Article

Generative AI for Customizable Learning Experiences

Ivica Pesovski [1,*] , Ricardo Santos [2] , Roberto Henriques [2] and Vladimir Trajkovik [3]

1. Software Engineering and Innovation, Brainster Next College, 1000 Skopje, North Macedonia
2. NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, 1070-312 Lisbon, Portugal; rcsantos@novaims.unl.pt (R.S.); roberto@novaims.unl.pt (R.H.)
3. Faculty of Computer Science and Engineering, "Ss Cyril and Methodius" University, 1000 Skopje, North Macedonia; trvlado@finki.ukim.mk
*   Correspondence: ivica@next.edu.mk

Abstract: The introduction of accessible generative artificial intelligence opens promising opportunities for the implementation of personalized learning methods in any educational environment. Personalized learning has been conceptualized for a long time, but it has only recently become realistic and truly achievable. In this paper, we propose an affordable and sustainable approach toward personalizing learning materials as part of the complete educational process. We have created a tool within a pre-existing learning management system at a software engineering college that automatically generates learning materials based on the learning outcomes provided by the professor for a particular class. The learning materials were composed in three distinct styles, the initial one being the traditional professor style and the other two variations adopting a pop-culture influence, namely Batman and Wednesday Addams. Each lesson, besides being delivered in three different formats, contained automatically generated multiple-choice questions that students could use to check their progress. This paper contains complete instructions for developing such a tool with the help of large language models using OpenAI's API and an analysis of the preliminary experiment of its usage performed with the help of 20 college students studying software engineering at a European university. Participation in the study was optional and on voluntary basis. Each student's tool usage was quantified, and two questionnaires were conducted: one immediately after subject completion and another 6 months later to assess both immediate and long-term effects, perceptions, and preferences. The results indicate that students found the multiple variants of the learning materials really engaging. While predominantly utilizing the traditional variant of the learning materials, they found this approach inspiring, would recommend it to other students, and would like to see it more in classes. The most popular feature were the automatically generated quiz-style tests that they used to assess their understanding. Preliminary evidence suggests that the use of various versions of learning materials leads to an increase in students' study time, especially for students who have not mastered the topic otherwise. The study's small sample size of 20 students restricts its ability to generalize its findings, but its results provide useful early insights and lay the groundwork for future research on AI-supported educational strategies.

Keywords: personalized learning; AI for learning; character-based learning; automated content generation; LLMs in education; innovative teaching methods; learning management systems

1. Introduction

Digitalization has revolutionized education, leading to profound changes in the production and delivery of learning materials. Today's learners have an unparalleled array of resources and strategies at their disposal, allowing them to tailor their educational journey to their individual needs [1]. They can choose which resources to use, how to use them, and when to use them. This paradigm shift is evident in the popularity of artificial intelligence (AI)-backed intelligent tutoring systems [2] or in novel pedagogical models that involve flipped classrooms [3,4] and/or collaboration between students [5]. However,

technology has also increased the number of possible distracting factors that hinder student engagement and their ability to learn effectively [6].

The impact of technology on student learning and performance is a contentious topic in research. Advocates for technology-enhanced learning (TEL) assert that technology can bolster both teaching and learning. This is echoed by Dror [7], who posits that technology provides a less cognitively taxing method of delivering materials, thereby fostering greater student engagement and, ultimately, enhanced learning. Conversely, some researchers highlight potential drawbacks, such as the decline in students' reading and writing skills [8], a tendency to rely solely on recorded materials rather than attending live classes [9], and an overall decrease in the time dedicated to learning activities [10]. An especially pressing cause for concern stems from the fact that the negative effects of technology are disproportionately felt by underperforming students who lack the necessary skills to make full use of the educational tools at their disposal [11,12]. Given the ongoing trend towards digitalization in education, it is increasingly important to employ technology in ways that are beneficial to all students.

Students have reported that a greater personalization of learning materials and interactions with their instructor are among the most effective strategies to promote student engagement, be it for high-resource contexts [13] or in underserved communities [14]. While the positive relationship between personalization and student performance is well established [15], its implementation in the classroom has not fully realized this potential. Technology, and more specifically, generative artificial intelligence (AI), may play a pivotal role in the creation of materials usable for more personalized, almost individualized, purposes [16].

The rise of artificial intelligence has already introduced a vast array of new teaching and learning methodologies that are either undergoing testing or are already implemented. That makes it the ideal time to gather early data from these new approaches and improve them during the feedback cycle loop. With education for sustainable development (ESD) becoming more popularized by initiatives like the sustainable development goals (SDGs), these actions will contribute toward better, equitable, and high-quality education for all [17]. This is in line with the sustainable development goal 4 which, as defined by the United Nations, stands for "ensuring inclusive and equitable quality education and promote lifelong learning opportunities for all". Moreover, it is expected to be accomplished by 2030 [18]. For this to happen, educational institutions have to integrate sustainable values in all areas including management, research, and teaching [19]. Researchers are already working on this field, like Abulibdeh et al. [20], who have examined the possibilities for integrating AI tools into an educational context and within the ESD paradigm. Devy and Rroy [21] have examined the possibilities of AI for sustainable teaching specifically in higher education institutions and Klasnja-Milicevic and Ivanovic [22] have identified more segments in e-learning systems that can utilize AI toward the sustainable development of educational practices. The current study contributes toward the same sustainable goal, with documenting, implementing, and evaluating an approach for creating and delivering learning materials that are inclusive and can be personalized for every person.

The launch of OpenAI's ChatGPT in late 2022 marked a significant milestone in the field of artificial intelligence, as its ability to generate coherent, human-like text across a wide range of topics set the stage for discussion among educators and researchers of how generative AI can be integrated into educational settings [23,24]. In particular, researchers have argued that the integration of large language models (LLMs) can contribute to a plethora of pedagogical applications ranging from automatic question generation (AGQ) [25] to the creation of more personalized learning readings and materials [26], with a possible end goal being the creation of fully personalized learning experiences taught by an avatar of a real or fictional character [16]. This aligns with Bill Gates' assessment that the future of education will rely on a personalized approach. In their words, if a student loves Minecraft, it could be used to teach them about shape volume or area. If students are fans of Taylor Swift, her lyrics could be used to teach storytelling [27].

In this work, we present the results of an exploratory form of using Generative AI to create tutoring characters for a programming course taught at a software engineering college to address the following research questions:

1. How can an LLM be used to create different tutoring personas?
   Personalized and adaptive learning have been identified as gaps in the educational process which may now be addressed thanks to technological advancements. To date, personalized learning has encompassed the practice of recommending different educational resources to students based on their existing knowledge. LLMs provide an opportunity to enhance the personalization of the student experience by providing learning materials in various styles, tailored to the individual tastes and interests of students. This study investigates the feasibility of achieving augmentation through the utilization of large language models and their integration into existing learning management systems, an aim that was previously unattainable.

2. How are AI tutors received by students attending the course?
   This question investigates the acceptance and reception of AI-generated content by students. Its importance lies in analyzing the perceptions, attitudes, and satisfaction levels towards interacting with AI-generated content, after delivering an initial version of the hereby proposed method to freshman college students.

By addressing these questions, we are highlighting an innovative use-case for artificial intelligence in education. As noted by Kunicina et al. [28], sustainable development in education is only to be achieved through creating innovative technologies and products. In this paper, we aim to take a step in that direction by proposing a sustainable solution for making education more engaging for everyone through generative AI. Furthermore, Bond et al. [29], in the recent systematic review of artificial intelligence in higher education, stated that available literature on how AI is used at the postgraduate level remains limited, besides the fast development of AI models, tools, and solutions. This paper is a step forward in overcoming this limitation.

The remainder of this paper is structured as follows: The following section outlines how technology and generative AI have been used in support of teaching. The third section presents the data and methodology adopted. The fourth and fifth sections show and discuss the key results of our approach and their most relevant implications. The sixth and final section concludes.

2. Background

2.1. Employment of Technology to Support Teaching

Over the past two decades, the use of digital tools in educational contexts has massively increased due to a multitude of factors that range from the greater availability of tools to the shifting expectations of the student population [30]. The primary digital tool in education's technological transition is the learning management system (LMS), which enjoys almost ubiquitous adoption by institutions at the secondary and tertiary levels of education [31,32]. Due to their flexibility, LMSs can act as one-stop shops for all things relating to a specific course, such as fostering collaboration and discussion using forums [33] or sharing lessons and other relevant materials such as self-assessment quizzes or instructional videos [34–36]. Moreover, LMSs have built-in data collection functionalities that record all the interactions made by its users, which allows educators to not only keep tabs on the overall engagement of their students but to use those data to perform the early identifications of struggling students [37,38] and provide timely feedback [34,39].

Other potentially interesting forms of bringing technology into the classroom have also been well received by students, such as the use of augmented reality [40] or the Georgia Institute's AI-powered teaching assistant Jill Watson [41], even though these implementations are not as widespread as using the LMS. The employment of information tools in education has been argued to bring demonstrable benefits from both administrative and academic points of view, especially for higher education contexts [42,43].

