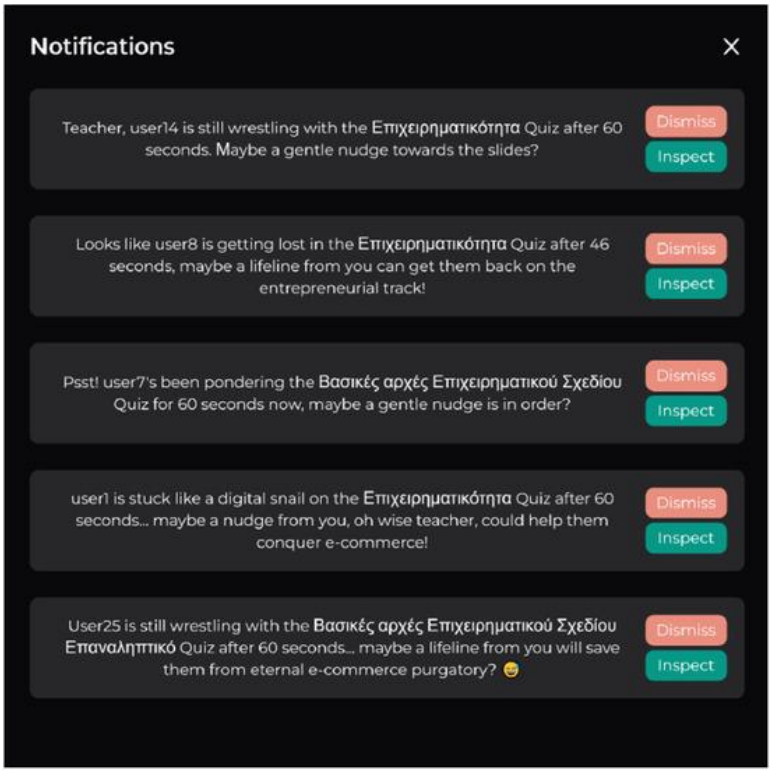

Figure 5: Notifications panel.

- a *personalized message* automatically generated with the support of the Gemini LLM, which described the detected issue (e.g., if a learner appeared to be "stuck" in a specific stage for an extended period);
- two action buttons: *Dismiss*, which removed the notification from the list, thus clearing the alert; and *Inspect*, which redirected the tutor to the *Inspect* page of the specific learner in question, enabling immediate review and analysis of the case.

This mechanism ensured that the tutor maintained a centralized, real-time overview of emerging difficulties faced by learners, while also enabling rapid, targeted intervention when and where it was most needed.

5. Learning Process and Learning Management

5.1. Game-Based Learning Process

The educational content of the VW was based on the content developed for the Entreality project (mentioned above), which was developed in the OpenSim platform. In this study, a new world was developed in Unity with different characteristics, but on the same educational subject, which concerned basic issues of entrepreneurship. Unity allowed for the use of more advanced capabilities in the visualization of the world and flexibility in the development of gamification features.

The virtual learning experience began with user authentication, involving options to sign up or log in. The system accommodated both single-player and multiplayer modes. The scene of the game, at this stage, consisted of five consecutive virtual rooms within a virtual building. Each room comprised either study material in the form of presentations on various panels (study room) or one or more quizzes (quiz room). A learner-player would have to go through all five rooms to be able to finish the game. There were two study rooms (first and third) and three quiz rooms (the rest). From a study room, the learner could proceed to the corresponding quiz room and attempt the quiz. Each time the user gave the right answer to a question, they received a point towards their score. To pass a quiz, more than 50% of the total quiz score should be achieved. Regardless of whether or not the learner attempts the quiz and, if so, to the result thereof, the learner could proceed to the next study room or return to the current study room. The fifth room housed the final quiz. At the end, each player received a total score of all quizzes, which represented his/her overall performance. Figure 6 illustrates the game-based learning process for the general case of N rooms. The solid lines indicate recommended paths.

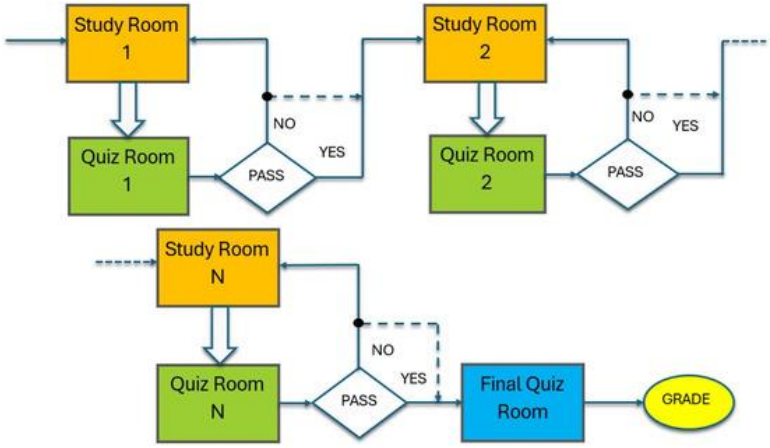

Figure 6: Game-based learning process flow.

As soon as a player entered a room, a notecard popped up at the lower right corner of the screen, displaying information about the educational process to be followed in the room. Additionally, in the upper right corner, a box appeared, displaying the current player's score and the highest score achieved so far by a player. During an intermediate quiz execution evolution, there were two NPCs whom the player could ask for help. The one to the right of the quiz panel was a SCRIPT-NPC, while the one to the left of the panel was an LLM-NPC.

5.2. Types and Use of NPCs

NPCs acted as helpers or guides within the game. They could provide advice, additional information, and facilitate the understanding of quiz questions, making the experience more interesting and interactive. To better guide and support learners, NPCs with two different levels of support were added: SCRIPT-NPC and LLM-NPC. Interaction with an NPC started when the learner approached the character, which triggered an action from the system.

5.2.1. SCRIPT-NPC

Approaching a SCRIPT-NPC triggered an interface displaying a set of predefined questions (see Figure 7). These questions were designed to align with the educational objectives of the game. The learner could select a specific question of interest, and the NPC responded with detailed information or guidance tailored to the selected query. This structured interaction ensured that learners received accurate and relevant information, helping them progress effectively in the game.

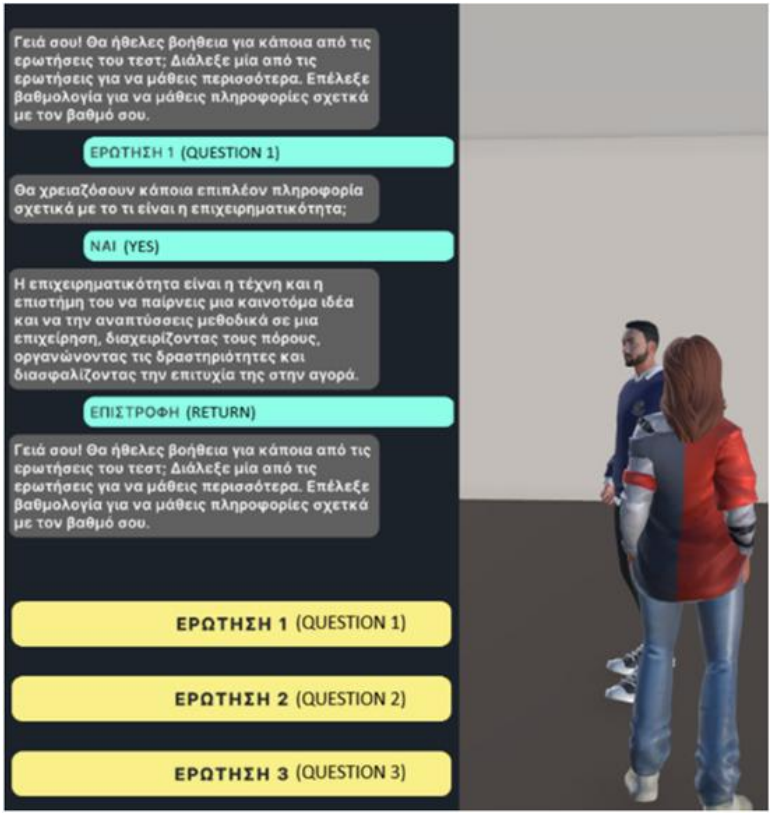

Figure 7: Conversation panel of a player with SCRIPT-NPC (the language displayed is Greek).

