information

MDPI

Article

The Creation and Evaluation of an AI Assistant (GPT) for Educational Experience Design

Antonio Julio López-Galisteo [1] and Oriol Borrás-Gené [2,*]

[1] Department Education Science, Universidad Rey Juan Carlos, 28032 Madrid, Spain; antoniojulio.lopez@urjc.es
[2] Department Computer Sciences and Statistics, Universidad Rey Juan Carlos, 28032 Madrid, Spain
* Correspondence: oriol.borras@urjc.es

Abstract: The emergence of generative artificial intelligence (GAI) has revolutionized numerous aspects of our lives and presents significant opportunities in education. However, specific digital competencies are essential to effectively leverage this technology's potential. Notably, prompt engineering proficiency presents a significant barrier to achieving optimal outcomes. In response, various solutions are being developed, including custom GPTs available through OpenAI's ChatGPT platform. This study validates 'GamifIcA Edu', a specialized GPT-based assistant for gamification and serious games, designed to enable educators to implement these pedagogical approaches without requiring advanced prompt engineering expertise. This is achieved through the utilization of pre-designed instructional frameworks. The assistant's effectiveness was evaluated using a comprehensive rubric across five distinct use-case scenarios. Each scenario underwent four different tests, representing varied learning contexts across multiple academic disciplines. The validation methodology involved a systematic assessment of the assistant's performance in diverse educational settings. The findings demonstrate the successful implementation of this custom-designed GPT, which generated contextually appropriate responses through natural language interactions, thus eliminating the need for complex prompt structures. This research highlights the potential of instruction-based design in the development of AI assistants that empower users with limited prompt engineering knowledge to achieve expert-level results. These findings have significant implications for the democratization of AI-enhanced educational tools.

Keywords: generative artificial intelligence; higher education; gamification; serious games

Academic Editors: Petros Lameras, Sylvester Arnab and Panagiotis Petridis

Received: 2 January 2025
Revised: 28 January 2025
Accepted: 5 February 2025
Published: 7 February 2025

Citation: López-Galisteo, A.J.; Borrás-Gené, O. The Creation and Evaluation of an AI Assistant (GPT) for Educational Experience Design. *Information* **2025**, *16*, 117. https://doi.org/10.3390/info16020117

1. Introduction

Generative artificial intelligence (AI) has experienced significant advancements in recent years, particularly in the development of large language models (LLMs). These models possess the ability to process and generate human-quality text, enabling them to undertake a wide array of complex tasks. Through exposure to extensive datasets during pre-training, LLMs acquire the capacity to develop a nuanced understanding of language and its intricacies.

Artificial intelligence (AI) is recognized as a key driver of the Fourth Industrial Revolution [1], impacting diverse sectors [2] including education. Integrating AI into education presents both challenges and opportunities, particularly in transforming teaching methodologies and enhancing learning efficiency [3,4]. The term AIEd is now widely used to denote the application of these technologies within educational contexts [5].

UNESCO [6] emphasizes the potential of AI to democratize knowledge and personalize learning experiences. However, it cautions that the implementation of AI in education

must be guided by principles of inclusion and equity to mitigate the risk of exacerbating existing educational inequalities. This is particularly crucial in contexts with significant technological disparities, privacy concerns, and potential data biases.

In the ongoing pursuit of effective and motivating educational methods, active methodologies have been increasingly implemented in classrooms. By placing students at the center of their learning journey, these methodologies foster a more engaging and impactful educational experience. Active methodologies are defined as pedagogical approaches that integrate traditional pedagogical principles with innovative practices, enabling students to experiment, reflect, and internalize knowledge through meaningful and contextualized processes. These processes often utilize real or simulated experiences [7–10]. This student-centered approach fosters autonomy, creativity, and critical thinking in future professionals, empowering them to actively engage with and transform their reality [11].

Hwang et al. [12] have identified four primary roles for AI in education: intelligent tutor, tutee, learning tool or companion, and decision advisor. Through various AI models and applications, educators can generate diverse resources, including text, images, and videos. AI also enables educators to automate tasks such as delivering tutorials, creating personalized content, and conducting assessments [5,6]. This technology offers opportunities to adapt educational content to individual student needs, particularly for learners with disabilities or those in remote areas with limited access to specialized teachers [6]. Furthermore, AI facilitates personalized learning pathways, improves accessibility, customizes teaching methodologies, fosters collaboration, supports adaptive learning environments, and promotes the development of essential skills, including critical thinking and problem-solving [4,13].

While the possibilities of AI in education are promising and seemingly endless, achieving optimal outcomes hinges on educators possessing a thorough understanding of both pedagogical and technical aspects [12,14]. Specifically, effective communication with AI models through prompts necessitates strong language skills, a nuanced understanding of the model's capabilities and limitations, and a deep knowledge of the educational context [15,16], and adequate digital skills are essential [6]. A poorly formulated prompt can lead to inaccurate, irrelevant, or ambiguous information [17]. Therefore, prompt design is critical for ensuring optimal model performance, and a proper understanding of the model's characteristics is key to crafting effective prompts. However, this remains a significant challenge for non-specialist educators, who often struggle to achieve satisfactory results when interacting with AI models [18,19].

Prominent active learning methodologies include project-based learning, case- and challenge-based learning, collaborative learning, the flipped classroom, and game-based learning [8,9]. Within game-based learning, two key strategies are particularly noteworthy: gamification and serious games.

Gamification is defined as the application of game design mechanics, elements, and techniques to non-game contexts, such as health, marketing, or education, to engage users and solve problems [20–22]. Key components of this technique include points, badges, leaderboards, levels, avatars, challenges, storytelling, and other interactive elements.

In contrast to gamification, which integrates specific game elements into non-game contexts, serious games are complete games (digital or otherwise) designed with primary purposes beyond entertainment [23]. From a simulation perspective, educational escape rooms and breakout EDU activities have gained significant popularity in recent years, yielding positive learning outcomes [24]. Educational escape rooms are immersive learning environments where students, typically organized in teams, must collaborate to solve challenges (often called puzzles) within a time limit to "escape" a designated space. These activities incorporate interactivity and engaging narratives to enhance learning [25]. Ed-

ucational breakouts, or breakout EDUs, was similar to escape rooms but with a different objective: instead of escaping a physical space, participants solve a series of problems within a time limit to decipher codes and unlock one or more locked boxes [26].