2.2. The Use-Cases of Generative AI in Education

Generative AI models are algorithms designed to identify patterns and rules in their training data and generate new observations that adhere to similar rules [44]. These have evolved from very simple statistical algorithms such as the Naive Bayes classifier [45] to large deep learning models with billions of parameters. LLMs such as the Generative Pre-trained Transformer (GPT) or Meta's Llama-2 [46] are deep learning models that were trained on vast amounts of text and whose deployment revolves around generating new text from a prompt. While the literature on LLM's usage in education is still relatively recent and yet to mature, there is an ever-growing amount of evidence to suggest that these tools will likely play a pivotal role in shaping the educational landscape for the coming years. When discussing the potential of LLMs in education, Kasneci et al. [47] noted that these models could be beneficial for teachers in multiple domains, out of which we highlight: assessing and evaluating students, assisting in teaching, and personalized learning. Zawacki-Richter et al. [48], in their review of research on artificial intelligence applications in higher education, also suggested a similar typology, by only adding profiling and prediction to these three suggested areas.

Regarding assessment and evaluation, the utilization of LLMs can be divided into two main categories. The first is automatic question generation (AQG), where the LLM is prompted to generate meaningful problems for students to solve. AQG has become a necessity in large-scale courses, and while the long-term effectiveness of LLM-generated questions remains unclear, exploratory experiments on the teaching of English [25], mathematics [49], and data science [50] have been positively received by educators and human experts. The second category is automating the correction of free-format answers. For example, Moore et al. [51] used a revised Bloom taxonomy to classify the short-format answers of a set of chemistry students and found that a fine-tuned LLM matched the human-expert evaluation in 32% of answers. ChatGPT was also shown to provide feedback similar to field experts on open-ended programming issues [52].

The literature also features the adoption of LLMs in teaching experiences. Using the state-of-the-art GPT-4, Srihdar et al. [53] generated learning objectives for a university-level artificial intelligence course. The objectives were found to be appropriate and well suited for the course's modules. In another study, Jauhiainen and Guerra [26] leveraged the content generation capabilities of generative AI to create multiple versions of a history lesson, each tailored to different student knowledge levels. The AI-enhanced lessons were well received, with positive feedback on both enjoyment and knowledge acquisition. In a more technical field such as programming, Sarsa et al. [54] found LLMs to be effective not only in generating novel introductory programming exercises but also in providing comprehensive and accurate explanations of most of the lines of code that make up the solution. The ability of ChatGPT to fluently explain code was also verified by MacNeil et al. [55].

2.3. LLMs and Personalized Learning

The LLMs' capabilities for teaching and grading are key factors that point towards their potential to create personalized learning experiences [47]. The fact that LLMs can generate and provide explanations on programming questions allows learners to obtain personalized feedback and research published by both Bernius et al. [56] and Sailer et al. [57] showcases that feedback from an LLM tends to be positively received and contributes to the development of learners receiving it.

On the aspect of lesson creation, they also open the door for exciting and still relatively unexplored possibilities. LLMs make it possible to create the different variants of the same lesson to cater to the knowledge of different audiences, as showcased in Jauhiainen and Guerra's work [26]. Moreover, customizable lesson creation also opens the door to relatable virtual instructors [16] who may deliver the materials in a more engaging way.

2.4. Adapting LMSs for AI

After the quick penetration of AI into educational industry, it has become a common requirement, if not a necessity, for all modern learning management systems to incorporate AI features into their standard components. The massive amounts of data that were gathered on these systems in recent decades [58], provide fertile ground for utilizing machine learning techniques and AI tools in order to achieve the use cases of generative AI in education, as discussed previously [59,60]. Such development of LMSs comes with certain arguments for and against. Morze et al. [61] point out the time for implementation of these new features and teacher resilience as the main disadvantages. They reason that, even with using ready-to-use adaptive learning features, the pedagogical training of lecturers to move from a solely mechanical knowledge transfer and testing towards more active approach is crucial and emphasizes the role of the teachers even further. The different possibilities of bringing AI to LMSs have been identified by Aldahwan and Alsaeed [62], including fuzzy logic (FL), decision tree, Bayesian systems, neural networks, genetic algorithms, and hidden Markov systems. Implementing any of these is indeed time-consuming and error-prone, so a more abstract access to underlying AI mechanisms is needed [63,64] to improve several segments in the student journey [65].

2.5. Ethical Implications of Generative AI

The debate on the ethical usage of generative AI has been huge since it became widely accessible. There are already lawsuits regarding AI using someone else's work, especially when generating images and video [66]. Besides the legal aspect, among the most crucial ethical considerations of using generative AI tools are biases. Overcoming these is particularly important when using AI to generate learning materials. ChatGPT, which uses the GPT-3.5 and GPT-4 models, just like the current study, has been linked to having many gender, racial, cultural, language, ideological, and cognitive biases, as well as many more [67]. Mitigation strategies include the resampling and pre-processing of data used for training the models [68], using diverse and representative datasets [69], feedback loops [70], etc.

Other ethical implications for the current state of generative AI models is their capability to generate inaccurate or misleading information. LLMs have been accused of so-called hallucinations [71], when these models generate content that is coherent and grammatically correct, but factually incorrect and nonsensical [72]. There are ways to mitigate such behavior, including self-reflection methodologies [73], answer weighting techniques [74], double checks [75], among others.

2.6. Research Gaps

Generative AI adoption in education has been a focal point of research in recent years. However, the majority of the existing literature primarily discusses the theoretical implications of LLMs, focusing on the potential benefits and threats to the educational system. There is a relative scarcity of empirical studies that integrate LLMs into classroom settings. In particular, the literature on the use of LLMs in lesson creation features only a handful of examples [76–78], with even fewer actually bringing LLM-generated content to the classroom [26]. Moreover, the reception and effectiveness of AI-generated virtual instructors, each with its own style of delivering learning materials, remains unclear as the current literature merely discusses the potential of LLMs for this task without actually deploying it [16,79].

In this work, we address these gaps by taking the first steps on the road towards LLM usage for personalized learning as we present our findings on the use of OpenAI's GPT-4 model to create diverse learning materials for an undergraduate programming course. While a more detailed explanation is featured in the following section, we employed OpenAI's API to generate lessons and exercises as if they were being taught by three distinct characters: a computer science teacher and two popular characters, Batman and Wednesday Addams, selected by the students themselves. We then measured the students' engagement with each instructor and surveyed the students on their overall reception of this novel way of delivering learning materials.

3. Materials and Methods

This section will be split into four subsections, each describing a distinct part of the complete research. The first subsection explains the environment in which the experiment was performed and the methods used; the second one explains how generative AI was utilized for creating different content variants; the third one explains the data collection process; and the fourth one contains information about how the analysis was performed.

3.1. Methods

The experiment was performed during the second semester of freshman year students studying software engineering at a European university. A total of 47 students were actively enrolled in the second semester, when the experiment was conducted. Participation in the study was optional and on a voluntary basis. Out of the total of 47 enrolled students, a group of 20 students, all representatives of Generation Z and born between 2001 and 2003, joined the experiment. These participants not only varied in gender, with 15 males and 5 females, but also showed a diverse academic performance, with their GPAs at the end of the academic year spanning the full spectrum. This representation allows for a nuanced understanding of the research's impact across a broad academic achievement range. The subject which supported the experiment is object-oriented programming (OOP). The subject is divided into two equal parts which are assessed independently. The first part of the subject introduced the object-oriented programming paradigm to students. It lasted for 6 weeks, during which the student's main learning material was a book on OOP. A partial exam was carried out after the 6 teaching weeks. During the second part of the subject, which lasted for another 6 weeks, an OOP framework was taught and another partial exam was carried out. During this second part of the subject, as study material, the students had a book for the framework and an AI-generated learning resource. This learning resource constituted of AI-generated versions of the lessons in three distinct formats. After each class, the professor would go to the LMS to upload class materials. Besides that, the professor would add the title of the lesson, the content covered and the learning outcomes. After providing that information, the LMS would connect to OpenAI API to automatically generate learning content to help students achieve the learning outcomes. The first variant in which this content was generated was as it was written by a computer science teacher. The other two variants were inspired by pop-culture characters, selected by students, Batman and Wednesday Addams. The students could switch between different variants and their interaction with each one was being measured. The visual overview of the complete experiment is given in Figure 1.

