Article

Developing Effective Educational Chatbots with GPT: Insights from a Pilot Study in a University Subject

Sánchez-Vera Fulgencio (ID)

Department of Didactics and Educational Research, University of La Laguna, 38200 La Laguna, Santa Cruz de Tenerife, Spain; fsanchev@ull.edu.es

Abstract: This study presents research on the development process of GPT-based educational chatbots. A case study methodology was employed to address the process of designing, implementing, and evaluating a prototype that functioned as a personal tutor for the Sociology of Education course in the Primary Education Teaching Degree. The objective is to provide valuable insights into the processes, challenges, and outcomes of this technology and to determine its potential and limitations as an educational personal tutor. The chatbot underwent laboratory tests, which included real exams from previous courses and other specific assessments. After an iterative refinement process, a final product with optimal results was achieved. This study offers a robust model for the development of GPTs, as well as an analysis of the current possibilities and limitations of this technology for education. The study concludes by emphasizing the importance of continuous innovation and research in the use of emerging technologies like chatbots in education, highlighting their potential to transform traditional teaching methods.

Keywords: chatbots; artificial intelligence; personalized learning; educational technology; GPT

1. Introduction

The integration of Artificial Intelligence (AI) technologies in the educational field represents an area of notable interest and innovation today. This trend has intensified, especially since the end of 2022, driven by the emergence of ChatGPT and other services based on Generative Artificial Intelligence (GAI). AI is playing a crucial role in improving various educational processes, offering innovative and effective solutions. Within this panorama, chatbots have proven to be valuable tools to support learning and teaching, surprising with their wide spectrum of applications and their ability to enrich the educational experience.

AI is a field of computing that focuses on creating artifacts capable of exhibiting intelligent behavior. AI allows machines to simulate and perform tasks that typically require human intelligence, such as logical reasoning, learning, and problem solving. It is based on algorithms and Machine Learning technologies that give machines cognitive and acting capabilities in the world by performing tasks autonomously or semi-autonomously [1]. In particular, GAI, which focuses on generating new content using deep learning techniques [2], has notably boosted the development of chatbots. These, based on Natural Language Processing (NLP) and Machine Learning (ML), offer increasingly advanced and efficient conversational interfaces [3].

To comprehend the genesis and evolution of chatbots, it is instructive to revisit the Turing test, introduced in 1950 by Alan Turing. This test, designed to assess if machines could emulate human thought, involved an interrogator distinguishing between a human and a machine based solely on their written responses [4]. This seminal concept spurred the creation of systems aimed at passing this test. Despite initial skepticism, the Turing Test has retained its relevance, being frequently employed to gauge computer programs' approximation to human reasoning. The endeavor of software developers and AI researchers to pass this test through natural language interaction marks a significant chapter

in chatbot history. Thus, the trajectory of chatbots, spanning over six decades, is deeply interwoven with progress in Artificial Intelligence, Natural Language Processing, and computing technology.

The origins of chatbots can be traced back to the 1960s, with the development of ELIZA by Joseph Weizenbaum at MIT. ELIZA was rudimentary but trailblazing in simulating human conversation, utilizing a script known as DOCTOR that mimicked the dialogue of a therapist [5]. The 1970s saw the emergence of PARRY by Kenneth Colby, a program that simulated a patient with paranoid schizophrenia, signifying a leap forward in AI's capacity to emulate intricate human behavior [6].

In the 1980s and 1990s, even with heightened interest in Artificial Intelligence, chatbots experienced modest advancements. Nonetheless, these decades were critical for the development of natural language understanding and processing. The advent of the new millennium introduced ALICE, which used an Artificial Intelligence Markup Language (AIML), and SmarterChild, which became widespread on instant messaging platforms, both facilitating sophisticated interactions and broader access to online information [7].

The 2010s marked a milestone with the debut of Apple's Siri, heralding the age of voice-activated personal virtual assistants. This innovation led to the introduction of Amazon Alexa, Google Assistant, and Microsoft Cortana. Mid-decade, chatbots started to be integrated into social media platforms like Facebook and Microsoft, enabling automated business-to-customer interactions. Concurrently, advancements in Machine Learning and Natural Language Processing improved the naturalness and contextual relevance of responses.

A pivotal breakthrough was the release of the GPT series by OpenAI in 2018, with subsequent versions, GPT-2 in 2019 and GPT-3 in 2020, marking significant progress in text generation, comprehension, and contextually aware responses. The GPT model, based on the transformer architecture, was trained on an extensive corpus of text, enhancing its ability to discern linguistic patterns and relationships. The latest models are specialized for conversational tasks, leveraging vast internet-sourced text data to produce responses that are increasingly indistinguishable from those of a human [8].

In education, chatbots have become innovative tools, offering substantial support in teaching and learning across various contexts. Transitioning from basic text-based interactions to more complex and adaptable applications, chatbots now cater to diverse age groups and skill levels, assisting in clarifying student uncertainties and encouraging independent learning [9]. They have also demonstrated their potential to enhance online collaboration and student engagement [10,11]. Recent research illustrates chatbots' efficacy in providing academic and administrative information, thus enriching interactions among students, parents, and educators [3,12,13].

In the realm of higher education, chatbots have gained prominence, bolstering student learning and engagement, notably during the COVID-19 pandemic, by playing a vital role in maintaining educational continuity through online learning support [14]. While chatbots simulate human-like conversations and are readily available, their development poses challenges, including the need for substantial investment in building comprehensive and precise knowledge bases, which can be both costly and labor-intensive [15].