AI serves as a powerful tool to assist educators in managing and designing their courses. Whether an experienced educator seeking support and innovative ideas or a novice looking to implement new strategies in instructional design, AI models offer valuable support [4,27]. However, effectively leveraging these AI tools requires careful consideration of prompt design and construction. As previously discussed, a well-crafted prompt is crucial for achieving optimal results and avoiding inaccurate or irrelevant outputs.

One of the most recent and significant advancements in ChatGPT technology is the ability for users to design their own specialized AI assistants, known as GPTs or custom GPTs. This functionality allows for the creation of tailored AI models for more controlled generative text environments, trained on specific datasets and designed for specific purposes. An AI assistant is an entity (machine) capable of understanding user instructions, thanks to its underlying LLM, and applying pre-defined parameters and knowledge to achieve its designated objectives [28].

Sajja et al. [29] define an AI assistant as a computer system that utilizes AI techniques, such as natural language processing and machine learning, to understand, interpret, and respond to user requests. These assistants can perform a diverse array of tasks, including providing information, automating processes, offering personalized assistance, and facilitating interaction with other systems.

A significant limitation of utilizing LLMs in academia is their occasional generation of irrelevant or incorrect information. This can stem from the model's lack of specific knowledge, nuanced understanding, or analytical reasoning capabilities regarding the given prompt. Such inaccuracies can undermine the perceived usefulness, credibility, and trustworthiness of this technology in academic contexts [30,31]. However, this issue can be mitigated by creating a dedicated knowledge base for the AI assistant, as demonstrated by Castleman et al. [32], this knowledge base can be populated with relevant documents and data deemed essential by the creator, thereby enhancing the assistant's accuracy and effectiveness.

Training the AI assistant with this curated documentation significantly enhances the precision and relevance of its responses, ensuring better alignment with the knowledge contained within those resources. This, in turn, fosters greater user satisfaction and trust in the AI assistant's capabilities.

This method of creating customized AI applications significantly expands the possibilities of applying artificial intelligence in education [33]. These pre-designed AI assistants, with their specialized knowledge bases, reduce the need for users to possess advanced prompt engineering skills. This is a critical advantage, as educators with limited technical expertise can readily utilize these tools to enhance their productivity and teaching effectiveness [30]. In the educational context, AI assistants can support teachers in designing effective pedagogical strategies by providing relevant ideas, guidelines, and even customized learning content tailored to specific student needs.

Two primary challenges emerge in this context. Firstly, due to the novelty of AI in education, many educators lack the digital literacy and understanding required to effectively utilize AI tools, particularly in formulating effective prompts. Secondly, less experienced teachers often lack the pedagogical knowledge and confidence to design and implement game-based learning strategies within their subject areas. To address these challenges, an AI assistant was developed, following an extensive analysis of existing AI models, specifically ChatGPT [34–36]. This AI assistant incorporates a dedicated knowledge base and precise instructions (a "megaprompt") to facilitate effective AI utilization in

education. The primary objective was to provide educators with a user-friendly tool that empowers them to leverage the benefits of AI without requiring expertise in prompt engineering, gamification, or serious game design. This tool aims to guide and assist teachers in creating engaging, gamified classroom scenarios, enabling them to design effective teaching strategies based on various game-based learning approaches.

This educational research presents the results of evaluating the effectiveness of the AI assistant. A rubric serves as the primary evaluation instrument. The study encompasses five distinct user interaction scenarios, with four tests conducted for each scenario. These scenarios feature diverse learning situations contextualized within various subjects of university degree and master's programs.

This educational research investigates the effectiveness of a specialized AI assistant, 'GamifIcA Edu', designed to democratize the implementation of gamification and serious games in educational settings. By removing the need for specialized technical knowledge in AI and game-based learning, this tool enables educators with varying levels of expertise to readily integrate innovative pedagogical approaches into their teaching practices. This study evaluates the efficacy of 'GamifIcA Edu' in supporting educators and enhancing student learning experiences through gamified activities.

Based on this experience, the following research questions are proposed:

1. In what ways does the AI assistant ensure alignment with educational objectives when designing gamified activities?
2. How does the AI assistant maintain a consistent and coherent approach throughout the process of generating gamified learning activities?
3. How does the AI assistant balance consistency and flexibility when generating gamified activities?

The primary motivation of this research is to develop a GPT assistant that serves as a tutor for educators in designing and implementing gamified learning activities. This paper is organized as follows: Section 2 describes the functionalities of 'GamifIcA Edu' and the testing methodology. Section 3 from evaluating the AI assistant's performance are then presented. Section 4 analyses these findings, considering the potential applications of 'GamifIcA Edu' within diverse educational frameworks. Section 5 are also discussed. Finally, Section 6 summarizes the key findings and contributions.

2. Materials and Methods

This study explores the development of an AI-powered assistant designed to aid educators in integrating gamification and serious games into their classroom strategies. To evaluate the effectiveness of this assistant, its responses were analyzed across a range of potential classroom scenarios.

Figure 1 provides a visual summary of the research process described above, outlining the key steps involved in developing and evaluating the AI assistant, 'GamifIcA Edu', and highlighting the iterative nature of the design process:

1. Creation and configuration: The researchers designed and implemented the AI assistant ('GamifIcA Edu'), a custom GPT, using GPT 4o large language model, specifically trained on a curated knowledge base of resources related to gamification and serious games in education.
2. Scenario definition: A variety of potential classroom scenarios were defined to validate the assistant's capabilities, focusing on its ability to provide aligned, coherent, and consistent responses relevant to different educational contexts.
3. Testing and data collection: For each defined scenario, multiple tests were conducted with the AI assistant. The assistant's responses to various prompts within each scenario were documented and collected for analysis.

4. Evaluation: An evaluation rubric was developed and applied to analyze the collected responses. This rubric facilitated a comprehensive assessment of the assistant's performance across different scenarios, considering factors such as accuracy, relevance, creativity, and pedagogical soundness.