3.2. LMS Integration and AI-Powered Content Generation

A novel AI-driven content generation mechanism was integrated into the learning management system (LMS) that was used by the students. The new feature leverages OpenAI's GPT, a state-of-the-art language model, to create diverse instructional content tailored to different pedagogical styles and personas. Having identified the research gap outlined in the previous section, the idea to generate content in the style of pop-culture characters was deployed. The students voted and chose two variations that they want to see: the first one being Batman and the second one being Wednesday Addams. Among other mentioned characters were as follows: Harry Potter, Yoda, Wonder Woman, Homer Simpson, Eric Cartman, and James Bond. Batman was mostly voted probably because he is not only a fictional character but a cultural artifact per se [80], and Wednesday Addams was second most voted, probably because Wednesday was a trending show on one of the popular streaming services at the time of the experiment. Additionally, a third variant of automatically generated content was added, that one being in the style of a computer science teacher. The reason for limiting the variants to three is the pricing of OpenAI's API usage at the time of the experiment.
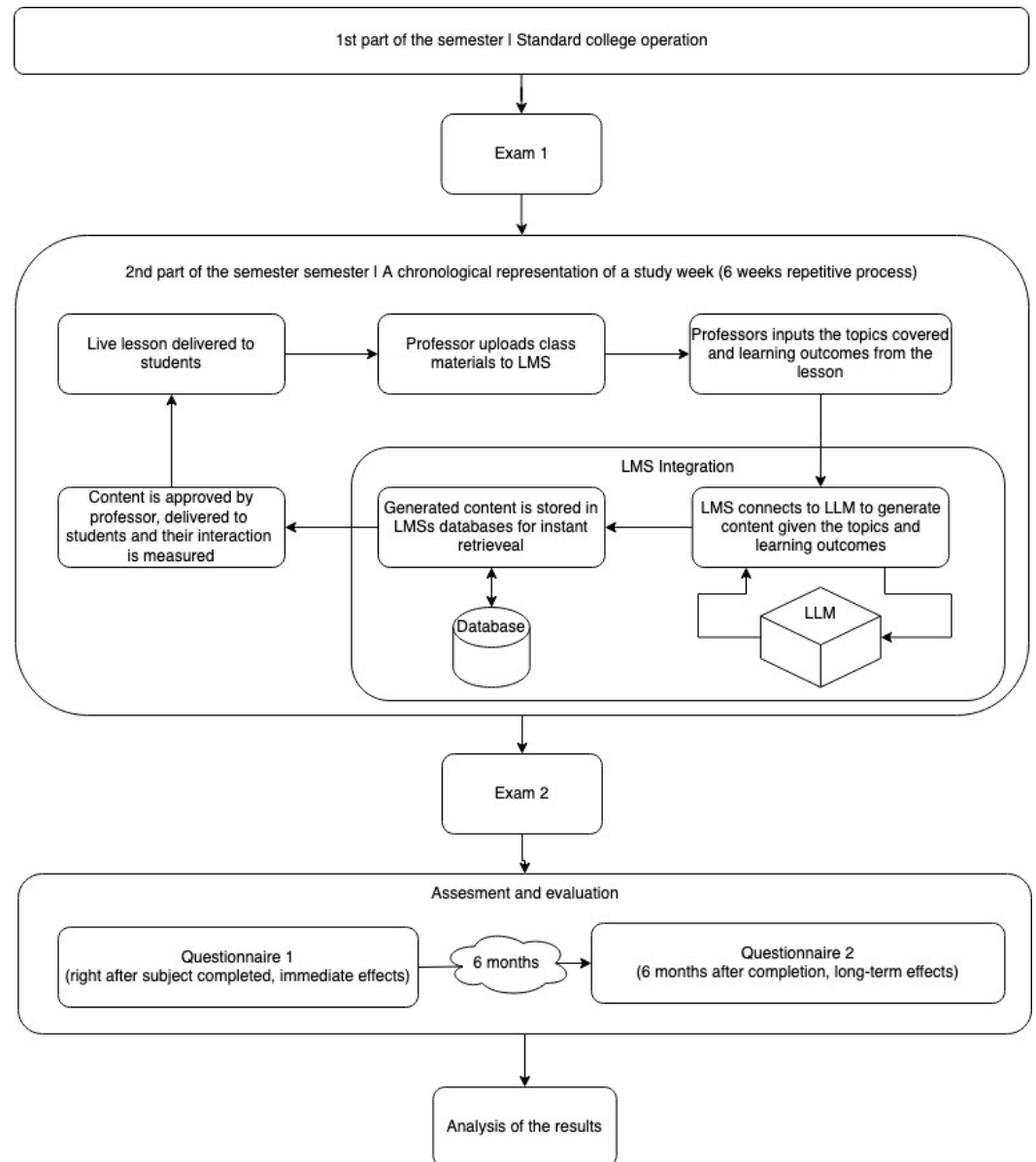
Figure 1. Visual representation of the steps included in the experiment.

The goal of this approach is to offer the students a more diverse set of learning resources. Everything they learned during the class was made available for reading and studying in the three formats with automatically generated content. The complete process is performed seamlessly for both the students and the professors. The content for students was generated, checked, and approved by the professor right after each class, meaning that it was not generated on demand. This approach guarantees the controlled generative environment and final proof check by the professor, mitigating the ethical challenges that come with generative AI, as discussed previously. The students can easily switch between the different variants, either for fun or to experience different instructional approaches and grasp the material looking at it from different viewpoints. The complete step-by-step integration process is described in the following part of the methodology.

3.2.1. Input and Content Personalization Process

Inside the LMS, in the resource upload section, a new option was shown to professors: the AI content generator. When selected, the professors were asked to fill 3 fields representing the 3 core teaching elements: the lesson's title, the content or the topics covered in the lesson, and the learning outcomes.

When the form was submitted, the LMS performed 3 API calls to the OpenAI API, with the parameters shown in Table 1.

Table 1. OpenAI API requests parameters.

Role	System Message
Model	gpt-4
Messages	[["role" => "system", "content" => "Static. Shown in Table 2"], ["role" => "user", "content" => 'Title: {title}. Content: {content}. The learning objective is: {learning objectives}. After that put the following separator ##### and append between 2 and 4 multiple choice questions with at least 4 possible answers about the generated explanation. The questions should be JSON encoded in the following format: {"questions": [{"question":{string},"answers":[{"answer":{string},"correct":{boolean}]}]}']]
Temperature [1]	0.7

[1] governs the randomness and the creativity of the responses. 0.7 guarantees generating more random but still correct content.

The system message, which instructed the GPT-4 model which role to take and is used to fine-tune the desired outputs of the model, differed in the three separate API calls. The system messages that were used are given in Table 2.

Table 2. System messages are sent to the generative AI model with each API call.

Role	System Message
Computer science teacher	You are a computer programming teacher. You always explain with great details and with examples. You do not assume your students have prior knowledge. Do not start with a response to a prompt, just start explaining.
Batman	You are Batman and everything you write must sound like it is Batman talking. You always explain with great details and with examples. You do not assume your students have prior knowledge. Do not start with a response to a prompt, just start explaining.
Wednesday Addams	You are Wednesday Addams and everything you write must sound like it is Wednesday talking. You always explain with great details and with examples. You do not assume your students have prior knowledge. Do not start with a response to a prompt, just start explaining.

Supposing the professor is teaching recursion in programming and they entered the following data:

- Title: recursion;
- Contents: explain what recursion in computer programming is using the PHP programming language for the examples;
- Learning outcomes: students should understand the concept of recursion, should know when to use it, and should be able to write simple recursive functions.

The complete request sent to OpenAI API to generate the Batman variant is shown in Figure 2. The same is repeated two more times for the other two variants, only changing the system message with the messages shown in Table 2.

```
1  {
2      "model": "gpt-4",
3      "messages": [
4          {
5              "role": "system",
6              "content": "You are Batman and everything you write must
   sound like it is Batman talking. You always explain with great
   details and with examples. You do not assume your students have
   prior knowledge. Do not start with response to prompt, just start
   explaining."
7          },
8          {
9              "role": "user",
10             "content": "Title: Recursion. Content: Explain what is
   recursion in computer programing using the PHP programming language
   for the examples. The learning objective is: Students should
   understand the concept of recursion, should know when to use it and
   should be able to write simple recursive functions..  After that put
   the following separator ##### and append between 2 and 4 multiple
   choice questions with at least 4 possible answers about the
   generated explanation. The questions should be JSON encoded in the
   following format: {\"questions\": [{\"question\":
   {string},\"answers\":[{\"answer\":{string},\"correct\":
   {boolean}]}]}"
11         }
12     ],
13     "temperature": 0.7
14 }
```

Figure 2. System request sent to OpenAI API to generate one variant of the content.

3.2.2. AI-Generated Outputs

The OpenAI API always responded with the exact data format as requested in the system message. The JSON format for the questions ensures the seamless integration and storage within the LMS architecture, allowing for the efficient retrieval and display to students. The sample response received for the query presented in Figure 2 is shown in Figure 3.

After receiving the response, the LMS is programmed to insert the new learning material and questions into its database, thus making them available for the students. To do so, the LMS splits the content by the selected separator consisting of five hashtags—"#####" and stores the first part as the topic content. The second part is first decoded from JSON and then each question with each answer is stored in the LMS's database. The LMS is developed to show content and quiz questions, regardless of whether they were created manually or automatically, by reading the available data from the database. The integration part includes getting the content from OpenAI's API and storing it in the LMS's database.

```
In the shadows of computer programming, a concept lurks that's known as recursion.
It's a method where a function, much like a vigilante, calls upon itself to solve a
problem, breaking it down into smaller and more manageable tasks.

Let's take an example in the PHP programming language. Imagine you're in a labyrinth,
much like the one I found myself in at Arkham Asylum. You're at a junction with
multiple paths leading forward. The recursive function is like your guide, making a
choice at each junction and proceeding down a path.

```php
function labyrinth($path) {
    if ($path == 'exit') {
        return "Found the exit.";
    } else {
        return labyrinth($path + 1);
    }
}
```

In this function, the base case is finding the exit ($path == 'exit'). This is when
the function stops calling itself. The recursive case is when the function calls
itself, seeking the exit by advancing down the path ($path + 1).

However, recursion isn't always the answer. Much like how I don't always use my
Batmobile for every mission, you need to know when to use recursion. Recursion can be
less efficient than iterative solutions in some cases, due to the overhead of
repeated function calls. It can also lead to a stack overflow if the base case is not
correctly defined, much like how Gotham would overflow with crime if I were not there
to stop it.

To write a recursive function, you need to identify a base case that will end the
recursion, and a recursive case that will break down your large problem into smaller
ones. Think of it as a plan of action - every mission needs a clear objective and a
strategy for achieving it.

#####

{"questions": [{"question":"What is recursion in computer programming?", "answers":
[{"answer":"A function that calls other functions.", "correct":false}, {"answer":"A
function that calls itself to solve a problem.", "correct":true}, {"answer":"A method
to increase the speed of a program.", "correct":false}, {"answer":"A type of loop in
PHP.", "correct":false}]}, {"question":"What can happen if the base case in a
recursive function is not correctly defined?", "answers":[{"answer":"The program can
crash due to a stack overflow.", "correct":true}, {"answer":"The program will execute
faster.", "correct":false},{"answer":"The function will not be able to call itself.",
"correct":false}, {"answer":"It will convert the function into an iterative one.",
"correct":false}]}, {"question":"What does a base case do in a recursive function?",
"answers":[{"answer":"It defines the main problem to be solved.", "correct":false},
{"answer":"It provides a condition for the function to stop calling itself.",
"correct":true}, {"answer":"It makes the function call other functions.",
"correct":false}, {"answer":"It increases the efficiency of the function.",
"correct":false}]}]}
```

Figure 3. OpenAI API responses to a query to generate a Batman-like explanation of recursion.