Incorporating an NPC with pre-defined questions and answers is a technique that offers significant benefits in educational games [5,9]. This method is particularly popular due to its ability to ensure consistency and provide clear guidance to players. The effectiveness of this type of NPC had been proven repeatedly, which determined the decision to implement them in the game. The main advantages were as follows: (a) using fixed questions and answers simplified interaction, (b) providing help and hints kept players constantly active, and (c) pre-defined questions and answers could ensure that specific educational concepts were systematically covered.

The SCRIPT-NPC played a critical role in guiding players through quizzes and providing assistance for educational activities. By limiting its responses to predefined content, the NPC ensured consistency and clarity in delivering instructions. The interaction prioritized usability, allowing players to easily navigate through questions and receive focused support without overwhelming them with unnecessary details.

Overall, an NPC with pre-defined questions and answers provided a simplified yet effective solution to player interaction. Using this NPC ensured that players always had the necessary support available for the game's questions. This approach guaranteed that every learner-player could proceed with confidence and overcome any questions related to the quizzes. The functionality of a SCRIPT-NPC was defined via a Unity script. The development of the NPC script was a time-consuming process that necessitated close collaboration with the creator of the educational content to ensure accuracy and proper support. It was necessary that the content creator contribute to the construction of a dialog tree, like the one depicted in Figure 8. This dialog tree was subsequently implemented within the script.

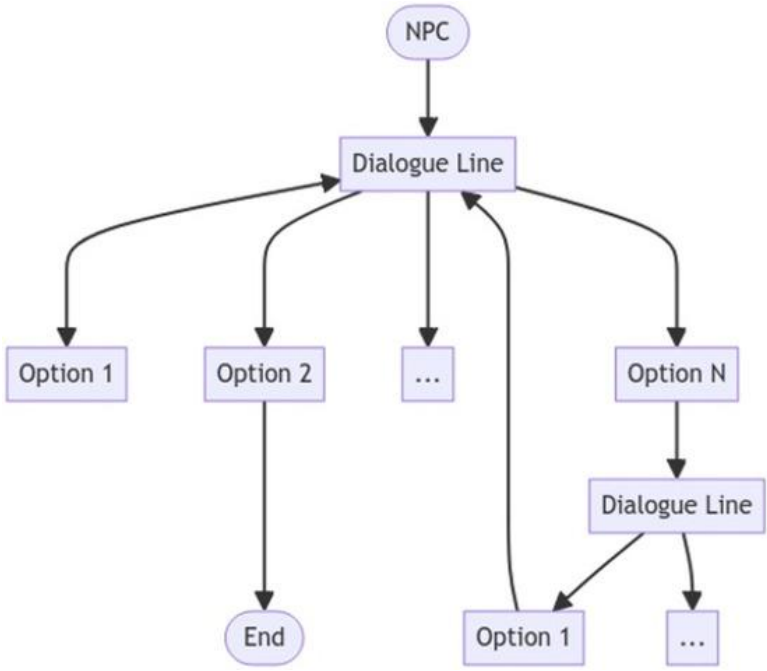

Figure 8: General dialog structure of NPCs.

The structure of a dialog in the database was organized in a multi-directional tree format. The root of the tree was an NPC, which was connected to a dialog line that had an unlimited number of options. These options could either be closing options that completed the flow of the dialog or options that led to the next dialog line, which in turn had its own options. Additionally, some options could return to a previous dialog line. This structure allowed for flexible and multidimensional management of dialogs, where each choice could lead to new paths or conclude the dialog session.

5.2.2. LLM-NPC

The interaction with the LLM-NPC was more dynamic and conversational. When the learner approached the NPC, a dialogue panel was activated, which allowed for free-form communication (see Figure 9). The LLM-NPC, powered by the advanced language model ChatGPT, could understand and

respond to a wide range of queries posed by the learner. This feature enhanced the immersive experience, as players could engage in natural language conversations, seek clarifications, and explore the subject matter in greater depth.

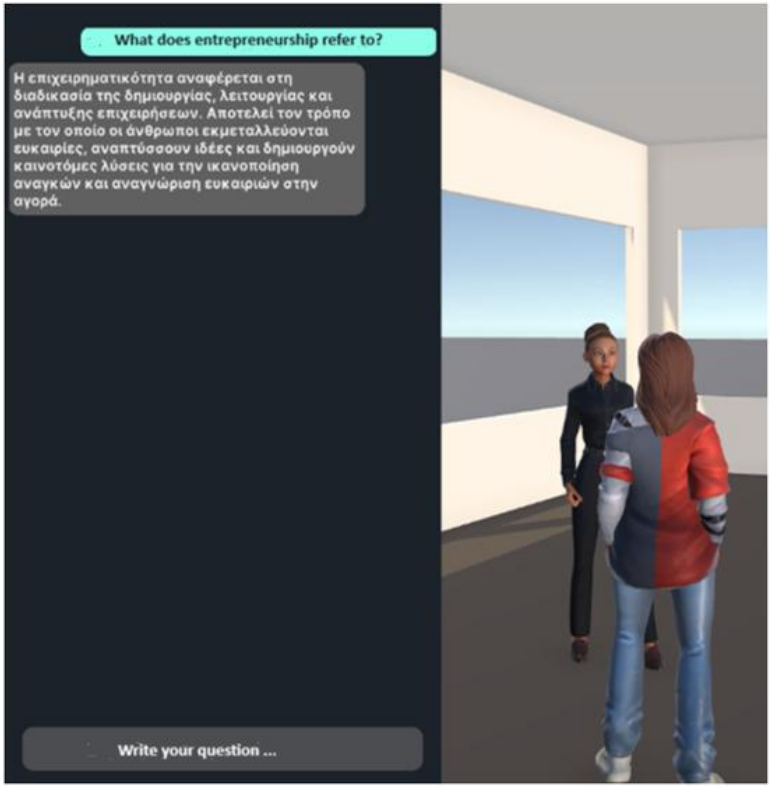

Figure 9: Conversation panel of a player with LLM-NPC (the language displayed is Greek).

Unlike the SCRIPT-NPC, the LLM-NPC adapted its responses based on the context of the conversation and the learner's individual needs. This flexibility allowed for addressing diverse learning styles and questions that might not be covered by structured NPCs. However, to maintain alignment with the educational objectives, the LLM-NPC was trained to provide responses relevant to the learning material and activities of the game.

For the LLM-NPC responses, the following prompt was given to ChatGPT: "You behave like a human being. You are a tutor specializing in business. Avoid answering questions unrelated to business content. If you are asked about your performance, if it is up to 50% you will recommend repeating the theory and a test, if it is from 51 to 70% you will recommend only theory and if it is above 71% you will congratulate."

The integration of an AI-powered NPC offered significant dynamics and advantages, which contributed to the decision to develop it. The main advantages were as follows: (a) enhanced realism, personalization, and engagement, (b) providing help and hints kept learners constantly active, and (c) acted as a source of knowledge, offering clarifications and further explanations to questions related to both the quizzes and the educational material, beyond predefined answers.

Overall, an AI-powered LLM-NPC could provide a more interactive, educational, and enjoyable gaming experience, which is especially important at a time when technology and digital interactions are becoming increasingly central to entertainment.