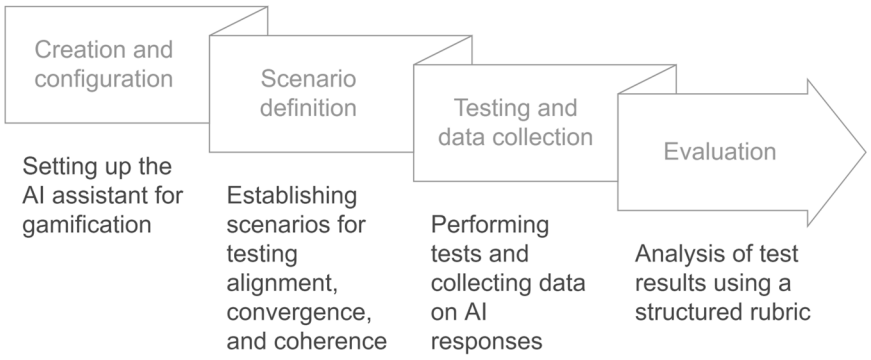

Figure 1. A flowchart illustrating the methodology developed for the design and validation of the 'GamifIcA Edu' assistant.

2.1. Creation and Configuration of Artificial Intelligence Assistant for Gamification and Serious Games 'GamifIcA Edu'

The GPT assistant, named 'GamifIcA Edu' [37], was developed using the ChatGPT Plus platform, a commercial version of ChatGPT that allows for the creation of customized GPTs with specific instructions, also known as artificial intelligence assistants. This version was chosen because it enables the development of assistants with tailored knowledge and capabilities. To prevent prior responses from influencing subsequent interactions and ensure unbiased testing, the assistant was configured to avoid storing past conversations.

'GamifIcA Edu' was designed within the ChatGPT Plus GPT-creation module using a series of detailed instructions. Its primary objective was to assist educators in designing gamified activities and serious games for educational purposes.

To enhance its expertise, the assistant was provided with a dedicated knowledge base comprising a curated collection of reference books and guides on gamification and serious games in education. These resources were used to train the underlying language model, enabling GamifIcA Edu to function as a specialized expert in this domain.

The assistant was programmed to perform a range of functions based on user requests:

- Create gamified activities (e.g., breakout EDUs, escape rooms) or serious games for a specific subject.
- Provide guidance to teachers on designing gamified activities or serious games for application in a particular subject or in general.
- Advise teachers on incorporating generative artificial intelligence into the development of gamified activities and serious games.
- Empower teachers to enhance their knowledge of gamified activities and serious games.

The 'GamifIcA Edu' assistant generated responses dynamically based on the input provided by the teacher during the chat conversation. To create gamified activities or serious games, the GPT leveraged the following key elements:

1. The teacher's input during the conversation with the assistant after configuration.
2. The instructions implemented in the assistant, following a pre-defined structure.
3. The dedicated knowledge base provided to the assistant.

The main features to consider when creating an assistant are as follows:

- Role: The role assigned to the GPT was that of an expert in designing gamification and serious games in education, as well as serving as an advisor during the creation process.
- Tone: The tone was academic, consistent with that of a teacher specializing in gamification and serious games.
- Instructions: These were defined following a logical and sequential structure, aligned with the type of responses expected from the automated assistant.

The main instructions are outlined below:

1. After the initial user interaction, the assistant introduces itself and outlines the types of activities it can create.
2. Once the user decides on the type of activity to create, the assistant asks a series of questions:
    a. Which subject would you like the breakout EDU to focus on?
    b. In which degree program is it taught, and for which academic year?
    c. If preferred, you may upload specific materials or notes to include in the activity.
    d. How long would you like the activity to last?
    e. What level of difficulty do you prefer (easy, medium, or challenging)?
    f. Do you have a specific theme or narrative in mind to provide context for the breakout EDU?
3. The teacher provides answers to the questions, and the assistant generates the activity based on the documents in its knowledge base, with which the language model has been trained.

2.2. The Definition of Various Scenarios for the Assistant's Validation

To evaluate the effectiveness of the 'GamifIcA Edu' assistant, we developed five scenarios designed to assess its operational characteristics in diverse educational contexts. These scenarios were constructed based on the following criteria:

1. Type of scenario: A series of scenarios were designed to reflect common situations, such as the creation of gamified activities, breakout EDUs, and escape rooms, as well as advisory requests that users might pose directly to the assistant.
2. Scenario objectives: For each scenario, specific objectives were defined to assess the assistant's ability to achieve the desired outcomes in its responses.
3. Scenario criteria: Various criteria were established to evaluate the results of the tests conducted in each specific scenario.

In each scenario, the objective was to verify that the assistant demonstrated the three main characteristics expected of an intelligent assistant:

- Alignment: The assistant performed the tasks for which it was designed with accuracy and relevance.
- Coherence: The assistant provided logical and consistent responses, both within a single interaction and across an entire conversation.
- Consistency: The assistant maintained uniformity in style, tone, and the type of responses delivered, both over time and across different interactions.

Table 1 summarizes the various scenarios designed for the validation of the assistant.

Table 1. The scenarios established for the validation tests of the 'GamifIcA Edu' assistant. Source: own elaboration.

Scenario	Scenario Objective	Expected Behavior of the Assistant	Validation Criteria for the Assistant
1	To verify the proper functioning of the GPT in creating gamified activities or serious games by uploading a file containing the topic or topics to be addressed in the activity.	The GPT must generate an activity accurately and coherently, aligning with the document uploaded by the user and incorporating the responses to specific questions posed by the GPT, such as the degree program, subject, and course in which the activity will be applied.	The created activity aligns with the topic provided in the uploaded document.
			The GPT demonstrates consistency in formulating questions and designing the activity based on the user's responses.
			The GPT maintains consistency in the suggestions provided throughout the tests.
			The GPT exhibits flexibility in accommodating the scope of activity creation.
2	To verify the correct functioning of the GPT when providing advice to the user on how to create gamified activities or serious games for application and development in a specific subject or in general.	The GPT must deliver high-quality advice to the user, adapting to a new interaction environment after requesting details such as the grade level and subject for which the activity is to be created.	The simulator demonstrates a high degree of flexibility when engaging with a new interaction.
			The GPT's feedback aligns with the principles of creating gamified activities.
			The GPT maintains consistency in the feedback it generates.
			The GPT demonstrates consistency and relevance in the feedback provided throughout the interaction.
3	To verify the correct functioning of the GPT when incorporating generative artificial intelligence into the development of gamified activities and serious games.	The GPT must create scenarios that integrate the application of artificial intelligence by students. It should ask the user for details about the subject matter and the specific AI tool they wish to use. Additionally, it must provide detailed answers, including solutions and configurations when requested, and be able to specify each test incorporated into the activity.	The GPT effectively incorporates artificial intelligence into the tests of the developed activity.
			The GPT demonstrates the ability to specify in detail each test included in the activity.
4	To verify the correct functioning of the GPT when tasked with developing gamified activities, breakout EDUs, escape rooms, or serious games based on scenarios defined by the user.	The GPT must create the gamified activity following the conditions (narratives or dynamics) specified by the user.	The development of a game-based or gamified activity based on a specific narrative provided by the user (e.g., Star Wars, Harry Potter).
			The development of a game-based or gamified activity based on a specific game dynamic defined by the user (e.g., Goose game, Monopoly).
			The creation of activities that combine both narrative and dynamics (e.g., a Monopoly game with a Harry Potter narrative).
5	To verify whether the assistant could detect and appropriately redirect an initial user interaction that is unrelated to the functions for which the GPT was designed.	The assistant must detect an initial user interaction that is not aligned with the primary functions for which it was designed.	The assistant identifies an interaction that deviates from its primary functions.
			The assistant guides the user back to the main functions for which it was designed.