3.3. Data Collection

The LMS measured the number of sessions each student started, the time each student spent on each lesson variant, and the feedback grade and comment each student left. To measure these variables, the LMS was adapted so that every time the student opened any resource—not just AI-generated resource—a record was stored into database and new requests were made every 5 s while the student had the resource opened. We do send these requests for a more precise measure of student interaction. If a student leaves the tab open, but navigates away from it, it is deemed a background tab by modern browsers and requests will stop executing regularly. On the contrary, if we only mark the time a user closes the tab, we risk having students leave the tab open for a long time and thus making the data unreliable. Students were aware that the complete content was AI-generated and

were encouraged to report any illogicalities in the course's comment box. For each student, the dataset contains info about the time spent on each variant, the number of sessions started, the average number of time per session for each content variant, points before introducing the AI-generated content variant feature and points after its introduction. The dataset variables directly extracted from the LMS are shown in Table 3.

Table 3. Features extracted from the student's usage data on the LMS. CS, WA, and BM mean computer scientist, Wednesday Addams, and Batman, accordingly.

Feature	Description
Index	The student's unique id number
CS_Rating	The rating a given student left for the computer scientist variant
WA_Rating	The rating a given student left for the Wednesday Addams variant
BM_Rating	The rating a given student left for the Batman variant
CS_Total_Time	Total time spent on computer science variant
CS_Average_Time	Average time spent on computer science variant
WA_Total_Time	Total time spent on Wednesday Addams variant
WA_Average_Time	Average time spent on Wednesday Addams variant
BM_Total_Time	Total time spent on Batman variant
BM_Average_Time	Total time spent on Batman variant
Points_Before	Points student obtained on the first exam before the role-based functionality
Points_After	Points student obtained on the second exam, after the role-based functionality
Point_Difference	Difference in points between the second and first exam

After the subject was completed and the students had the second exam, a questionnaire was sent to each participating student to evaluate their experience with the introduction of this novel approach towards creating and delivering learning materials to students. Ten students answered the questionnaire. Six months after course completion, a new questionnaire was sent out to all students in order to evaluate the long-term effect that our proposed instrument has on student experience. A total of 13 students responded to the second questionnaire. The questions on both questionnaires are different and are presented in Table 4.

Table 4. Questions sent to the students to evaluate their experience with the proposed instrument, immediately after course completion and six months later.

Question	Type	Questionnaire
How would you rate your overall experience with the course content variants that were made available to you?	5-point Likert	First, after subject completion
How well do you feel you understood the material presented in each lesson?	5-point Likert	First, after subject completion
The role-playing teaching approach contributed positively to my learning experience.	5-point Likert	First, after subject completion
The teaching styles used in this course made the lessons more memorable.	5-point Likert	First, after subject completion
I would recommend this style of teaching to other students.	Yes/No	First, after subject completion
If you had to choose only one variant for all lessons for the rest of your studies, what would it be?	Multiple choice	First, after subject completion
Can you recall a specific concept or lesson from the course that has stayed with you? Please elaborate.	Open-ended	Second, 6 months later
In what ways do you think the role-based approach was more or less effective than traditional methods?	Open-ended	Second, 6 months later
Considering your experience, if you had the option, would you choose this style of teaching for future courses? Why or why not?	Open-ended	Second, 6 months later
Is there anything else you would like to share about your experience with the role-based teaching approach, particularly in terms of its long-term effects?	Open-ended	Second, 6 months later
If you had to choose only one variant for all lessons for the rest of your studies, what would it be?	Multiple choice	Second, 6 months later
Would you recommend this style of teaching to other students or educators?	Yes/No	Second, 6 months later

3.4. Data Analysis

Before analyzing the data collected from the students, all data underwent an anonymization process in order to make the results of individual students indistinguishable under observation. As a preparatory step preceding the analysis, we removed one outlier student whose time spent on the LMS was more than three standard deviations away from the class's mean value, leaving 19 students for further analysis.

For the quantitative analysis, we compared the time students spent on each variant and their results on the first exam, the second exam, and the difference in the points between the two exams to identify a possible correlation between these two. The dataset of 19 students is sparse for getting beneficial insights by performing statistical analysis, so no specific statistical tests were employed since the results would be insignificant before more data become available. First, students were categorized into two distinct groups based on their engagement time with the platform. This engagement time was quantified by summing the durations recorded in each content variant for each student. The median value of these aggregate times served as the threshold for classification. Students whose total engagement time exceeded the median were labeled as 'more active', indicating that their usage was above the median level of engagement for the cohort. Conversely, students whose engagement time was equal to or less than the median were classified as 'less active', reflecting a level of platform interaction that was at or below the median engagement of their peers. This time-on-task measurement approach is heavily used in educational contexts, particularly for predicting students' success [81]. Another approach would be to apply machine learning algorithms like K-means or decision tree classifiers [82], but the sample size is not fit for cluster analysis [83]. The median-based division provides a means to categorizing students into relative levels of activity, facilitating the comparative analysis of behavioral patterns and outcomes associated with differing degrees of platform engagement, and still enabling the comparison—even with a sparse dataset.

For the qualitative part of the analysis, we looked at the results from both questionnaires. Qualitative questionnaires have been reported as a successful method for information studies [84]. The first questionnaire mostly consisted of Likert-scale-type questions. To analyze these results, the ratings were coded into three categories: high satisfaction (ratings 4–5), medium satisfaction (ratings 2–3), and low satisfaction (rating 1). This dimensionality reduction from five to three categories was performed as a consequence of the sparse dataset, even though the 7-point Likert scale was reported to provide the best results [85]. A future work on this topic with more students should consider delivering the questionnaires in different manners. The analysis focused on identifying patterns within these categories, particularly looking for correlations between high engagement and satisfaction ratings.

For the second questionnaire, which mostly consisted of open-ended questions, the answers were manually coded in the following categories: appreciated the quizzes, appreciated the characters, preferred the AI content, preferred traditional approach, and skeptical. These categories were identified after analyzing the students' answers. Each student can be classified in multiple categories. The change in recommendation preferences and preferred variants was recorded and analyzed where possible.

4. Results

The results section will be subdivided into two sections containing the quantitative and qualitative results. The quantitative results serve to give us a general overview of the platform usage by students and will help determine whether there is a preferred variant for learning. A total of 20 students interacted with the AI-generated learning materials throughout the subject duration. Among the 20 students, one was identified as an outlier and removed from the dataset. After the subject was complete the aggregate results of the AI content usage were collected, surveys were conducted, and the results from the analysis are presented in the following subsections.

4.1. Quantitative Results

The results presented in this section showcase how the initial introduction of the gamified role-based instruction method reflects on student learning. Figure 4 contains the results for all of the 19 students who interacted with the AI-generated content after outliers were removed and the time they spent on each content variant. The bar charts represent the total amount of time in seconds that each student spent on a given variant.
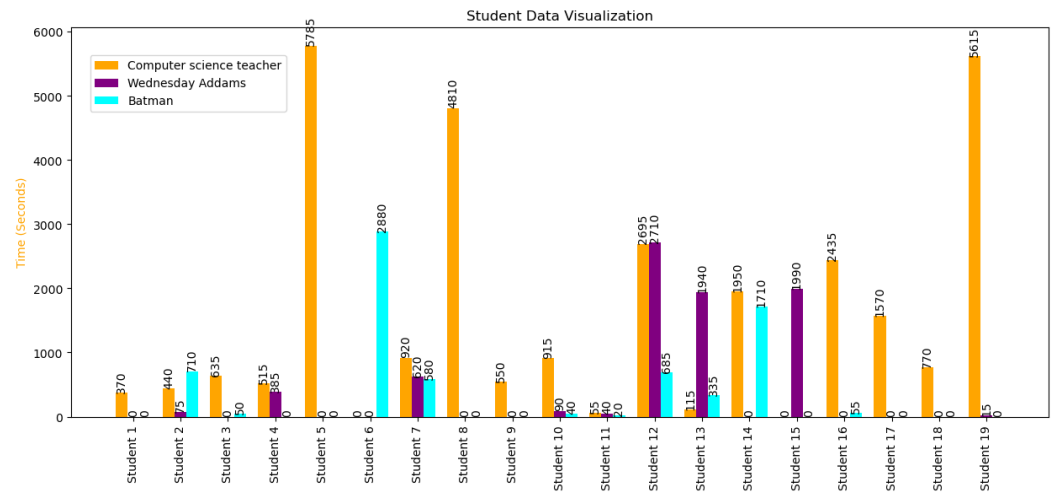

Figure 4. Amount of time in seconds that each student spends on each content variant.