5.3. Tutor-Supported Learning

Upon receiving a difficulty notification, tutors could access an inspection interface within their dashboard that provided detailed information about the struggling student's current state. From this interface, tutors could request a real-time screenshot of the student's game screen to understand the exact challenge the student faced. The screenshot request would flow through the WebSocket connection as a "ScreenshotUnityRequest" type of message to the specific student's game instance. The WebSocket Manager, continuously monitoring for incoming messages, would receive this request and trigger the ScreenshotController component within the game to capture the current screen state.

The screenshot capture process occurred seamlessly without disrupting the student's gameplay experience. Once captured, the image was converted to base64 format for efficient transmission

and sent back through the WebSocket connection as a "ScreenshotUnityResponse" message. This visual feedback provided tutors with immediate insight into whether students were stuck on specific problems, misunderstanding instructions, experiencing technical issues, or simply needed encouragement to proceed. The non-intrusive nature of this screenshot mechanism preserved the student's autonomy while enabling targeted tutor support.

After assessing the student's situation through the screenshot, tutors could provide personalized assistance through the hint system. Tutors composed contextual hints or guidance messages tailored to the specific challenge observed (see Figure 10), which were transmitted through the WebSocket connection as "IncomingHint" messages. The WebSocket Manager processed these incoming hints by activating the game's help interface. When a hint arrived, the system activated the help panel UI element, making it visible to the student, and updated the text content with the tutor's guidance. Simultaneously, the system logged this help interaction through the LoggingService, recording which phase of the game triggered the assistance request, for later analysis of learning patterns and common difficulty points. This logging provided valuable data for curriculum improvement and identified content areas requiring additional instructional support.

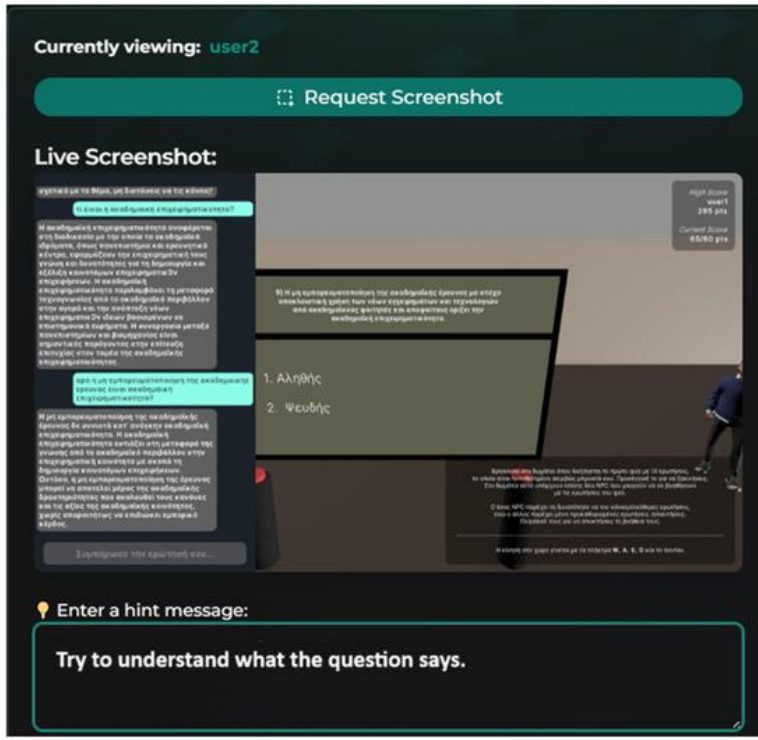

Figure 10: Tutor interaction screenshot for hint provision.

6. Experimental Evaluation

We conducted two evaluation experiments. In the first one, at an earlier stage of system development when the LMU was not ready, the participants used the tutoring system without the support of the tutor, i.e., having only the NPCs' support. In the second one, the participants had support from the tutor via the LMU as well. Given that the system was intended for university students and young people, we issued a relevant call for volunteers between 18 and 35 years old.

6.1. Early Evaluation Experiment

The evaluation procedure was implemented with the help of 34 participants/users, which allowed for the collection of an adequate number of answers and data to draw conclusions. From those, 47% (16) were males and 53% (18) were females, while

61.8% were between the ages of 18 and 25, and the rest were between 26 and 35. Half of the participants had prior experience with activities in virtual worlds. Also, 25 participants (73.5%) had limited knowledge of entrepreneurship aspects, and only 9 participants (26.5%) reported having substantial knowledge in the field.

The experiment took place in the presence of a tutor, so each participant was in the same room (face-to-face) as the tutor. Each participant was asked to complete a learning scenario, consisting of guided presentations and corresponding quizzes delivered through the virtual environment. At the beginning of the experiment, participants received both in-system instructions and verbal guidance. The verbal guidance focused on the user's movement and the need to interact with both types of NPCs. After the end of each participant's session, a questionnaire was completed by the participant. It was designed to obtain data on their experience with aspects of the tutoring system, as defined in Figure 1.

The answers' data revealed that most participants (94%) found the support provided by the environment and the NPCs to be either very or extremely helpful, with only two participants considering the support partially helpful. This indicates that the NPCs and the gaming feature were highly effective and positively received by most users. Furthermore, 97% of participants (33 participants) used both types of NPCs during their learning process.

More specifically, the questionnaire was targeted to evaluate the two types of NPCs across functional aspects, like usage and helpfulness, and affective aspects, like interest and friendliness. The results of the answers to the corresponding questions of the questionnaire are visually presented, in the form of pie charts, in Figure 11 and Figure 12. To extract valid conclusions, we applied a binomial (exact) test for each aspect, given their binary choices (for interest and usage, 'Equally' responses were excluded) and the small-to-modest size of the participant groups. The results are depicted in Table 1 and discussed below.

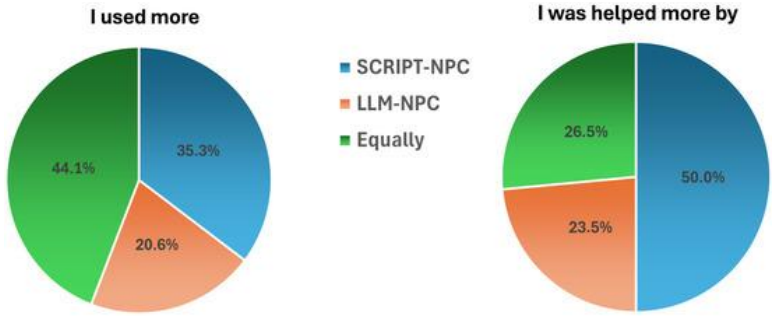

Figure 11: Results on functional aspects: usage (**left**) and helpfulness (**right**) of NPCs.

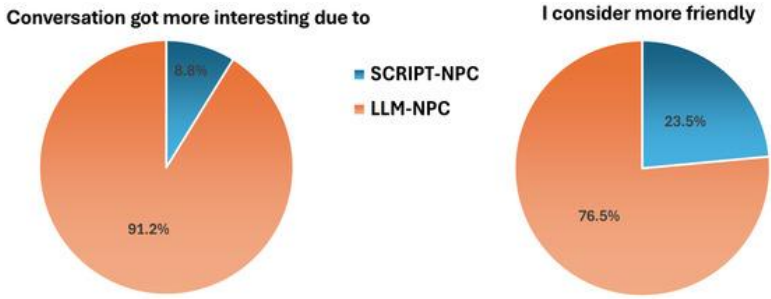

Figure 12: Results on affective aspects: interest (**left**) and friendliness (**right**) of NPCs.

Table 1: Early evaluation experiment and binomial test results for aspects of NPCs ($N = 34$).