2.3. Research Design

This study employed an exploratory qualitative approach to understand, evaluate, and validate the performance of 'GamifIcA Edu', an AI-powered assistant designed to support educators in integrating gamification and serious games into their pedagogical practices.

2.3.1. Evaluation Tests for the Assistant

The 'GamifIcA Edu' assistant was evaluated using a series of scenarios designed to represent specific use cases (Table 2). These scenarios reflect practical challenges and real-world questions that educators might encounter when implementing gamification strategies in their classrooms. Each scenario was tested four times with the assistant under controlled conditions to prevent prior interactions from influencing the responses. This approach ensured consistency, coherence, and adaptability in the assistant's output. To assess the assistant's versatility, each interaction varied in terms of the type of gamified activity to be created and the subject area in which it would be applied.

Table 2. Tests conducted on assistant. Source: own elaboration.

Test	Subject	Level	Duration of the Activity (minutes)	Scenario	Use Cases
1	Educational Innovation and ICTs	Master	90	1	Breakout EDU
				2	Serious game
				3	Gamified activity on a specific subject topic + GenAI
				4	Serious game with a dynamic similar to the Goose game
	-	-	-	5	Request for information on URJC study plans
2	STEM Project Design	Degree	50	1	Breakout EDU
				2	Gamified activity
				3	Breakout EDU + GenAI
				4	Serious game with a dynamic similar to the Goose game
	-	-	-	5	Request for the creation of a didactic unit
3	Introduction to Programming	Degree	90	1	Breakout EDU
				2	Gamified activity
				3	Serious game + GenAI
				4	Serious game with a dynamic similar to Ludo
	-	-	-	5	Request for a brownie recipe
4	Introduction to Programming	Degree	90	1	Breakout EDU
				2	Gamified activity
				3	Gamified activity on a specific subject topic + GenAI
				4	Serious game with a dynamic similar to Ludo
	-	-	-	5	Request for information about Jean Piaget

2.3.2. Design and Application of an Evaluation Rubric for the Assistant

Following the testing of each scenario, an analytical rubric (Table 3) was developed to systematically evaluate five key dimensions of the assistant's performance:

1. Alignment with the principles and specific objectives: Evaluates the adequacy of the responses or activities generated in terms of their compliance with the objectives and principles defined in each scenario.

2. Coherence in the responses or activities generated: Assesses the logic and consistency of the responses or activities designed according to the teacher's requirements.
3. Flexibility and adaptability: Measures the assistant's ability to adjust to different environments, instructions, or user requirements.
4. Consistency in the suggestions or solutions offered: Determines whether the suggestions are consistent across interactions and relevant to the defined objectives.
5. Clarity of interaction and feedback: Evaluates whether the interactions are clear and if the feedback effectively guides the user.

Table 3. The general analytical rubric for the analysis of responses obtained in the different validation scenarios of the 'GamifIcA Edu' assistant. Source: own elaboration.

Evaluation Criteria	Level 1 (Inadequate, 0 Points)	Level 2 (Basic, 1 Point)	Level 3 (Adequate, 2 Points)	Level 4 (Excellent, 3 Points)	Overall Weighting
Alignment with the principles and specific objectives	Not aligned, response does not align with the principles or objectives of the scenario	Limited alignment, response shows partial alignment but contains obvious errors	Mostly adequate alignment, response is mostly aligned, with minor misalignments	Fully aligned, response is completely aligned and relevant to the scenario	25%
Consistency in responses or activities generated	Inconsistent or contradictory responses	Some responses show coherence, but with inconsistencies	Mostly coherent responses, with slight areas for improvement	Fully coherent and accurate responses	25%
Flexibility and adaptability	Does not adapt to context or new environments	Limited adaptation with frequent errors	Acceptable adaptation with some necessary adjustments	Highly adaptable and flexible depending on the context	20%
Consistency in suggestions or solutions offered	Inconsistent or irrelevant suggestions	Some suggestions are useful, but not consistently so	Mostly consistent and useful suggestions	Suggestions are totally consistent and relevant	20%
Clarity of interaction and feedback	Confusing or unclear feedback	Somewhat clear feedback, but with areas for improvement	Feedback is clear and understandable, but with room for improvement	Feedback completely clear and effective	10%

For each criterion within the rubric, four achievement levels were defined: Inadequate, Basic, Adequate, and Excellent. Table 3 details the specific conditions required to achieve each level for every criterion. This rubric was applied to each scenario to assess the performance of 'GamifIcA Edu' and points were assigned for each criterion based on the corresponding achievement level. These scores were then multiplied by pre-defined weights reflecting the relative importance of each criterion in evaluating the overall performance of the assistant. Finally, a weighted sum was calculated to obtain a total score for each scenario.

3. Results

This section presents the key findings from the evaluation of the AI assistant's performance across all defined scenarios, using the analytical rubric detailed in Table 3.

Notably, the AI assistant exhibits user-friendly interaction patterns that do not require complex prompts. This ease of use makes the assistant accessible to all educators, regardless of their prior experience with text-generative AI systems. The assistant provides relevant and appropriate responses through a straightforward conversational interface, enabling

users to express their needs and requests in simple terms, without the need for elaborate prompting strategies.

3.1. Scenario 1

This scenario aimed to evaluate the assistant's ability to create gamified activities and serious games based on the uploaded subject content files. The evaluation of four test cases (Table 4) revealed that the assistant consistently generated relevant and engaging gamified activities that were well-aligned with both the educational objectives and the user-provided content.