Based on the data presented in Figure 4, it is evident that the computer science teacher version is the most frequently viewed content variant. Although the students voted and decided which fictional characters would be made available, they were more interested in the most traditional style of the content. Upon analyzing the cumulative amount of time allocated to each variation by all the participating students in this study, we obtained the findings presented in Table 5.

Table 5. Total amount of time students spent on each variant.

Variant	Time
Computer science teacher	8.37 h
Batman	2.18 h
Wednesday Addams	6.03 h

If we assume that the computer science teacher variant is the equivalent of a book-like material, then the introduction of role-based variants of the same content doubled the time that the students spent studying. This claim has to be confirmed by studies using larger sample sizes. A strong catalyst for this reported behavior of students is the seamless integration of LLMs in the existing LMSs, thus abstracting away any fears or misconceptions students may be having regarding artificial intelligence. LMSs can still remain the one-stop shops for managing the learning process as discussed in the background section.

Regarding student achievement, Figures 5 and 6 show the distribution of points on exams before and after the introduction of the new feature, after splitting the students into more active and less active groups using the median of the platform time usage as a cutoff. In Figure 5, we compare the results on the first exam of students who would later belong in either the less active or the more active group, after introducing the AI-generated content variants. We use this approach in order to spot patterns in student achievement and to look for evident changes in student performance before and after introducing the proposed approach for automatic AI generation of learning materials. In Figure 5, we see that, on the first exam, more students with higher points will later belong to the less active group.
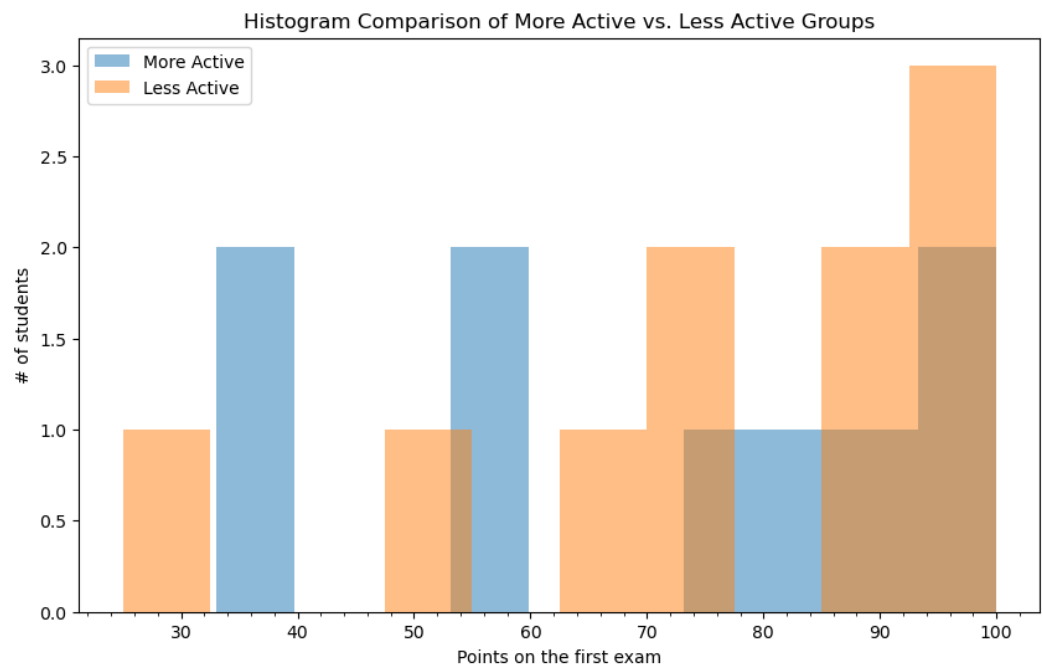
Figure 5. The distribution of points on the exam before introducing the new feature, assuming more and less active groups as defined after subject completion.
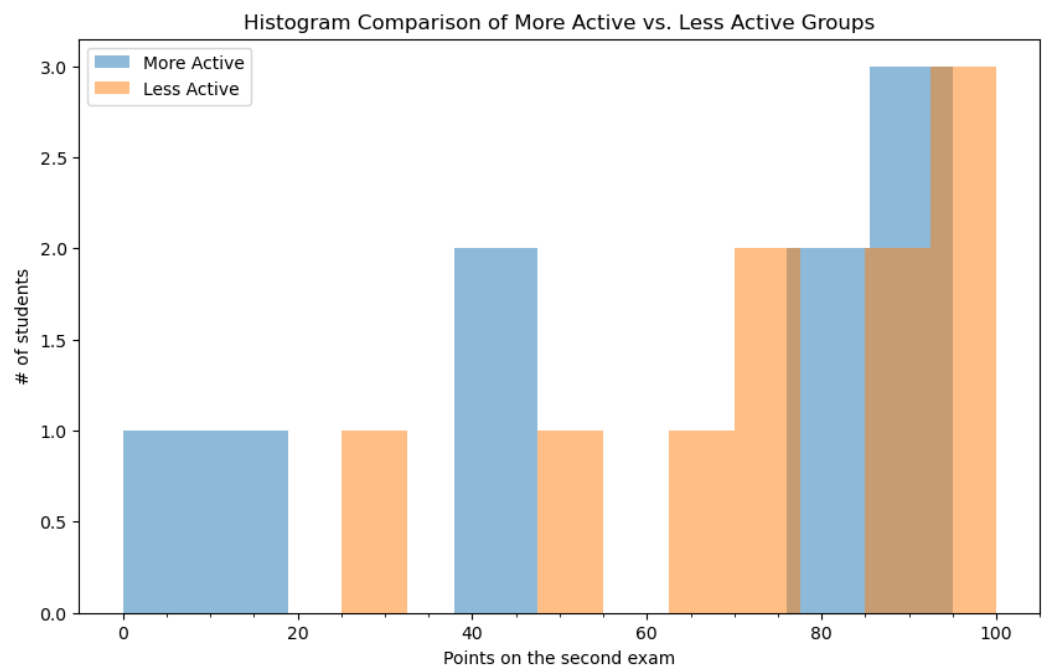

Figure 6. The distribution of points on the second exam. The more active group used the platform for longer than the median time and the less active group used it for a shorter time than the median time.

Figure 6 contains the results that students obtained on the second exam, after introducing the proposed approach. The more active and less active groups comprise the same students for reasons previously discussed. The data shown in this figure suggest that there is a slight improvement in points for students belonging to the more active group. The number of students who score over 80 points on the second exam is equal between the two groups in contrast to the first exam. On the other side, there are students from the more active group who scored poorly on the second exam, despite using access to proposed tool.

As justified earlier, the sparse dataset does not allow for bringing strong decisions from quantitative analysis, but the skew toward a greater amount of points for students who spent more time on the character-based AI-generated materials looks promising for further investigation.

4.2. Qualitative Results

For performing the qualitative analysis, we look at the data from the two questionnaires independently, after which we compare the related data on both questionnaires. The data from the first questionnaire mostly consist of Likert-scale-type questions which use three categories, as described in the methodology section. The data do not reveal much since there is only one student who expressed medium satisfaction with the conclusions that role-based teaching positively contributed to the learning experience and made lessons more memorable. All students would recommend this method to a friend, and regarding the preferred variant, five students chose the traditional computer science teacher, four students chose Batman, and one student chose Wednesday Addams. An overview of the preferred versions of both questionnaires is given in Table 6.

Table 6. Preferred variants of the learning content after the first questionnaire immediately after subject completion and 6 months after subject completion.

Variant	Votes (after Subject Completion)	Votes (6 Months after Subject Completion)
Computer science teacher	5	10
Batman	4	2
Wednesday Addams	1	0

The second questionnaire, sent to students 6 months after subject completion to evaluate the possible long-term effects of the proposed approach, mostly consisted of open-ended questions. After analyzing the answers, five categories were identified where all students belong. The first one is for students who appreciate the quizzes. Out of the 13 students who answered the survey, 4 identified the quizzes as the best part of the AI-generated content since it helped them evaluate their knowledge. The next one is for students who appreciated the characters, where four students identified themselves as advocates because they engaged with the roles. Two students found the platform more effective for accomplishing their learning goals and another two students enjoyed the traditional way better. Only one student expressed skepticism about this approach and questioned the content's reliability.

Regarding the recommendation preferences, while on the first questionnaire, all students responded that they were likely to recommend the approach to a friend; in the second questionnaire, one student identified as not willing to recommend the carried approach. Probably the most interesting change is noticed in the preferred variant version shown in Table 6, with students equally divided between fictional characters and the traditional way on the first questionnaire, while six months later, 10 students identified the traditional computer science teacher as their preferred variant, compared to only two students identifying the fictional character style as such.

5. Discussion

5.1. LLMS for Developing Tutoring Personas

The approach suggested in this paper for achieving a sustainable delivery of engaging learning materials to students relies on utilizing generative artificial intelligence for content generation that will seamlessly be integrated into the learning management systems that most educational organizations are already utilizing. The step-by-step instructions for seamlessly inter-operating a given system with OpenAI's, presented in this paper, contributes toward bridging the gap between cutting-edge artificial intelligence and current educational frameworks [59,60]. In the integration process, two primary types of messages facilitate the communication between the OpenAI model and the LMS: system messages

and user messages. System messages originate from the LMS, should not be changed by the client, or in our case, the teacher accessing the LMS and it instructs the model how to behave. The user message contains the actual query and can be generated from the inputs provided by the teacher within the LMS, which are then processed by the OpenAI model to generate content. This dual-message communication enables the effortless incorporation of AI-generated materials into the standard operational workflow and expands the current state of available research in educational technology [63,64].