Aspect	*SCRIPT-NPC n (%)*	*LLM-NPC n (%)*	*Equal n (%)*	*p Value*
Usage	12 (35.3)	7 (20.6)	15 (44.1)	0.18
Helpfulness	17 (50.0)	8 (23.5)	9 (26.5)	0.054
Interest	3 (8.8)	31 (91.2)	—	<<0.001
Friendliness	8 (23.5)	26 (76.5)	—	0.0015

Usage: Fifteen students (44.1%) reported equal usage of both NPCs, twelve (35.3%) reported using SCRIPT-NPC more, and seven (20.6%) reported using LLM-NPC more. Among students who perceived a difference in usage (i.e., excluding those who chose 'Equally'), the exact binomial test showed no statistically significant preference between SCRIPT-NPC and LLM-NPC ($p = 0.18 > 0.05$). (RQ2)

Helpfulness: Nine students (26.5%) reported that both NPCs were equally helpful. Among the remaining students, seventeen (50.0%) rated SCRIPT-NPC as more helpful, likely due to its predefined, more targeted responses for quiz-related questions, while eight (23.5%) rated LLM-NPC as more helpful. The exact binomial test indicated that this difference *marginally* did not reach statistical significance ($p = 0.054 > 0.05$; one more vote required) in favor of SCRIPT-NPC. (RQ2)

Interest: Thirty-one students (91.2%) reported that LLM-NPC was more interesting, compared to three students (8.8%), who preferred SCRIPT-NPC. This difference is statistically significant according to the exact binomial test ($p << 0.05$). (RQ3)

Friendliness: Twenty-six students (76.5%) rated LLM-NPC as more friendly, while eight students (23.5%) preferred SCRIPT-NPC. The exact binomial test indicated that this difference is statistically significant ($p = 0.0015 < 0.05$), suggesting a clear preference for LLM-NPCs in terms of perceived friendliness, probably appreciating its ability to enhance conversational engagement and interaction with the virtual world. (RQ3)

Interest motivating sources: The results of a multiple-response question about which aspect of the tutoring system (content, quizzes, SCRIPT-NPC, LLM-NPC, scoring, environment, high score) increased the interest of the learners are depicted in **Figure 13**. Notice that in this type of question, the participant could pick more than one of the possible options. Given that the number of selections ($17 + 17 + 6 + 12 + 23 + 1 + 1 = 77$) was larger than the number of participants (34), it meant that most (if not all) of them made more than one selection. From the results, the scoring system accounted for the largest proportion of selections (29.8%), equally followed by content and quizzes (22.1%), next followed by LLM-NPC (15.6%) and SCRIPT-NPC (7.8%). Percentages refer to the total number of selections. This distribution suggests that gamification-related elements and assessment activities, alongside content, played a prominent role in sustaining learner interest during learning activities. (RQ4)

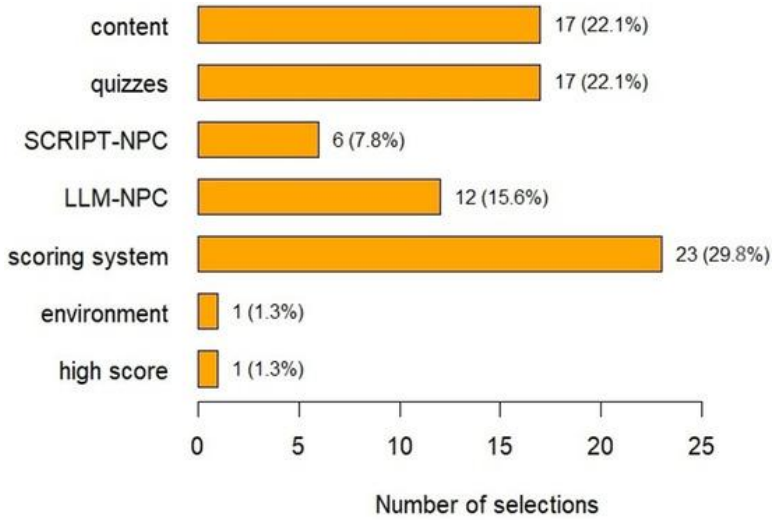

Figure 13: Results on the reasons for enhancing interest.

On the other hand, the data recorded in the database provided valuable information about users' interactions with NPCs. Specifically, it showed the total number of messages users sent to LLM-NPCs and the number of users who selected options from SCRIPT-NPCs. In addition, it tracked how many times specific questions were selected by SCRIPT-NPCs, allowing the tutor to identify which questions may have been challenging for the learners. The average score of each quiz was also available, revealing that the two quizzes with the highest average scores were supported by both types of NPCs, not just LLM-NPCs as in the other two quizzes. This suggests that SCRIPT-NPCs significantly helped learners to answer the quiz questions correctly, as confirmed by the questionnaire results, probably because the material focused specifically on quiz-related questions.

6.2. Late-Evaluation Experiment

A total of 30 individuals participated in the second evaluation, selected to ensure diversity in digital proficiency and academic background, consistent with the sampling criteria used in the first evaluation. Of those, 90% were between the ages of 18 and 25,

and the rest were between 26 and 35. Also, 70% identified as male and the rest as female. In terms of prior familiarity with 3D virtual environments, 66.3% reported high or complete familiarity, while 36.7% indicated moderate familiarity. Regarding the domain of entrepreneurship, a significant majority (66.7%) stated that they had limited, partial, or no prior knowledge, suggesting minimal subject-matter background before engaging with the experience.

The experiment utilized a dual-computer setup: one workstation for the participant and one for the tutor. In this iteration, the educational platform included the LMU monitoring system, which allowed the tutor to observe learner activity and intervene in real time, based on real-time analytics and live notifications, deliver personalized hints, and track progress visually.

Each participant was asked to complete the same instructional sequence as in the initial evaluation, consisting of guided presentations and embedded quizzes delivered through the virtual environment. While participants engaged with the content, the tutor used the monitoring system to observe their progress. In cases where learners appeared to struggle—based on system alerts or tutor judgment—contextual hints were sent through the interface to assist them. These interventions were non-intrusive and maintained the immersive nature of the activity.

In addition, after completing their sessions, participants filled out a structured questionnaire, similar to the one in the first evaluation, but containing extra questions, designed to assess perceived usability, clarity of instruction, effectiveness of support mechanisms (both NPC and instructor-based), and overall learning experience. The results of the answers to select questions from the questionnaire are visually presented, in the form of pies, in Figure 14, Figure 15, Figure 16, Figure 17 and Figure 18.

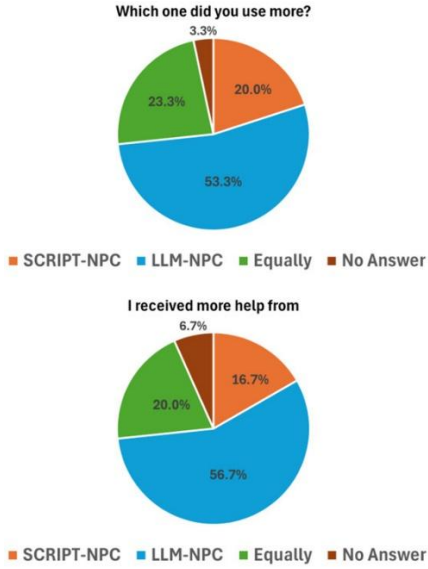

Figure 14: Results on functional aspects: usage (**top**) and helpfulness (**bottom**) of NPCs.

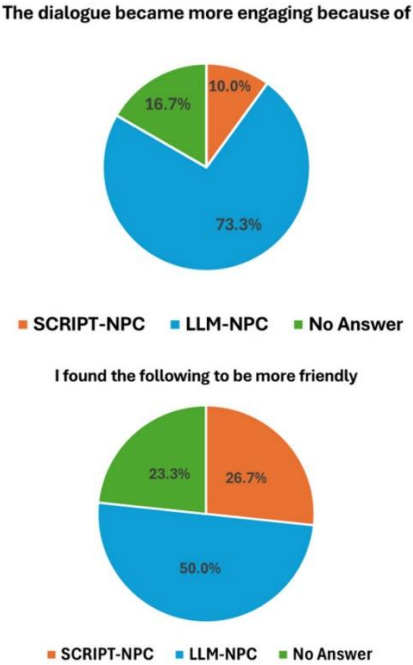

Figure 15: Results on affective aspects: interest (**top**) and friendliness (**bottom**) of NPCs.