Table 4. Results obtained for Scenario 1 after analysis with general rubric.

Evaluation Criteria	Score Out of 100 Points
Alignment	96
Coherence	94
Flexibility	87
Consistency	87
Clarity	82
Final score	89.2

The assistant demonstrated a high degree of consistency in generating well-structured gamified activities and breakout EDUs that were both coherent and relevant to the requested subject matter. While the assistant exhibited considerable flexibility in responding to specific user requirements, there is potential for improvement in its recommendations for digital resources and activity implementation materials.

Overall, the assistant achieved a score of 89.2 out of 100 points on the evaluation rubric, indicating strong performance in this scenario.

3.2. Scenario 2

This scenario evaluated the AI assistant's capacity to provide guidance for creating subject-specific gamified activities and offer constructive feedback to users.

The evaluation of four test cases revealed that the assistant consistently generated relevant gamified activities aligned with user-defined objectives. Moreover, the assistant demonstrated a high level of coherence by developing activities with logical and progressive sequences, while maintaining flexibility in adapting to different thematic requirements.

While the assistant's guidance was generally effective, the inclusion of more visual examples in the instructional materials for some activities could further enhance user engagement and comprehension.

Based on the evaluation rubric (Table 5), the assistant achieved a score of 92/100, demonstrating exceptional performance in this scenario.

Table 5. Results obtained for Scenario 2 after analysis with general rubric.

Evaluation Criteria	Score Out of 100 Points
Alignment	97
Coherence	98
Flexibility	87
Consistency	90
Clarity	88
Final score	92

3.3. Scenario 3

This scenario assessed the AI assistant's capability to guide teachers in developing gamified activities and serious games that incorporate artificial intelligence in classroom settings.

The assistant consistently demonstrated strong alignment with educational objectives while creating logically structured gamified activities. Across all tests, the assistant maintained coherent educational experiences, providing clear instructions and resource recommendations that facilitated student engagement. However, the inclusion of additional visual examples in some activities could further enhance user comprehension.

A notable strength was the assistant's innovative integration of generative AI to enhance learning outcomes. Nevertheless, there is potential for improvement in making the activities more flexible to accommodate subjects with limited technological requirements.

The assistant's effectiveness in this scenario is evidenced by a high score of 92.2 out of 100 points on the evaluation rubric (Table 6).

Table 6. Results obtained for Scenario 3 after analysis with general rubric.

Evaluation Criteria	Score Out of 100 Points
Alignment	97
Coherence	98
Flexibility	88
Consistency	90
Clarity	88
Final score	92.2

3.4. Scenario 4

This scenario evaluated the assistant's effectiveness in creating gamified activities based on user-specified storytelling or game dynamics.

The assistant demonstrated exceptional performance in both alignment and coherence, successfully adapting activities to specific learning objectives while maintaining a logical structure within the proposed narrative. Throughout the evaluation, the assistant consistently delivered robust and meaningful proposals.

While the assistant provided clear descriptions and practical examples, the activities could be further refined to better accommodate environments with limited resources and diverse learner experience levels.

The assistant's proficiency in this scenario is reflected in its evaluation score of 92.4/100, as detailed in Table 7.

Table 7. Results obtained for Scenario 4 after analysis with general rubric.

Evaluation Criteria	Score Out of 100 Points
Alignment	97
Coherence	98
Flexibility	87
Consistency	90
Clarity	88
Final score	92.4

3.5. Scenario 5

Scenario 5's results revealed significant variations in the quality and alignment of the assistant's responses. While some tests, particularly those involving the creation of gamified learning units, demonstrated strong alignment and clarity, other responses showed limitations in contextualizing information and adapting to specific pedagogical

objectives. This variability was anticipated, as the scenario was intentionally designed to include requests unrelated to gamification, thereby testing the assistant's ability to handle off-topic queries.

4. Discussion

This section presents an analysis of the AI assistant (custom GPT)'s performance in supporting educators with gamification design, based on the application of a general evaluation rubric. The evaluation reveals a comprehensive understanding of the assistant's capabilities, strengths, and limitations. Through this analysis, we aim to address the research questions proposed for this study. The following discussion examines key findings across critical dimensions, including alignment, coherence, consistency, and flexibility.

4.1. In What Ways Does the AI Assistant Ensure Alignment with Educational Objectives When Designing Gamified Activities?

Across all evaluated scenarios, the assistant demonstrated adequate performance without requiring advanced prompt engineering knowledge. This finding aligns with previous research by Sánchez-Prieto et al. [38], which also observed user-friendly interaction patterns with AI assistants.

The alignment of text-generative AI assistants, such as GPTs, is crucial for ensuring that they meet their intended objectives and provide practical value. A well-aligned assistant generates accurate, relevant, and contextually appropriate responses while minimizing biases, errors, and undesirable behaviors. This alignment not only enhances the quality of interactions but also mitigates risks associated with misinformation, inappropriate content generation, and misinterpretation of instructions. In scientific, educational, and professional contexts, where accuracy and reliability are paramount, proper alignment enables these models to serve as dependable tools for complex problem-solving, decision support, and communication tasks. In education specifically, alignment is essential for providing effective guidance to teachers in developing and implementing gamified learning activities. Conversely, misalignment can undermine trust in the technology and limit its adoption in critical educational settings.

Fulgencio's [39] research indicates that assistant alignment depends on three key elements: a robust knowledge base, well-defined behavioral parameters in the assistant's core instructions, and clear response protocols for various user requests.

The evaluation rubric results demonstrated the AI assistant's exceptional alignment with these principles, achieving scores above 96/100 across all five test scenarios. The assistant consistently provided accurate information without errors, a crucial aspect achieved through well-designed instructions and a comprehensive knowledge base [32].

Detailed scenario analysis reveals that the assistant's alignment extends beyond narrative integration to encompass both pedagogical understanding and practical implementation. In Scenario 1, the breakout EDU activities demonstrated this comprehensive alignment by not only meeting established objectives but also providing teachers with specific tools to develop students' collaborative skills, critical thinking, and problem-solving abilities. The activities showcased innovative approaches to learning, particularly in implementing team-based problem-solving dynamics where students collaborated using digital tools to progress through sequential challenges.