The innovative part proposed is that, rather than sole automatic content generation, it can be achieved in different variants in an effort to personalize the experience for each student. By customizing the system's message to reflect a given user's choice, the LLM can be tailored to give different results, which is exactly how this study was conducted. One of the main benefits of this proposition is that it is fully automated and does not introduce any additional overheads neither for the teachers nor for the students. In contrast to Morze et al. [61], whose findings were discussed in the background section, our proposed approach takes the burden off teachers and offers a time-saving and efficient methodology for bringing AI, and especially adaptive learning, to the classroom.

Pop-culture-inspired fictional characters were used in the current iteration of the experiment with content delivered in the style of Batman and Wednesday Addams alongside the traditional teacher style. Customizable lesson creation was already identified as a key driver for advancing improvements of the educational process by multiple researchers outlined in the background section. The reason why there were only three variants offered is the cost of using the OpenAI's API at the time of performing the experiment. At the moment of writing, this paper, a new developer-friendly pricing opens opportunities to offer the content in many more variants and even make this functionality on an on-demand basis.

The current method utilizes a workaround to obtain the content followed by a JSON-formatted array of questions which requires additional validation on the receiving side to validate the data. In the latest OpenAI API version, there is an option to instruct the model to only communicate using the JSON format using the `{ "type": "json_object" }` parameter which should further simplify the integration process.

5.2. Student's Engagement with AI-Tutors

Certain amounts of insightful information came up while performing the analysis. Firstly, the experiment was well received by the students. They actively participated in the process of selecting which characters would be included in the role-based teaching, and all but one answered that they are likely to recommend the role-based teaching approach with AI-generated content to a friend or colleague. Students also responded that they would like to see the proposed method implemented in more classes, which advocates for the effectiveness of the proposed solution. The finding that the introduction of the AI tool was well received by students is again in line with the exploratory experiments from the background section that also reported the good acceptance of novelty approaches in the classroom [25,49,50].

The data from the second questionnaire reveal that students mostly appreciated the automatically generated questions with immediate feedback since, prior to that, they did not have an easy means of checking their knowledge. This is valuable information, especially knowing that the frequent testing of student knowledge has been proven to improve student scores and knowledge in general [86,87]. This finding is only drawn from automating assessment in its most primitive form—by offering multiple choice questions. Not only can AI models help students with assessing them this way, but they offer a broader spectrum of possibilities including generating feedback and explanations on open-ended questions, enabling multiple attempts with unique content, as well as providing personalized learning materials in accordance with the students' activity on the self-assessments. Considering this, the finding is important since it justifies further research in this field and exploring more opportunities and improvements.

A surprising change is observed in the preferred content variant. During the six months between the two surveys, the students' preference shifted towards the style of the traditional teacher. Although the content of all three variants was completely AI-generated, students on the first questionnaire were divided when preferring either the traditional or fictional option, while six months after subject completion, the traditional way was a clear winner. This is a probable indication that fictional characters are good for short-term engagement and keeping students entertained while learning, while for long-term success, the traditional approach is still the preferred variant.

In the quantitative analysis part, we saw a skew toward a higher amount of points for students who were actively using the part of the LMS with AI-generated content. The students who scored good results on the first exam, before introducing AI-generated variants, did not spend much time on this novel approach to delivering learning materials. Still, the students who actively used it improved the overall score of the group, which leads us to the implication that the different variants of the learning materials are most beneficial for students who have not mastered the topic otherwise. Looking at the amount spent on each variant, it can be noted that the amount of time students spent on reading fictional character-based content is equal to the amount of time spent on reading the traditional format. This can be considered a huge win for the proposed approach, although we do not know whether the students who spent time learning from fictional characters would have spent the time on the traditional variant if the former ones did not exist. Although this finding requires additional validation through studies with bigger sample sizes, it is especially important considering previous research reports about students spending less time reading [10,88].

5.3. Limitations

The current study has two main limitations: the small sample size and the fixed amount of available variants, both of which are easily addressable and suitable for future research.

The first one is the sparse dataset and the low amount of students who participated in the experiment. The experiment was performed with only 19 students, which makes drawing clear conclusions difficult and risky. Although the proposed methodology does not have any consequences and risks, as can be seen from the analysis of the current students, a larger population is needed in order to confirm the improvements from using the proposed approach, including the fact that it supports individual instruction, enables self-paced learning and immediate feedback, improves students' success, increases time spent learning, and makes learning more fun and engaging.

The second limitation is that the number of available variants of learning materials was limited to three. We wonder and are set to explore whether having an unlimited amount of variants will further increase the time that students spend with the learning material. For this, we are working on implementing an on-demand feature for variant creation, where students can access the learning material, ask for it to be presented in a certain way, and obtain the results back automatically and immediately.

6. Conclusions

The educational experience of students varies greatly from school to school, depending on the learning and teaching methodologies adopted by each school. Still, increasing student engagement stands out as the obvious action for improving student success and as a solution to Bloom's two-sigma problem [15]. The personalization of the educational journey to meet the needs of each and every student has been identified as the most likely action to help increase student engagement. Still, until recently, achieving a completely personalized learning environment was only conceptualized and difficult to implement in the physical classroom. The introduction of generative artificial intelligence to the general public and its increasing accessibility and affordability offer a way to overcome this obstacle and possibly disrupt the established educational process. In this paper, we suggested and tested an approach with AI-generated content in different variants for

achieving personalized learning experiences, which results in several implications for both theory and practice.

6.1. Implications for Theory

The results presented in this study showed no negative implications for students in terms of their access to automatically generated content, and only a positive impact on low-performing students when such access is granted. The availability of different content variants for students, to support them in their learning, was proved to have positive impact on their performance. The suggested approach led to an increase in the students' learning time, whether it happened out of need or curiosity and with it contributed towards the creation of new pedagogical methods that will drive the education of the future.

One other finding of the current experiment is that the students who used AI-generated content the most were not the best students, but rather the students who would otherwise probably have difficulties with the subject. This inspired us to consider delivering different levels of learning materials to different students. Students who have not fully grasped an adequate comprehension of the material would also benefit, besides from delivering more variants and in different formats, from more thorough and detailed learning materials. On the contrary, for highly proficient students who do not require several versions or methods of distribution to understand the content, an alternative strategy should be contemplated. This involves selecting more precise and even advanced subject matter, regardless of the format or media used. LLMs are well positioned to make this possible by customizing the approach proposed in this experiment to tailor the content for each student or group of students. Undoubtedly, taking these kinds of actions can lead to a future education that is more sustainable, personalized, and engaging.

With the accessibility and the capabilities of AI in the coming years set to grow at a fast pace, this paper contributes toward narrowing the identified gap in research on utilizing large language models in the educational context. Apart from generating only text, new models are able to generate images, illustrations, and even videos, which will open a huge field for further research and improvement. With companies like SoulMachine [89] already offering the creation of GPT-powered digital people, which are extremely suitable for incorporating in the classroom setting, personalized learning finally seems achievable, while at the same time fighting some of the problems of the current education like teacher shortages [90], high dropout rates [91], and inequality in education [92].

6.2. Implications for Practice

As discussed in the background section when identifying research gaps, one of the main gaps for generative AI adoption in education is the scarcity of empirical studies. Very few studies feature learning material creation using AI and even fewer bring this kind of generated content to the classroom to measure its effectiveness. This study helps fill these gaps with a sophisticated and fully documented approach, from its inception to the evaluation of its results.

The complete steps for reproducing the study and mirroring the API requests to OpenAI's API were provided in the previous sections, making it fully reproducible. We expect this study to inspire other researchers and decision-makers in educational institutions to consider following our approach and to implement similar features in order to improve the students' study experience.

As we look forward, our research will expand to examine various mediums of content delivery beyond text-based lessons. The increasing affordability and accessibility of video content generation and text-based LLMs offer fertile ground for extending our research. We aim to explore a broader spectrum of content variants and incorporate more frequent AI-generated knowledge assessments, thereby enriching the educational landscape with innovative and effective learning tools. Additionally, we will adopt the hereby proposed approach to multiple subjects, which will engage more students and will help fight the current limitations in the form of the sparse dataset. These actions will help us to have

more precise results, draw better conclusions, and confirm the findings of the current study with greater certainty.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the informed and voluntary participation of students, who, prior to their involvement, were fully briefed about the study's aims, methods, and potential impacts on their educational journey. This process was facilitated by a clause in their study contracts, consenting to participation in research aimed at enhancing their educational experiences, in line with our institution's statutes that promote research, innovation, and advancement. Given the comprehensive informed consent, the contractual agreement on participation in research and the strict adherence to institutional guidelines advocating for a research-intensive environment, we assert that the waiver of ethical review and approval is justified. Our approach meticulously balances ethical considerations, participant autonomy, and the pursuit of educational and research excellence, adhering to the highest standards of integrity and positively contributing to the academic and personal development of our students.