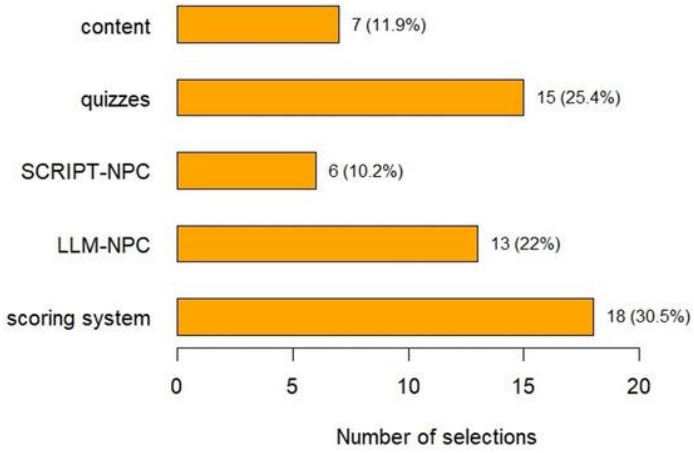

Figure 16: Results on the reasons for increasing interest.

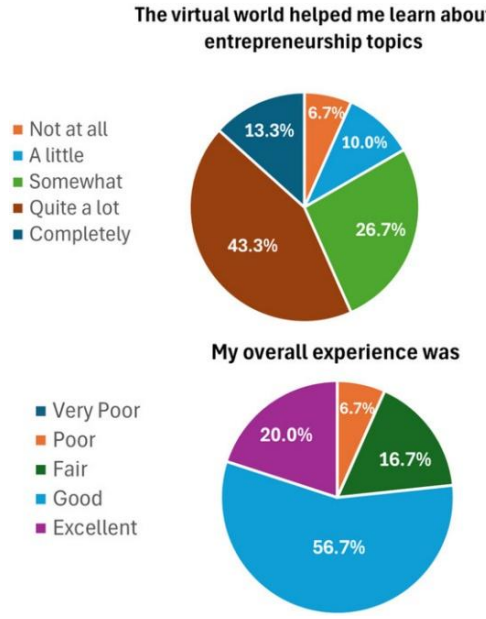

Figure 17: Results on system-level aspects: learning outcome (**top**) and overall experience (**bottom**).

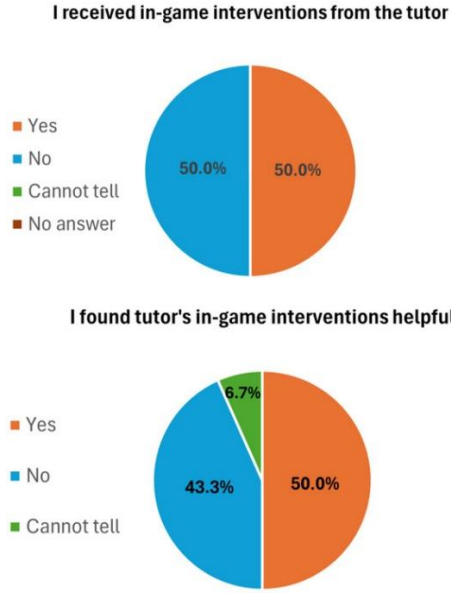

Figure 18: Results on tutor-related aspects: intervention (**top**) and helpfulness (**bottom**).

Again, as in the early evaluation, the questionnaire included four questions to evaluate the two types of NPCs across usability aspects, like friendliness and usage, and affective aspects, like interest and helpfulness. The results of the answers to the corresponding questions from the questionnaire are visually presented, in the form of pie charts, in Figure 14 and Figure 15. To extract valid conclusions, we again applied a binomial (exact) test. Notice that binomial tests were conducted, excluding 'Equally' and 'I prefer not to answer' responses to get samples with binary choices. The results are depicted in Table 2 and discussed below.

Table 2: Late-evaluation experiment and binomial test results for aspects of NPCs ($N = 30$).

Aspect	*SCRIPT-NPC n (%)*	*LLM-NPC n (%)*	*No Answer n (%)*	*Equal n (%)*	*p Value*
Usage	6 (20.0)	16 (53.3)	1 (3.3)	7 (23.3)	≈0.023
Helpfulness	5 (16.7)	17 (56.7)	2 (6.7)	6 (20.0)	≈0.008
Interest	3 (10.0)	22 (73.3)	5 (16.7)	—	<<0.001
Friendliness	8 (26.7)	15 (50.0)	7 (23.3)	—	0.105

Usage: Sixteen students (53.3%) reported using LLM-NPC more, six students (20.0%) reported using SCRIPT-NPC more, seven students (23.3%) reported equal usage, and one student (3.3%) selected "prefer no answer." Among students who perceived a difference in usage, an exact binomial test revealed a statistically significant preference between the NPCs ($p \approx 0.023 < 0.05$) in favor of LLM-NPC. (RQ2).

Helpfulness: Seventeen students (56.7%) rated LLM-NPC as more helpful, five students (16.7%) rated SCRIPT-NPC as more helpful, six students (20.0%) reported equal helpfulness, and two students (6.7%) selected "prefer no answer." Among students who expressed a preference, an exact binomial test indicated a statistically significant preference for LLM-NPC ($p \approx 0.008 < 0.05$). (RQ2).

Interest: Twenty-two students (73.3%) reported that LLM-NPC was more interesting, three students (10.0%) preferred SCRIPT-NPC, and five students (16.7%) selected "prefer no answer." Among students who expressed a preference, an exact binomial test showed a statistically significant preference for LLM-NPC ($p << 0.05$). (RQ3).

Friendliness: Fifteen students (50.0%) rated LLM-NPC as more friendly, eight students (26.7%) preferred SCRIPT-NPC, and seven students (23.3%) selected "prefer no answer." Among students who expressed a preference, an exact binomial test indicated no statistically significant difference between the friendliness rating of SCRIPT-NPC and LLM-NPC ($p = 0.105 > 0.05$). (RQ3).

Interest motivating sources: In Figure 15, the results of a similar multiple-response question, as in the early evaluation, about which aspect of the tutoring system (content, quizzes, SCRIPT-NPC, LLM-NPC, scoring system) increased the interest of the learners are depicted. Notice that the two almost ignored aspects from the early evaluation results (environment, high score) were removed. Given that the number of selections ($7 + 15 + 6 + 13 + 18 = 59$) was larger than the number of participants (30), it meant that most (if not all) of them made more than one selection. However, this time, there were fewer average selections per participant ($59/30 \approx 2$ vs. $77/34 \approx 2.3$).

From the results, a quite similar pattern is observed, where the scoring system was assigned the largest proportion of selections (30.5%), followed by quizzes (25.4%), LLM-NPC (22%), and content (11.9%), leaving SCRIPT-NPC last (10.2%). Again, these findings indicate a stable contribution of gamified elements across experiments, alongside an increased role of LLM-NPC and instructional content in supporting learner interest. (RQ4).

Additionally, to assess the effectiveness of the system, questions concerning the learning outcome and the overall experience of learners in dealing with the tutoring system were included, using a 5-level ordinal-type answer selection. The questions, their possible answers, and the results are depicted in Figure 17. To validate the results, we used a one-sample, one-tailed, exact Wilcoxon signed-rank test.

Learning Outcome (Subjective): To be able to apply the Wilcoxon test, we assigned ordered integer values to answer levels: 1–5 (Not at all–Completely), to simulate a 5-level Likert-like scale, with median = 3 (Somewhat) playing the role of neutral level. Our null and alternative hypotheses are as follows:

H_0. *Subjective learning outcome is equal to the median value (=3).*
Ha. *Subjective learning outcome is greater than the median value (>3).*

The data and the results of the Wilcoxon test are presented in Table 3. As can be clearly seen, participants reported that the system effectively supported learning about entrepreneurship topics. A one-sample Wilcoxon signed-rank test versus the neutral midpoint (Somewhat → 3) showed that perceived learning support was significantly above neutral (median = Lot → 4) ($p \approx 0.002 < 0.05$). In total, 57% of the participants rated the learning support as Lot or Completely, indicating a generally positive perception of learning. (RQ5).