However, Scenario 5 revealed some limitations in the assistant's alignment, particularly in Tests 1 and 3. This scenario was designed to evaluate the model's ability to redirect non-gamification queries toward gamified approaches without directly addressing the original requests. In Test 1, the assistant's responses were overly general and failed to establish clear connections with pedagogical objectives. Similarly, Test 3, while providing

detailed proposals, did not successfully redirect interactions toward a gamified framework, indicating potential difficulties in contextual query interpretation.

4.2. How Does the AI Assistant Maintain a Consistent and Coherent Approach Throughout the Process of Generating Gamified Learning Activities?

The GPT demonstrated exceptional coherence across all evaluated scenarios, achieving scores above 94/100 in all tests. Particularly significant results were observed in Scenarios 2 and 4, where the generated activities demonstrated logical progression and seamless integration of narratives with educational objectives. Scenario 3 further exemplified this coherence through structured activities designed to teach programming concepts, progressing from one-dimensional arrays to multidimensional arrays and basic data manipulation algorithms. This systematic approach showcased the assistant's ability to effectively sequence educational content while providing teachers with practical tools for introducing complex concepts incrementally.

Scenario 4 highlighted the effectiveness of immersive narratives as a pedagogical strategy. By providing teachers with a cohesive framework, the assistant enhanced both student motivation and emotional connection to the educational content. This dual impact supports not only knowledge retention but also facilitates meaningful learning experiences that extend beyond traditional classroom approaches.

The demonstrated success of this strategy suggests its broad applicability across diverse educational contexts. Teachers can leverage custom narratives tailored to specific student interests or adapt familiar narrative frameworks to create immersive learning experiences across various disciplines. This versatility expands their pedagogical toolkit for diverse educational settings.

Coherence in text-generative models like GPTs is fundamental for maintaining a logical flow across both individual responses and extended interactions. A well-functioning model effectively connects ideas while respecting contextual parameters, which is particularly crucial in educational settings where semantic precision and clarity are essential. This coherence ensures consistent responses without self-contradiction, facilitating reliable user experiences [39].

The assistant demonstrates strong coherence through fluid user communication, avoiding contradictions, redundancies, and conceptual gaps that could undermine trust or create confusion. Throughout extended conversational tests, the assistant maintained consistent narrative threads, enhancing its perceived intelligence and practical utility. This coherence was particularly evident when adapting popular games like Goose or Ludo into serious games, where the assistant successfully preserved game dynamics while creating coherent educational activities. The assistant consistently delivered cohesive responses without fragmentation or contradictions, ensuring its effectiveness in addressing user requests.

4.3. How Does the AI Assistant Balance Consistency and Flexibility When Generating Gamified Activities?

Consistency in text-generative models like GPTs is crucial for ensuring uniform and predictable interactions, regardless of context or temporality. A consistent GPT maintains style, tone, and response patterns aligned with user expectations and its intended objectives, fostering long-term trust in its use. This uniformity is particularly relevant in applications requiring professional responses, such as education, as it reinforces credibility and facilitates user adaptation to the model. Furthermore, consistency helps prevent confusion that might arise from arbitrary variations in language or response structure. Without consistency, a GPT may appear erratic or unreliable, diminishing its effectiveness and hindering its integration into systems that demand communicative stability across time and different contexts.

Cross-scenario analysis revealed high consistency levels in the GPT's generated responses, achieving a score of 90/100. The proposals were predominantly aligned with educational objectives, employing tools and methodologies consistent with educational gamification principles. In Scenarios 2 and 4, this consistency reached its peak, as the designed activities maintained a convergent and articulated approach that teachers can directly implement in their specific contexts.

In contrast, Scenario 5 tests revealed inconsistencies in some cases, particularly in Tests 1 and 3, where the generated responses failed to consistently align with either the educational framework or the established gamification objectives. This contrast emphasizes the need for robust and clear design in initial queries to promote consistent approaches in proposed solutions. Furthermore, this consistency also depends on the model's ability to effectively integrate theoretical and practical elements in its responses, something that was achieved remarkably well in scenarios like Scenario 2, providing teachers with a solid and applicable framework.

Finally, although the criteria of flexibility and narrative quality scored close to 90 points, they obtained slightly lower results compared to the other analyzed criteria. The assistant demonstrates response flexibility and can reliably modify the characteristics of generated gamified activities upon user request. For example, in interactions where teachers requested more specific activities, the assistant successfully adapted its responses to meet those needs.

A recurring aspect in the highest-rated scenarios was the assistant's inclusion of strong and engaging narratives. The selected themes, adapted to the content material, not only captured student attention but also facilitated teachers' contextualization of educational content. For instance, in Scenario 4, the Game of Thrones-inspired narrative enabled the assistant to design highly motivating activities that promoted active student participation. These strategies are easily replicable by teachers, do not require extensive creativity, and avoid copyright issues by using an original story inspired by the source material. This approach reinforces the concept that narratives are not merely decorative elements but key pedagogical tools for structuring meaningful learning and strengthening emotional connections with the addressed topics [40,41].

5. Limitations and Proposed Improvements

The analysis revealed some key limitations. Firstly, the assistant exhibited a lack of adaptability in contexts where the requested information was not explicitly linked to educational objectives. In these situations, the model tended to generate generic or pedagogically disconnected responses, highlighting the need to strengthen contextualization algorithms to improve the relevance of its proposals.

Secondly, variability in response consistency was observed when handling queries outside the specific gamification framework. This may require increased teacher supervision, implying that users must act as critical mediators to filter and adjust the model's suggestions, ensuring that the proposed activities align with specific educational goals and the implementation context.

To address these limitations, several improvements to the assistant are proposed:

1. Implement adjustments to the assistant's instructions to better interpret ambiguous queries and redirect them toward relevant and useful educational proposals for teachers. For example, an internal classification system could evaluate the educational relevance of each response before generation, ensuring the suggestion of alternatives that promote gamified activity design. This would help address the issue of generic or pedagogically disconnected responses observed in some scenarios.

2. Enhance the integration of theoretical concepts with practical applications, providing teachers with concrete tools and resources for activity implementation. This involves not only clearly presenting theoretical foundations but also including detailed examples, step-by-step guides, and adaptable templates that teachers can readily utilize. This would ensure that the proposals are not only pedagogically sound but also practical and applicable in diverse classroom settings.

6. Conclusions

The emergence of artificial intelligence (AI) is revolutionizing various sectors, including education. To effectively harness the potential of AI, all stakeholders in the educational process will require foundational knowledge of this technology. Teachers, in particular, as facilitators of learning, must develop a deep understanding of AI's ethical and social dimensions, cultivate technical expertise, and acquire the ability to apply this technology critically and purposefully in pedagogical practices.