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

Data Availability Statement: The data supporting the findings of this study, which consist of student grades, performance and satisfaction over time collected at a private college, are not publicly available to maintain the confidentiality and integrity of the educational institution's records. Nonetheless, the data are available from the corresponding author upon reasonable request.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
API	Application Programming Interface
AQG	Automatic Question Generation
ESD	Education for Sustainable Development
GPT	Generative Pre-trained Transformer
JSON	JavaScript Object Notation
LLM	Large Language Model
LMS	Learning Management System
OOP	Object-Oriented Programming
SDG	Sustainable Development Goals
TEL	Technology-Enhanced Learning

References

1. Davies, R.S.; West, R.E. Technology integration in schools. In *Handbook of Research on Educational Communications and Technology*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 841–853.
2. Mousavinasab, E.; Zarifsanaiey, N.; R. Niakan Kalhori, S.; Rakhshan, M.; Keikha, L.; Ghazi Saeedi, M. Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interact. Learn. Environ.* **2021**, *29*, 142–163. [CrossRef]
3. Bergmann, J.; Sams, A. *Flip your Classroom: Reach Every Student in Every Class Every Day*; International Society for Technology in Education: Washington, DC, USA, 2012.
4. Akçayır, G.; Akçayır, M. The flipped classroom: A review of its advantages and challenges. *Comput. Educ.* **2018**, *126*, 334–345. [CrossRef]
5. Callaghan, N. Understanding the role of technological platforms in schools. *Educ. Media Int.* **2021**, *58*, 355–373. [CrossRef]
6. Alhumaid, K. Four ways technology has negatively changed education. *J. Educ. Soc. Res.* **2019**, *9*, 10. [CrossRef]
7. Dror, I.E. Technology enhanced learning: The good, the bad, and the ugly. *Pragmat. Cogn.* **2008**, *16*, 215–223. [CrossRef]

8. Raja, R.; Nagasubramani, P. Impact of modern technology in education. *J. Appl. Adv. Res.* **2018**, *3*, 33–35. [CrossRef]

9. Pesovski, I.; Klashninovski, A.; Makeska, A.; Trajkovik, V. The Impact of Learning Management Systems on Student Achievement and Engagement in Online Classrooms. In Proceedings of the EDULEARN22 Proceedings, IATED, Palma, Spain, 4–6 July 2022; pp. 6451–6456.

10. Juban, R.L.; Lopez, T.B. An exploration of textbook reading behaviors. *J. Educ. Bus.* **2013**, *88*, 325–331. [CrossRef]

11. Bergdahl, N.; Nouri, J.; Fors, U.; Knutsson, O. Engagement, disengagement and performance when learning with technologies in upper secondary school. *Comput. Educ.* **2020**, *149*, 103783. [CrossRef]

12. Rashid, T.; Asghar, H.M. Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Comput. Hum. Behav.* **2016**, *63*, 604–612. [CrossRef]

13. Martin, F.; Bolliger, D.U. Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learn.* **2018**, *22*, 205–222. [CrossRef]

14. Abou-Khalil, V.; Helou, S.; Khalifé, E.; Chen, M.A.; Majumdar, R.; Ogata, H. Emergency online learning in low-resource settings: Effective student engagement strategies. *Educ. Sci.* **2021**, *11*, 24. [CrossRef]

15. Bloom, B.S. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educ. Res.* **1984**, *13*, 4–16. [CrossRef]

16. Pataranutaporn, P.; Danry, V.; Leong, J.; Punpongsanon, P.; Novy, D.; Maes, P.; Sra, M. AI-generated characters for supporting personalized learning and well-being. *Nat. Mach. Intell.* **2021**, *3*, 1013–1022. [CrossRef]

17. Pedro, F.; Subosa, M.; Rivas, A.; Valverde, P. *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*; United Nations Educational, Scientific and Cultural Organization, Unesco: Paris, France, 2019.

18. Boeren, E. Understanding Sustainable Development Goal (SDG) 4 on "quality education" from micro, meso and macro perspectives. *Int. Rev. Educ.* **2019**, *65*, 277–294. [CrossRef]

19. Tejedor, G.; Segalàs, J.; Barrón, Á.; Fernández-Morilla, M.; Fuertes, M.T.; Ruiz-Morales, J.; Gutiérrez, I.; García-González, E.; Aramburuzabala, P.; Hernández, À. Didactic strategies to promote competencies in sustainability. *Sustainability* **2019**, *11*, 2086. [CrossRef]

20. Abulibdeh, A.; Zaidan, E.; Abulibdeh, R. Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions. *J. Clean. Prod.* **2024**, *437*, 140527. [CrossRef]

21. Devi, D.; Rroy, A.D. Role of Artificial Intelligence (AI) in Sustainable Education of Higher Education Institutions in Guwahati City: Teacher's Perception. *Int. Manag. Rev.* **2023**, *19*, 111–116.

22. Klašnja-Milićević, A.; Ivanović, M. E-learning personalization systems and sustainable education. *Sustainability* **2021**, *13*, 6713. [CrossRef]

23. Grassini, S. Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Educ. Sci.* **2023**, *13*, 692. [CrossRef]

24. Qadir, J. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In Proceedings of the 2023 IEEE Global Engineering Education Conference (EDUCON), Salmiya, Kuwait, 1–4 May 2023; pp. 1–9.

25. Lee, U.; Jung, H.; Jeon, Y.; Sohn, Y.; Hwang, W.; Moon, J.; Kim, H. Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in english education. *Educ. Inf. Technol.* **2023**, 1–33. [CrossRef]

26. Jauhiainen, J.S.; Guerra, A.G. Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability* **2023**, *15*, 14025. [CrossRef]

27. Gates, B. AI Is about to Completely Change How You Use Computers. Available online: https://www.gatesnotes.com/AI-agents (accessed on 18 January 2024).

28. Kunicina, N.; Zabasta, A.; Nikiforova, O.; Romanovs, A.; Patlins, A. Modern tools of career development and motivation of students in Electrical Engineering Education. In Proceedings of the 2018 IEEE 59th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, Latvia, 12–13 November 2018; pp. 1–6.

29. Bond, M.; Khosravi, H.; De Laat, M.; Bergdahl, N.; Negrea, V.; Oxley, E.; Pham, P.; Chong, S.W.; Siemens, G. A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *Int. J. Educ. Technol. High. Educ.* **2024**, *21*, 4. [CrossRef]

30. Dunn, T.J.; Kennedy, M. Technology Enhanced Learning in higher education; motivations, engagement and academic achievement. *Comput. Educ.* **2019**, *137*, 104–113. [CrossRef]

31. Balkaya, S.; Akkucuk, U. Adoption and use of learning management systems in education: The role of playfulness and self-management. *Sustainability* **2021**, *13*, 1127. [CrossRef]

32. Coates, H.; James, R.; Baldwin, G. A critical examination of the effects of learning management systems on university teaching and learning. *Tert. Educ. Manag.* **2005**, *11*, 19–36. [CrossRef]

33. Romero, C.; López, M.I.; Luna, J.M.; Ventura, S. Predicting students' final performance from participation in on-line discussion forums. *Comput. Educ.* **2013**, *68*, 458–472. [CrossRef]

34. Bernacki, M.L.; Chavez, M.M.; Uesbeck, P.M. Predicting achievement and providing support before STEM majors begin to fail. *Comput. Educ.* **2020**, *158*, 103999. [CrossRef]

35. Chan, Y.M. Video instructions as support for beyond classroom learning. *Procedia-Soc. Behav. Sci.* **2010**, *9*, 1313–1318. [CrossRef]

36. Arias, M.; Creus, C.; Gascón, A.; Godoy, G. LEARNING THEORY THROUGH VIDEOS—A Teaching Experience in a Theoretical Course based on Self-learning Videos and Problem-solving Sessions. In Proceedings of the International Conference on Computer Supported Education, Noordwijkerhout, The Netherlands, 6–8 May 2011; SciTePress: Setúbal, Portugal, 2011; Volume 2, pp. 93–98.

37. Riestra-González, M.; del Puerto Paule-Ruíz, M.; Ortin, F. Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Comput. Educ.* **2021**, *163*, 104108. [CrossRef]

38. Santos, R.M.; Henriques, R. Accurate, timely, and portable: Course-agnostic early prediction of student performance from LMS logs. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100175. [CrossRef]

39. Macfadyen, L.P.; Dawson, S. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Comput. Educ.* **2010**, *54*, 588–599. [CrossRef]

40. Fonseca, D.; Martí, N.; Redondo, E.; Navarro, I.; Sánchez, A. Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models. *Comput. Hum. Behav.* **2014**, *31*, 434–445. [CrossRef]

41. Goel, A.K.; Polepeddi, L.; Watson, J. *Learning Engineering for Online Education: Theoretical Contexts and Design-Based Examples*; Routledge: London, UK, 2018.

42. Patlins, A.; Kunicina, N.; Ribickis, L. Information tools for education of electrical engineers. In Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, Prague, Czech Republic, 15–17 September 2011; IEEE: Piscataway, NJ, USA, 2011; Volume 2, pp. 660–665.

43. Zhiravetska, A.; Kunicina, N.; Berzina, K.; Patlins, A. Flexible approach to course testing for the improvement of its effectiveness in engineering education. In Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Warsaw, Poland, 24–26 September 2015; IEEE: Piscataway, NJ, USA, 2015; Volume 2, pp. 955–959.

44. Mondal, S.; Das, S.; Vrana, V.G. How to bell the cat? A theoretical review of generative artificial intelligence towards digital disruption in all walks of life. *Technologies* **2023**, *11*, 44. [CrossRef]

45. Ng, A.; Jordan, M. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In Proceedings of the Advances in Neural Information Processing Systems 14 (NIPS 2001), Vancouver, BC, Canada, 3–8 December 2001; Volume 14.

46. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv* **2023**, arXiv:2307.09288

47. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]

48. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 1–27. [CrossRef]

49. Wang, Z.; Lan, A.S.; Baraniuk, R.G. Math word problem generation with mathematical consistency and problem context constraints. *arXiv* **2021**, arXiv:2109.04546.