Table 3: Data and Wilcoxon test results for subjective learning outcome assessment.

Aspect	**1** ***(Not at All)***	**2** ***(Little)***	**3** ***(Somewhat)***	**4** ***(Lot)***	**5** ***(Completely)***	***Median***	***Wilcoxon p***
Learning outcome	2	3	8	13	4	4	≈0.002

Table 4: Data and Wilcoxon test results for overall learner experience.

Aspect	**1** ***(Very Poor)***	**2** ***(Poor)***	**3** ***(Fair)***	**4** ***(Good)***	**5** ***(Excellent)***	***Median***	***Wilcoxon p***
Overall Experience	0	2	5	17	6	4	<<0.001

Overall Experience: To be able to apply the Wilcoxon test, we also assigned ordered integer values to answer levels: 1–5 (Very poor–Excellent), to simulate a 5-level Likert-like scale, again, with median = 3 (Fair) playing the role of neutral level. Our null and alternative hypotheses are as follows:

H_0. *Overall experience is equal to the median value (=3).*
Ha. *Overall experience is greater than the median value (>3).*

The data and the results of the Wilcoxon test are presented in Table 4. As can be clearly seen, the participants' overall experience with the system was highly positive. A one-sample Wilcoxon signed-rank test comparing responses to the neutral midpoint (Fair → 3) indicated that ratings were significantly higher than neutral (median = Good → 4) ($p << 0.05$). In total, 76% of participants rated their overall experience as Good or Excellent, indicating a generally positive perception of overall experience with the system. (RQ5).

Given the use of the LMU and the capability of the tutor to intervene and give support to the users, the questionnaire in this second experiment included extra questions about tutor-related aspects: frequency of intervention and helpfulness. Results show that 50% of the participants received some kind of support from the tutor (see Figure 18, top). On the other hand, 50% of the participants found the interventions of the tutor helpful, whereas 43.3% declared them to be not helpful, and the rest (6.7%) did not express any opinion (Figure 18, bottom). Looking at this result in relation to the previous one, it seems that from 50% of the participants who did not receive any support from the tutor, 43.3% of them answered "No" instead of "I cannot tell", misunderstanding the semantics of this option. Given this explanation, it means that 100% of those who received some support declared that it was helpful. (RQ6).

7. Discussion

7.1. Findings

Across both experiments, LLM-NPC was overwhelmingly perceived as more *interesting* than SCRIPT-NPC. In exp1 (N = 34), LLM-NPC was rated as more *friendly* by 76.5% of participants and more interesting by 91.2%, with both effects

reaching statistical significance. In exp2 ($N = 30$), the same directional trend was observed, although the friendliness difference did not reach statistical significance once "prefer no answer" responses were excluded, while the preference for LLM-NPC as more interesting remained statistically significant.

In contrast, perceptions related to *usage* and *helpfulness* were more nuanced. In exp1, no statistically significant differences were observed for either usage or helpfulness, despite the fact that there are descriptive trends favoring SCRIPT-NPC for helpfulness (only one more vote was missing to achieve statistical significance). In exp2, both showed statistically significant differences, favoring LLM-NPC.

To obtain a more holistic view of evaluation results, we aggregated the results from questions that were common between both questionnaires of the two experiments (Exp1, Exp2) and made two plots: a stacked bar plot (see Figure 19), concerning friendliness, interest, usage, and helpfulness related to the two types of NPCs; and a bar plot for the results of the question that concerned the sources of increased interest in the users, which are displayed in Figure 20. Notice that in Figure 19, only users who expressed opinions were considered. Also, in Figure 20, 'environment' and 'high score' aspects were not considered, given that they were not included in the second questionnaire.

Looking at Figure 19, in a combined mode, it is obvious that LLM-NPC dominates as far as friendliness and interest are concerned, although the result for friendliness in exp2 was not statistically significant (RQ3). However, the situation is different for usage and helpfulness, where LLM-NPC seems to have only a slight overall overhead (RQ2). Nevertheless, the results show that, although the introduction of LLM-NPCs is of great value, the involvement of traditional NPCs is still valuable in tutoring systems.

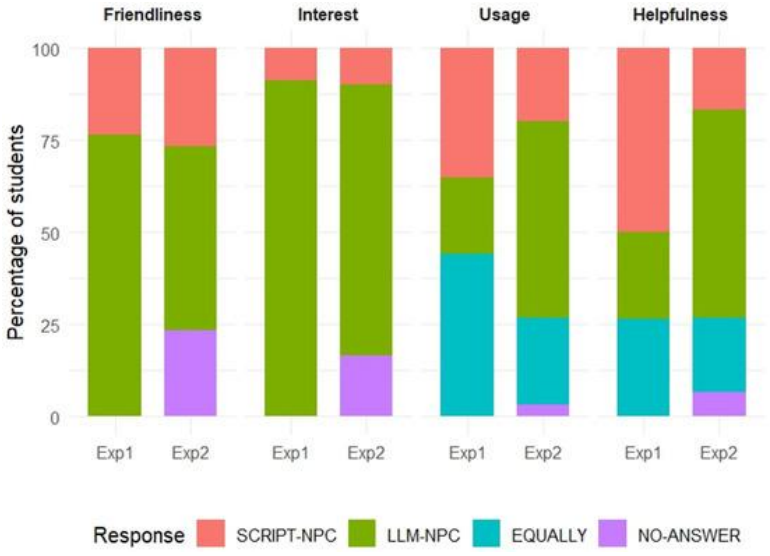

Figure 19: Aggregate results for 'friendliness', 'interest', 'usage', and 'helpfulness' of NPCs.

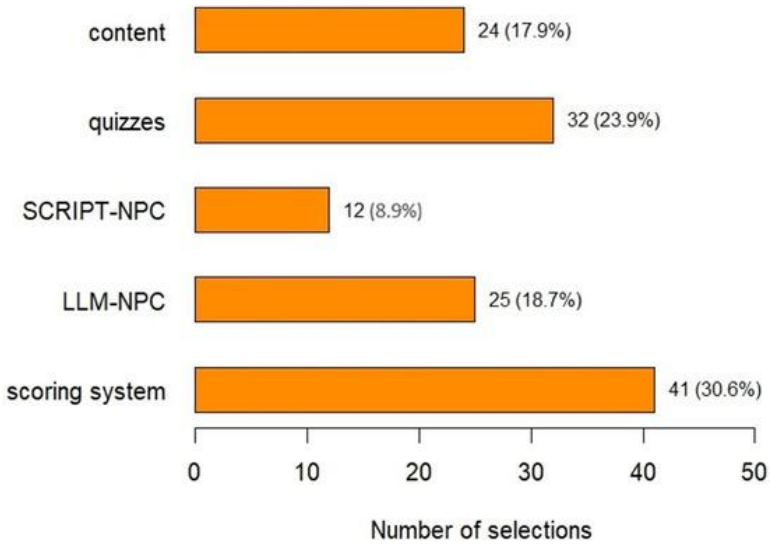

Figure 20: Aggregate results for educational features of the tutoring system.

From the bar chart in Figure 20, it is clear that gamification features of the tutoring system (scoring system, quizzes) dominated in increasing interest, hence user engagement with the LLM-NPCs and content came in second place (RQ4).

Overall, these findings suggest that while affective dimensions such as friendliness and interest were consistently associated with LLM-NPC, functional perceptions such as usage and helpfulness were more sensitive to contextual factors and questionnaire design.