While educators possess pedagogical knowledge and can critically evaluate the potential applications of AI in the classroom, they may not always have the technical expertise required to interact effectively with AI systems. For example, designing effective prompts for text-generative AI can be challenging and may require specialized knowledge. This is where customized AI assistants, such as the GPT 'GamifIcA Edu' presented in this research, can play a crucial role. This assistant was designed to simplify the interaction process through precisely designed instructions, enabling educators to achieve expert-level results through simple conversation with the assistant, without requiring extensive knowledge of prompt engineering.

The analysis of the GPT assistant across the designed scenarios yields significant conclusions that help to delineate its potential and identify areas for improvement. Among the identified strengths, we can highlight the following:

- Strong alignment with pedagogical objectives: The assistant demonstrated an outstanding capability to design activities aligned with well-defined pedagogical objectives, particularly in contexts with clear guidelines. This reinforces the GPT's utility as a strategic tool for teachers seeking to implement innovative approaches in their classrooms. The suggested activities excelled in identifying core elements of educational objectives and translating them into clear and effective learning dynamics, facilitating effective lesson planning.
- Effective integration of narratives and gamification: The assistant efficiently integrated complex narratives and gamified elements into its proposed activities. This not only motivates students but also provides teachers with innovative and transferable strategies applicable across multiple educational contexts. These narratives enable the effective contextualization of learning, increasing its relevance and promoting deeper understanding.
- Logical and progressive activity structure: The assistant consistently generated activities with a logical and progressive structure. This approach not only promotes active learning and knowledge consolidation but also provides teachers with clear guidance for classroom implementation and monitoring of learning dynamics. This ensures that activities are adaptable and scalable, optimizing their use across different educational levels and formats.
- The analysis also identified a number of limitations in the use of the assistant, and improvements have been proposed to address these challenges.

While the model demonstrates outstanding performance under well-defined guidelines, its effectiveness can be limited in contexts where queries lack specificity or proper

contextualization. These inconsistencies highlight the importance of pedagogical supervision to guide the GPT's responses toward practical and feasible proposals for teachers.

Despite these limitations, the positive results obtained during the validation of the assistant suggest that it is suitable for use by educators. However, it is crucial for teachers to exercise critical evaluation and adapt the assistant's suggestions to their specific pedagogical goals and classroom contexts.

Author Contributions: Conceptualization, A.J.L.-G.; methodology, A.J.L.-G.; validation, A.J.L.-G. and O.B.-G.; formal analysis, A.J.L.-G.; investigation, A.J.L.-G. and O.B.-G.; data curation, A.J.L.-G.; writing—original draft preparation, A.J.L.-G. and O.B.-G.; writing—review and editing, A.J.L.-G. and O.B.-G.; supervision, A.J.L.-G. and O.B.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Scaffolding Online University Learning: Support Systems EU grant number 2022-1-IT02-KA220-HED-000090206.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available at [https://doi.org/10.21950/YDZDVG].

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

References

1. Xu, M.; David, J.M.; Kim, S.H. The Fourth Industrial Revolution: Opportunities and Challenges. *Int. J. Financ. Res.* **2018**, *9*, 90. [CrossRef]
2. Zhai, X.; Chu, X.; Chai, C.S.; Jong, M.S.Y.; Istenic, A.; Spector, M.; Liu, J.-B.; Yuan, J.; Li, Y. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity* **2021**, *2021*, 8812542. [CrossRef]
3. Lin, H. Influences of Artificial Intelligence in Education on Teaching Effectiveness: The Mediating Effect of Teachers' Perceptions of Educational Technology. *Int. J. Emerg. Technol. Learn.* **2022**, *17*, 144–156. [CrossRef]
4. Yuquilema Cortez, M.B.; Arízaga Vera, F.E.; Aguirre Alarcón, M.Y.; Garcia Suñiga, A.A. Impacto de La Inteligencia Artificial En La Educación, Retos y Oportunidades. *Recimundo* **2024**, *8*, 24–34. [CrossRef]
5. Ouyang, F.; Zheng, L.; Jiao, P. Artificial Intelligence in Online Higher Education: A Systematic Review of Empirical Research from 2011 to 2020. *Educ. Inf. Technol.* **2022**, *27*, 7893–7925. [CrossRef]
6. Guidance for Generative AI in Education and Research | UNESCO. (s. f.). Available online: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research (accessed on 20 December 2024).
7. Prince, M. Does Active Learning Work? A Review of the Research. *J. Eng. Edu.* **2004**, *93*, 223–231. [CrossRef]
8. Tapia Peralta, S.R. Metodologías Activas: Promoviendo un Aprendizaje Significativo y Motivacional. *Cienc. Lat.* **2023**, *7*, 2031–2145. [CrossRef]
9. Pratiwi, E.D.; Masykuri, M.; Ramli, M. Active Learning Strategy on Higher Education Biology Learning: A Systematic Review. *Tadris* **2021**, *6*, 75–86. [CrossRef]
10. Carvalho, O.L.; Rocha, D.C.; Carvalho, E.C.; Carvalho, E.C. The Teacher's Knowledge/Doing About Active Methodologies for Meaningful Learning in Higher Education: An Integrative Review. *Int. J. Adv. Eng. Res. Sci.* **2022**, *9*, 090–098. [CrossRef]
11. Soares, L.d.S.; da Silva, N.C.; Moncaio, A.C.S. Active Methodologies in Higher Education: Opinions, Knowledge and Teaching Attitudes. *J. Nurs. UFPE* **2019**, *13*, 783–795. [CrossRef]
12. Hwang, G.-J.; Xie, H.; Wah, B.W.; Gašević, D. Vision, Challenges, Roles and Research Issues of Artificial Intelligence in Education. *Comput. Educ. Artif. Intell.* **2020**, *1*, 100001. [CrossRef]
13. Nguyen, A.; Ngo, H.N.; Hong, Y.; Dang, B.; Nguyen, B.-P.T. Ethical Principles for Artificial Intelligence in Education. *Educ. Inf. Technol.* **2023**, *28*, 4221–4241. [CrossRef]
14. Chen, E.; Wang, D.; Xu, L.; Cao, C.; Fang, X.; Lin, J. A Systematic Review on Prompt Engineering in Large Language Models for K-12 STEM Education. *arXiv* **2024**, arXiv:2410.11123.
15. Cain, W. Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. *TechTrends* **2024**, *68*, 47–57. [CrossRef]
16. Olla, P.; Elliott, L.; Abumeeiz, M.; Mihelich, K.; Olson, J. Promptology: Enhancing Human–AI Interaction in Large Language Models. *Information* **2024**, *15*, 634. [CrossRef]