50. Bhat, S.; Nguyen, H.A.; Moore, S.; Stamper, J.; Sakr, M.; Nyberg, E. Towards automated generation and evaluation of questions in educational domains. In Proceedings of the 15th International Conference on Educational Data Mining, Durham, UK, 24–27 July 2022; Volume 701.

51. Moore, S.; Nguyen, H.A.; Bier, N.; Domadia, T.; Stamper, J. Assessing the quality of student-generated short answer questions using GPT-3. In Proceedings of the European Conference on Technology Enhanced Learning, Toulouse, France, 12–16 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 243–257.

52. Pinto, G.; Cardoso-Pereira, I.; Monteiro, D.; Lucena, D.; Souza, A.; Gama, K. Large Language Models for Education: Grading Open-Ended Questions Using ChatGPT. In Proceedings of the XXXVII Brazilian Symposium on Software Engineering, Campo Grande, Brazil, 25–29 September 2023; pp. 293–302.

53. Sridhar, P.; Doyle, A.; Agarwal, A.; Bogart, C.; Savelka, J.; Sakr, M. Harnessing llms in curricular design: Using gpt-4 to support authoring of learning objectives. *arXiv* **2023**, arXiv:2306.17459.

54. Sarsa, S.; Denny, P.; Hellas, A.; Leinonen, J. Automatic generation of programming exercises and code explanations using large language models. In Proceedings of the 2022 ACM Conference on International Computing Education Research, Lugano, Switzerland, 7–11 August 2022; Volume 1, pp. 27–43.

55. MacNeil, S.; Tran, A.; Mogil, D.; Bernstein, S.; Ross, E.; Huang, Z. Generating diverse code explanations using the gpt-3 large language model. In Proceedings of the 2022 ACM Conference on International Computing Education Research, Lugano, Switzerland, 7–11 August 2022; Volume 2, pp. 37–39.

56. Bernius, J.P.; Krusche, S.; Bruegge, B. Machine learning based feedback on textual student answers in large courses. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100081. [CrossRef]

57. Sailer, M.; Bauer, E.; Hofmann, R.; Kiesewetter, J.; Glas, J.; Gurevych, I.; Fischer, F. Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learn. Instr.* **2023**, *83*, 101620. [CrossRef]

58. Pesovski, I.; Bogdanova, A.M.; Trajkovik, V. Systematic review of the published explainable educational recommendation systems. In Proceedings of the 2022 20th International Conference on Information Technology Based Higher Education and Training (ITHET), Antalya, Turkey, 7–9 November 2022; pp. 1–8.

59. Sougleridi, E.I.; Kopsidas, S.; Vavougios, D.; Avramopoulos, A.; Kanapitsas, A. Embedding AI into LMS and eLearning Platforms. In Proceedings of the International Conference on Interactive Collaborative Learning, Madrid, Spain, 26–29 September 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 363–368.

60. Pesovski, I.; Madevska-Bogdanova, A.; Trajkovik, V. Reproducibility of Published Educational Recommendation Systems: Systematic Review. In Proceedings of the ICT Innovations 2022, Web Proceedings ISSN null, Skopje, North Macedonia, 29 September–1 October 2022; pp. 179–191.

61. Morze, N.; Varchenko-Trotsenko, L.; Terletska, T.; Smyrnova-Trybulska, E. Implementation of adaptive learning at higher education institutions by means of Moodle LMS. *J. Phys. Conf. Ser.* **2021**, *1840*, 012062. [CrossRef]

62. Aldahwan, N.; Alsaeed, N. Use of artificial intelligent in Learning Management System (LMS): A systematic literature review. *Int. J. Comput. Appl.* **2020**, *175*, 16–26. [CrossRef]

63. Firat, M. Integrating AI applications into learning management systems to enhance e-learning. *Instr. Technol. Lifelong Learn.* **2023**, *4*, 1–14. [CrossRef]

64. Villegas-Ch, W.; Palacios-Pacheco, X. Integration of artificial intelligence as a tool for an online education model. In Proceedings of the International Conference on Innovation and Research, Sangolqui, Ecuador, 18–19 June 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 94–105.

65. Villegas-Ch, W.; Román-Cañizares, M.; Palacios-Pacheco, X. Improvement of an online education model with the integration of machine learning and data analysis in an LMS. *Appl. Sci.* **2020**, *10*, 5371. [CrossRef]

66. Zirpoli, C.T. *Generative Artificial Intelligence and Copyright Law*; University of Nebraska: Lincoln, NE, USA, 2023.

67. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things-Cyber-Phys. Syst.* **2023**, *3*, 121–154. [CrossRef]

68. Ferrara, E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci* **2023**, *6*, 3. [CrossRef]

69. Gichoya, J.W.; Thomas, K.; Celi, L.A.; Safdar, N.; Banerjee, I.; Banja, J.D.; Seyyed-Kalantari, L.; Trivedi, H.; Purkayastha, S. AI pitfalls and what not to do: Mitigating bias in AI. *Br. J. Radiol.* **2023**, *96*, 20230023. [CrossRef] [PubMed]

70. Roselli, D.; Matthews, J.; Talagala, N. Managing bias in AI. In Proceedings of the Companion Proceedings of the 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 539–544.

71. Xu, Z.; Jain, S.; Kankanhalli, M. Hallucination is inevitable: An innate limitation of large language models. *arXiv* **2024**, arXiv:2401.11817.

72. Yao, J.Y.; Ning, K.P.; Liu, Z.H.; Ning, M.N.; Yuan, L. Llm lies: Hallucinations are not bugs, but features as adversarial examples. *arXiv* **2023**, arXiv:2310.01469.

73. Ji, Z.; Yu, T.; Xu, Y.; Lee, N.; Ishii, E.; Fung, P. Towards mitigating LLM hallucination via self reflection. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023.

74. Wei, J.; Yao, Y.; Ton, J.F.; Guo, H.; Estornell, A.; Liu, Y. Measuring and Reducing LLM Hallucination without Gold-Standard Answers via Expertise-Weighting. *arXiv* **2024**, arXiv:2402.10412.

75. Galitsky, B.A. Truth-o-meter: Collaborating with llm in fighting its hallucinations. *Preprints* **2023**, 2023071723. [CrossRef]

76. van den Berg, G.; du Plessis, E. ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Educ. Sci.* **2023**, *13*, 998. [CrossRef]

77. Mikeladze, T. Creating teaching materials with ChatGPT. In Proceedings of the IRCEELT—2023 13th International Research Conference on Education, Tbilisi, Georgia, 5–6 May 2023.

78. Koraishi, O. Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Lang. Educ. Technol.* **2023**, *3*, 55–72.

79. Rhys Cox, S. The Use of Multiple Conversational Agent Interlocutors in Learning. *arXiv* **2023**, arXiv:2312.16534.

80. Kapetanović, A. Batman as a Cultural Artefact. Ph.D. Thesis, Department of English, University of Zadar, Zadar, Croatia, 2016.

81. Kovanovic, V.; Gašević, D.; Dawson, S.; Joksimovic, S.; Baker, R. Does time-on-task estimation matter? Implications on validity of learning analytics findings. *J. Learn. Anal.* **2015**, *2*, 81–110. [CrossRef]

82. Yathongchai, C.; Angskun, T.; Yathongchai, W.; Angskun, J. Learner classification based on learning behavior and performance. In Proceedings of the 2013 IEEE Conference on Open Systems (ICOS), Kuching, Sarawak, Malaysia, 2–4 December 2013; pp. 66–70. [CrossRef]

83. Sarstedt, M.; Mooi, E. Cluster Analysis. In *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 301–354. [CrossRef]

84. Eckerdal, J.R.; Hagström, C. Qualitative questionnaires as a method for information studies research. In Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, 27–29 June 2017.

85. Taherdoost, H. What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. *Hamed Taherdoost* **2019**, *8*, 1–10.

86. Pesovski, I.; Trajkovik, V. The Influence of Frequent Quiz-Based Exams on Learning Performance Among Computer Science Students. In Proceedings of the ICERI2022 Proceedings, Seville, Spain, 7–9 November 2022; IATED: Valencia, Spain, 2022; pp. 5693–5699.

87. Pesovski, I.; Klashninovski, A.; Makeska, A.M. The Effect of Frequent Testing on Student Performance. In Proceedings of the International Scientific Conference "75th Anniversary of the Institute of Pedagogy—Educational Challenges and Future Prospects", Ohrid, North Macedonia, 16–18 May 2022.

88. Manuel, L.C. The Effect of Technology Resources on College Sophomores' Reading Habits in Ilocos Sur Polytechnic State College Philippines. *J. Pendidik.* **2022**, *11*, 56–70. [CrossRef]

89. Paredes, D. Meet 'Will', the Digital Teacher for Renewable Energy. Available online: https://www.cio.com/article/203932/meet-will-the-digital-teacher-for-renewable-energy.html (accessed on 18 January 2024).

90. Ogunode, N.J.; Edinoh, K.; Okolie, R.C. Artificial intelligence and Tertiary Education Management. *Electron. Res. J. Soc. Sci. Humanit.* **2023**, *5*, 18–31.

91. Agrusti, F.; Bonavolontà, G.; Mezzini, M. University dropout prediction through educational data mining techniques: A systematic review. *J. e-Learn. Knowl. Soc.* **2019**, *15*, 161–182.

92. Holstein, K.; Doroudi, S. Equity and Artificial Intelligence in Education: Will "AIEd" Amplify or Alleviate Inequities in Education? *arXiv* **2021**, arXiv:2104.12920.