The present findings provide converging evidence that learners consistently perceived LLM-NPC as more friendly and more interesting than SCRIPT-NPC across the two independent samples. These affective dimensions are closely related to constructs such as social presence and agent likability, which have been shown to influence learner engagement in adaptive e-learning environments. The robustness of these effects across experiments suggests that LLM-NPC's design features may have successfully conveyed a more engaging or socially appealing flavor (RQ3).

In contrast, perceptions of usage and helpfulness exhibited greater variability. The absence of consistent differences in reported usage suggests that learners did not systematically rely more on one NPC than the other, even when affective preferences were pronounced. Perceived helpfulness appeared to be particularly sensitive to contextual and methodological factors. While exp1 showed no significant difference between NPCs, exp2 revealed a significant preference for LLM-NPC. This discrepancy may reflect differences in how learners interpreted helpfulness when given expanded response options, or it may indicate that helpfulness judgments require more deliberate reflection than immediate affective impressions (RQ2).

Taken together, the results highlight an important distinction between affective and functional evaluations of NPC tutors, suggesting that adaptive e-learning systems should consider

these dimensions separately when designing and assessing virtual agents.

7.2. Limitations

Several limitations should be considered when interpreting the findings. In exp1, the sample size was modest ($N = 34$), which may have limited statistical power, particularly for detecting differences in usage and perceived helpfulness. Also, the measures relied on self-reported perceptions rather than objective performance or learning outcomes, reflecting subjective user experience rather than instructional effectiveness. Additionally, the inclusion of an "equal" response option, while informative, reduced the number of observations available for inferential testing in some comparisons. Finally, the study focused on short-term interaction with the NPCs; longer-term use may yield different patterns of preference or perceived utility. Future work should replicate these findings with larger samples, longitudinal designs, and objective learning measures.

In exp2, although the sample size ($N = 30$) was comparable to that of exp1, statistical power remained limited for detecting small differences, particularly when responses were distributed across multiple categories. The inclusion of a 'prefer no answer' option, while allowing participants to avoid forced choices, reduced the number of observations available for inferential analyses. As in the first experiment, outcomes were based on self-reported perceptions rather than objective usage logs or learning outcomes. Finally, the study captured short-term impressions of the NPCs; longer-term exposure may lead to different usage patterns or perceptions of helpfulness. Future work should investigate these effects using larger samples, longitudinal designs, and objective learning measures.

The present study focused on learners' perceptions of and interactions with non-player character (NPC) tutors, rather than on direct measures of learning outcomes. This design choice was intentional and aligned with the study's primary objective of evaluating user experience dimensions—such as friendliness, interest, usage, and perceived helpfulness—that are known to

influence engagement and acceptance of adaptive e-learning technologies. Prior research suggests that affective and experiential factors play a critical role in shaping how learners interact with instructional agents, particularly during early stages of adoption. As such, understanding these perceptions represents an important prerequisite for effective instructional design. While objective learning outcomes are a crucial component of comprehensive system evaluation, they were beyond the scope of the current exploratory investigation and are planned as a focus of future work building on the present findings.

7.3. Practical Application

The proposed virtual tutoring system has been designed to support integration into real educational settings rather than functioning solely as a research prototype. The system operates in a non-immersive virtual world environment and can be accessed using standard desktop or laptop computers, avoiding the financial and logistical constraints associated with immersive VR hardware. This design choice lowers institutional barriers to adoption and facilitates deployment in typical classrooms, computer labs, or remote learning contexts.

The system is well-suited to blended instructional models, where virtual-world activities complement lectures or online coursework. Tutors can assign learners to engage with NPC-guided tasks for practice or exploration, while retaining the ability to intervene in real time when learners encounter difficulties. This approach aligns with established teaching workflows and reduces instructor burden by combining automated guidance with selective human intervention. Recent higher education research has shown that virtual worlds and related metaverse environments extend social learning beyond traditional classroom settings and support collaborative interactions that can positively influence learner engagement and performance, highlighting the pedagogical value of socially rich virtual environments [32].

Scalability is also supported by the combined use of SCRIPT-NPCs and LLM-NPCs. SCRIPT-NPCs can manage routine instructional interactions efficiently, ensuring predictable behavior and low computational cost. LLM-NPCs are employed selectively for open-ended dialogue, adaptive feedback, and higher-level reasoning, where flexibility and natural language interaction are pedagogically beneficial. Recent literature on AI-powered educational agents highlights how hybrid systems that combine automated LLM-based guidance with human instructional oversight can enhance adaptive support while maintaining pedagogical control, a design principle also embodied in our tutoring system [33]. This hybrid architecture allows for a single human tutor to supervise multiple learners simultaneously, intervening only when it is necessary, which makes the system suitable for larger class sizes without proportionally increasing instructional workload.

From the perspective of learner acceptance, the positive questionnaire-based evaluation results suggest favorable perceptions of engagement, usefulness, and interaction quality. Recent educational technology research confirms that TAM remains a leading explanatory framework for learners' acceptance and use of instructional technologies, including digital learning tools and virtual learning environments [34]. Together, these findings indicate that the proposed system is not only technically viable but also pedagogically and organizationally suitable for real-world teaching scenarios, supporting its practical application and dissemination potential.

8. Conclusions

In this paper, we present the design and implementation of a curriculum-related tutoring system based on 3D Virtual World technology. It uses some kind of gamification and, compared to existing similar systems, it uses, apart from traditional NPCs (called SCRIPT-NPCs), a kind of intelligent NPC, called LLM-NPCs, which employ the ChatGPT (free version) LLM to manifest more natural dialogues with avatars (users). Also, the system is provided with a database, which stores educational content and user transactions with the system entities and the

NPCs, and a user interface for the tutors/tutors to manage the quizzes and gain insights from gathered statistics. Furthermore, it provides a learning management unit that traces the learning process online, via which the tutor can interact online with the learners and support them by giving hints.

Evaluation of the system reveals that LLM-NPCs, by providing personalized guidance and assistance to learners, foster a more engaging and friendly atmosphere that increases their interest and involvement in the educational environment. On the other hand, SCRIPT-NPCs are almost equally effective to LLM-NPCs in addressing specific questions, such as those related to quizzes, offering targeted support, but much more time is required for their development compared to that required for LLM-NPCs. However, somewhat surprisingly, evaluation of the educational features of the system showed that 'scoring system' was the most interesting for learners, with 'tests/quizzes' coming in second and LLM-NPCs in the third, along with 'content'. So, it seems that the use of gamification is very important.

Additionally, online progress tracking through the recording and analysis of students' movements and performance provides valuable data to educators, facilitating assessment and the adaptation of the learning process to meet learners' needs as well as providing support during learning sessions. Evaluation showed that all users involved in a tutor intervention were very satisfied.

However, the application of statistical tests for the validation of the conclusions did not show statistical significance in all cases, mainly due to the modest size of the samples of participants in the evaluation experiments.

Our future work will focus on a more robust and detailed evaluation of the system. A first direction would be to employ more participants and make an objective evaluation of the learning outcomes, using pre-tests and post-tests. Another possible direction is trying and testing various methods for ChatGPT prompt engineering, which may lead to the introduction of other types of LLM-based NPCs (e.g., based on thematic training). Also, the introduction of an extra module that

uses machine learning techniques could exploit the data in the database to produce useful models for user behavior recognition and automatic provision of help. Finally, alternative LLMs, like Gemini, could be employed and tested.