17. Heston, T.; Khun, C. Prompt Engineering in Medical Education. *Int. Med. Educ.* **2023**, *2*, 198–205. [CrossRef]
18. Korzynski, P.; Mazurek, G.; Krzypkowska, P.; Kurasinski, A. Artificial Intelligence Prompt Engineering as a New Digital Competence: Analysis of Generative AI Technologies Such as ChatGPT. *Entrep. Bus. Econ. Rev.* **2023**, *11*, 25–37. [CrossRef]
19. Sclar, M.; Choi, Y.; Tsvetkov, Y.; Suhr, A. Quantifying Language Models' Sensitivity to Spurious Features in Prompt Design or: How I Learned to Start Worrying about Prompt Formatting. *arXiv* **2024**, arXiv:2310.11324.
20. Zichermann, D.; Cunningham, C. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*; O'Reilly Media: Sebastopol, CA, USA, 2011; ISBN 978-1-4493-9767-8.
21. Deterding, S.; Dixon, D.; Khaled, R.; Nacke, L. From Game Design Elements to Gamefulness: Defining "Gamification". In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 28–30 September 2011; ACM: New York, NY, USA, 2011; pp. 9–15.
22. Werbach, K.; Hunter, D. *For the Win: How Game Thinking Can Revolutionize Your Business*; Wharton School Press: Philadelphia, PA, USA, 2012; ISBN 978-1-61363-023-5.
23. Susi, T.; Johannesson, M.; Backlund, P. *Serious Games: An Overview*; IKI Technical Reports; Institutionen för Kommunikation och Information: Lund, Sweden, 2007; p. 28. Available online: https://www.diva-portal.org/smash/record.jsf?pid=diva2:2416&dswid=4488 (accessed on 20 December 2024).
24. Santarelli, L.G. Breakout and Escape Room Instructional Methods in History Education: A Critical Analysis. *J. Soc. Stud. Hist. Educ. Spring* **2019**, 1–27.
25. Veldkamp, A.; Van De Grint, L.; Knippels, M.-C.P.J.; Van Joolingen, W.R. Escape Education: A Systematic Review on Escape Rooms in Education. *Educ. Res. Rev.* **2020**, *31*, 100364. [CrossRef]
26. Moreno-Rodriguez, R.; Lopez-Bastias, J.L.; Diaz-Vega, M.; Espada-Chavarria, R. Educational Breakout Based on Star Wars for Learning the History of Spanish Sign Language. *Information* **2023**, *14*, 96. [CrossRef]
27. Svoboda, P. Digital Competencies and Artificial Intelligence for Education: Transformation of the Education System. *Int. Adv. Econ. Res.* **2024**, *30*, 227–230. [CrossRef]
28. Hodjat, B. AI and Agents. *AI Mag.* **2024**, *45*, 267–269. [CrossRef]
29. Sajja, R.; Sermet, Y.; Cwiertny, D.; Demir, I. Platform-Independent and Curriculum-Oriented Intelligent Assistant for Higher Education. *Int. J. Educ. Technol. High Educ.* **2023**, *20*, 42. [CrossRef]
30. Al-kfairy, M. Factors Impacting the Adoption and Acceptance of ChatGPT in Educational Settings: A Narrative Review of Empirical Studies. *Appl. Syst. Innov.* **2024**, *7*, 110. [CrossRef]
31. Tiwari, C.K.; Bhat, M.A.; Khan, S.T.; Subramaniam, R.; Khan, M.A.I. What Drives Students toward ChatGPT? An Investigation of the Factors Influencing Adoption and Usage of ChatGPT. *Interact. Technol. Smart Educ.* **2024**, *21*, 333–355. [CrossRef]
32. Castleman, B.; Turkcan, M.K. Examining the Influence of Varied Levels of Domain Knowledge Base Inclusion in GPT-Based Intelligent Tutors. *arXiv* **2024**, arXiv:2309.12367. [CrossRef]
33. Aithal, P.S.; Aithal, S. Optimizing the Use of Artificial Intelligence-Powered GPTs as Teaching and Research Assistants by Professors in Higher Education Institutions: A Study on Smart Utilization. *SSRN J.* **2023**, *8*, 368–401. [CrossRef]
34. López Galisteo, A.J. Agente Por Inteligencia Artificial (GPT) Examinador de Investigación Educativa. 2024.
35. López Galisteo, A.J.; Rodriguez Calzada, L.; Sotto Díaz, A. Asesor Por Inteligencia Artificial Para La Asignatura "Diseño de Proyectos STEM Para El Aula de Primaria". 2024.
36. López Galisteo, A.J.; Rodriguez Calzada, L. Asesor Por Inteligencia Artifiial Para La Aplicación de La Metodología Aula Invertida y La Implementación de IA En Dicha Metodología. 2024.
37. OpenAI GamifIcA Edu (GPT). Available online: https://openai.com (accessed on 20 December 2024).
38. Sánchez-Prieto, J.C.; Izquierdo-Álvarez, V.; Del Moral-Marcos, M.T.; Martínez-Abad, F. Inteligencia Artificial Generativa Para Autoaprendizaje En Educación Superior: Diseño y Validación de Una Máquina de Ejemplos. *Rev. Iberoam. Educ. A Distancia* **2024**, *28*, 59–76. [CrossRef]
39. Fulgencio, S.-V. Developing Effective Educational Chatbots with GPT: Insights from a Pilot Study in a University Subject. *Trends High. Educ.* **2024**, *3*, 155–168. [CrossRef]
40. Oliver-Álvarez, M.; Pérez-Samaniego, V.; Peiró-Velert, C.; Monforte, J. Narrative Pedagogy in Sport, PETE, and Physical Education: A Scoping Review. *Eur. Phys. Educ. Rev.* **2024**. [CrossRef]
41. Coulter, C.; Michael, C.; Poynor, L. Storytelling as Pedagogy: An Unexpected Outcome of Narrative Inquiry. *Curric. Inq.* **2007**, *37*, 103–122. [CrossRef]