References

1. Alqahtani AS, Daghestani LF, Ibrahim LF. Environments and System Types of Virtual Reality Technology in STEM: A Survey. Int J Adv Comput Sci Appl. 2017; 8: 77–89.
2. Makransky G, Borre-Gude S, Mayer RE. Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. J Comput Assist Learn. 2019; 35: 691–707.
3. Wikipedia. Virtual World. Available online (accessed 12 February 2025).
4. Nowak KL, Fox J. Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. Rev Commun Res. 2018; 6: 201–232.
5. Klüwer T, Adolphs P, Xu F, Uszkoreit H, Cheng X. Talking NPCs in a Virtual Game World. In: Proceedings of the ACL 2010 System Demonstrations; Uppsala, Sweden; 13 July 2010. pp. 36–41.
6. Christopoulos A, Mystakidis S. Gamification in Education. Encyclopedia. 2023; 3: 1223–1243.
7. Karagiorgas DN, Niemann S. Gamification and Game-Based Learning. J Educ Technol Syst. 2017; 45: 499–519.
8. Maratou V, Xenos M, Vuckovic D, Granić A, Drecun A. Enhancing Learning on Information Security Using 3D Virtual World Learning Environment. In: Proceedings of the ICIST 2015 5th International Conference on Information Society and Technology; Istanbul, Turkey; 21–23 March 2015. pp. 279–284.
9. Grivokostopoulou F, Kovas K, Perikos I. The Effectiveness of Embodied Pedagogical Agents and Their Impact on Students Learning in Virtual Worlds. Appl Sci. 2020; 10: 1739.
10. Barmpari A, Voyiatzaki E, Hatzilygeroudis I. An Educational Virtual World System with Gamification Features and LLM Guided NPCs. In: Grafand S, Markos A, editors. *Generative*

Systems and Intelligent Tutoring Systems (ITS 2025). LNCS 15773. Cham, Switzerland: Springer Nature; 2025. pp. 213–223.

11. Grivokostopoulou F, Perikos I, Hatzilygeroudis I. An Innovative Educational Environment Based on Virtual Reality and Gamification for Learning Search Algorithms. In: Proceedings of the 2016 IEEE Eighth International Conference on Technology for Education (T4E); Mumbai, India; 2–4 December 2016. pp. 110–115.
12. Nunes FB, Herpich F, Zunguze MC, Nichele AG, Antunes FF, et al. A Virtual World for the Teaching and Learning of Natural Sciences. In: Proceedings of the EDULEARN 2017 Conference; Barcelona, Spain; 3–5 July 2017. pp. 6–13.
13. Kim H, Ke F. Effects of game-based learning in an OpenSim-supported virtual environment on mathematical performance. Interact Learn Environ. 2017; 25: 543–557.
14. Bai J, Xia C, Zhou Z, Zhu Y. Design and Implementation of a Biological Virtual Display System Based on 3DMax and Unity3D. In: Proceedings of the 2nd International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII); Melbourne, Australia; 12–14 June 2024. pp. 641–646.
15. Grivokostopoulou F, Perikos I, Kovas K, Paraskevas M, Hatzilygeroudis I. Utilizing Virtual Reality to Assist Students in Learning Physics. In: Proceedings of the 2017 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (IEEE TALE 2017); Hong Kong, China; 12–14 December 2017. pp. 486–489.
16. de Mattos DP, Popovici D. VR4STEM—A 3D Virtual World for Assisting Young People to Gain Entrepreneurship Skill in the STEM and ICT Domains. In: Proceedings of the 12th International Technology, Education and Development Conference (INTED 2018); Valencia, Spain; 5–7 March 2018. pp. 9322–9330.
17. Krassmann AL, Nunes FB, Bessa M, Tarouco LMR, Bercht M. Virtual Companions and 3D Virtual Worlds: Investigating the Sense of Presence in Distance Education. In: Zaphiris P, Ioannou A, editors. Learning and Collaboration Technologies. Ubiquitous and Virtual Environments for Learning and Collaboration (HCII 2019). LNCS 11591.

Cham, Switzerland: Springer Nature; 2019. pp. 175–192.
18. Grivokostopoulou F, Perikos I, Kovas K, Nikolic S, Paraskevas M, et al. Examining the Impact of Pedagogical Agents on Students Learning Experience in Virtual Worlds. In: Proceedings of the IEEE International Conference on Teaching, Assessment, and Learning for Engineering (IEEE TALE 2018); Wollongong, Australia; 4–7 December 2018. pp. 602–607.
19. Christopoulos A, Conrad M, Shukla M. Learner Experience in Hybrid Virtual Worlds: Interacting with Pedagogical Agents. In: Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019); Heraklion, Greece; 2–4 May 2019. pp. 488–495.
20. Guerra-Mota M, Minas D, Xenos M, Sá MM. Development of a 3D Virtual World Tool for Sustainable Energy Education. In: Proceedings of the 1st International Conference on Sustainable Energy Education (SEED 2024); Valencia, Spain; 3–5 July 2024. pp. 646–654.
21. Puttinaovarat S, Pruitikanee S, Kongcharoen J, Saeliw A, Inthong P, et al. The Digital Game for Curriculum Public Relations (PR) and Learning Using Unity3D. Int J Interact Mob Technol (iJIM). 2023; 17: 81–100.
22. Athanasiou P, Voyiatzaki E, Hatzilygeroudis I. Evolving Non-Player Characters in Educational Games in Virtual Worlds. In: Proceedings of the 15th International Conference on Education and New Learning Technologies (EDULEARN-23); Palma, Spain; 3–5 July 2023. pp. 7053–7059.
23. Huang M, Ma J, Bo S. Design and Implementation of a Multi-level Personalized Teaching Framework Based on LLM. In: Proceedings of the 14th International Conference on Educational and Information Technology; Guangzhou, China; 14–16 March 2025. pp. 8–14.
24. Khosrawi-Rad B, Shahda F, Robra-Bissantz S. Towards Pedagogical Conversational Agents as Creativity Drivers in Virtual Worlds. In: Proceedings of the 26th International Academic Mindtrek Conference (Mindtrek-23); Tampere, Finland; 3–6 October 2023.
25. Guevarra M, Bhattacharjee I, Das S, Wayllace C, Epp CD, et al. An LLM-Guided Tutoring System for Social Skills

Training. Proc AAAI Conf Artif Intell. 2025; 39: 29643–29645.
26. El Hajji M, Ait Baha T, Berka A, Ait Nacer H, El Aouifi H, et al. An Architecture for Intelligent Tutoring in Virtual Reality: Integrating LLMs and Multimodal Interaction for Immersive Learning. Information. 2025; 16: 556.
27. Evwiekpaefe AE, Darius Chinyio DT, Tohomdet LK. The Llama–ARCS Adaptive Learning Framework: AI–VR Integration System for Real-Time Motivational Feedback in Higher Education. J Comput Theor Appl (JCTA). 2026; 3: 260–273.
28. Peng X, Quaye J, Rao S, Xu W, Botchway P, et al. Player-Driven Emergence in LLM-Driven Game Narrative. In: Proceedings of the 2024 Conference on Games (CoG-24); Milan, Italy; 5–8 August 2024. pp. 1–8.
29. Christiansen FR, Hollensberg LN, Jensen NB, Julsgaard K, Jespersen KN, et al. Exploring Presence in Interactions with LLM-Driven NPCs: A Comparative Study of Speech Recognition and Dialogue Options. In: Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology (VRST-24); Trier, Germany; 9–11 October 2024. pp. 1–11.
30. Ogunlesia D, Wang X. GPT-NPC: Enhancing NPC Human-Likeness and Autonomy in Video Games. In: Proceedings of the ECAI 2024 Workshop on eXtended Reality & Intelligent Agents (XRIA 24); Santiago de Compostela, Spain; 20 October 2024.
31. Xiao T, Zhu J. Foundations of Large Language Models. arXiv. 2025; arXiv:2501.09223.
32. Lin LH, Pryor MR, Beckmann N. Social opportunities, learning practices, and performance in metaverse and virtual world: A comparative scoping review in higher education. Comput Educ. 2025; 239: 105391.
33. Córdova-Esparza DM. AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. Information. 2025; 16: 469.
34. Zawacki-Richter O, Jung I, editors. Handbook of Open, Distance and Digital Education. Berlin/Heidelberg, Germany: Springer. 2023.