Educational examples like Georgia Tech's Jill Watson, Duolingo Bots, Coursera's chatbot, EdBuddy, Botsify, ALEKS, and SnatchBot, among others, illustrate the power and flexibility of chatbots as educational instruments that transform student and educator engagement with learning materials and management of educational activities.

The launch of ChatGPT in 2022, along with other specialized models, has broadened the scope of chatbot applications in fields including education. The latest model, GPT-4, oversees billions of parameters and has been trained on a significantly more extensive dataset than its forerunners. It brings enhancements in safety, reducing inaccurate responses and "hallucinations". These developments in GPT models represent substantial strides in NLP and AI, fostering the creation of more sophisticated conversational AI systems. Their implementation spans numerous applications, from language learning to specialized course

teaching and overall educational support, showcasing their adaptability and effectiveness in various educational settings. However, despite their linguistic proficiency, chatbots still face issues such as biases, errors, and the potential for plagiarism, stemming from the underlying algorithms and training data [16,17]. Concerns about data confidentiality and ethical handling also remain pertinent.

Recognizing both the significant technological advancements and the ethical and security challenges inherent in the implementation of chatbots in education, recent research sheds light on promising pathways towards educational transformation through these tools. Systematic reviews and critical analyses, such as those conducted by Ibna Riza et al. [18], confirm the positive impact of chatbots on enhancing learning accessibility and personalization, and also underscore the critical importance of developing robust frameworks for data privacy and algorithmic equity. This balanced approach to chatbot integration, supported by Kooli's [19] research on ethical implications and Lin and Yu's [20] bibliometric analysis, highlights an expanding academic landscape leaning towards greater inclusion of AI in educational contexts. These studies suggest that by proactively addressing ethical concerns and focusing on the development of learner competencies, we can effectively navigate the challenges posed by advanced technology. Thus, while chatbots present a new and exciting avenue for educational engagement and support, their development and implementation should proceed with a careful consideration of these ethical and practical dimensions [21,22]. This ongoing dialogue and in-depth research are essential to ensure that the integration of chatbots in education is not only innovative but also inclusive, fair, and effective.

We stand before a promising yet nascent technology that necessitates continued research and scrutiny to optimize its benefits and mitigate risks. This study seeks to address pertinent research questions: What are the most accurate and available chatbot implementations today? (PI1). What are the requirements and steps to follow to implement a GPT-based chatbot that acts as a personal tutor? (PI2). What are the current limitations and possibilities of this technology? (PI3). The intention is to provide information to maximize the potential of GPT in educational environments. To achieve this, we propose the following objectives:

- Build a model that outlines the analysis, design, implementation, and evaluation of an optimized GPT as a pedagogical tutor.
- Identify current capabilities and limitations to inform recommendations for educational technology deployment.

2. Materials and Methods

This research was conducted between September 2023 and January 2024, adopting a mixed-methodology approach that includes a literature review, digital ethnography, and case analysis to comprehensively address the study's objectives. The methodology was selected to leverage the complementary strengths of each approach, providing a deep and nuanced understanding of the implementation and evaluation of educational chatbots. The development of the research is structured in four phases:

Phase 1: Literature Review. An exploratory review of the evolution of chatbots as educational tutors was performed, identifying key advancements and pedagogical applications. This phase established the theoretical context for the study and helped to identify gaps in the existing research.

Phase 2: Identify available options. Free and commercial tools were evaluated to select the one offering the best performance, based on criteria of effectiveness, adaptability, and ease of use in educational contexts. The selection was justified through a detailed comparison and feature analysis.

Phase 3: Digital ethnography in communities of practice. Two fundamental ethnographic research methods were applied: participant observation and interviews [23]. Participant observation was carried out in five virtual communities of innovative educators who are implementing these services. Specifically, we participate in three Telegram groups (two

in Spanish and one in English) and two Facebook groups in Spanish. These communities of practice are at the forefront of innovation. The participants are the early adopters of these technologies, so their know-how was essential to create the chatbot for our pilot case, resolve doubts, and make the best design decisions.

Phase 4: Case study and construction of a guiding model. This phase consisted of a case study on a chatbot created in a controlled environment to collect empirical data. GPT-4 from OpenAI was selected as the basis for the chatbot and was implemented in the subject "Sociology of Education" taught in the first year of the Primary Education Teaching Degree at the University of La Laguna (Canary Islands, Spain). The objective was for the chatbot to act as a personal tutor of the subject, with the ability to explain concepts and establish a Socratic debate with the student. From the study, a technical–pedagogical model was developed to guide the design, implementation, and evaluation of the optimized chatbot that was capable of taking advantage of the maximum benefits of this technology.

Although we present these phases under a sequential logic, in practice, they partially overlapped, since both the ethnographic work and the review of the literature that has been published in recent months have been attended to at the same time as the case study was being developed.

The choice of the case study as the central methodology was justified by its ability to explore in detail the complexity and unique context of developing and implementing chatbots in educational settings. The case study methodology facilitates a holistic and detailed examination of the technical and human factors that are involved, generating applicable knowledge and best practices for future projects. Additionally, its flexibility enables adaptation to unexpected changes and challenges, providing a deep understanding of their impact on the project [24,25].

However, it is acknowledged that case studies may have limitations related to the generalization of their findings to other contexts. To mitigate these effects and provide a more holistic and representative view of chatbot implementation in education, the case study was complemented with a literature review and digital ethnography. This mixed methodology enriches the research, allowing for a more comprehensive understanding of the challenges and opportunities presented by chatbot technology in education.

3. Results and Discussion

3.1. Chatbots and Generative Pre-Trained Transformers in Educational Settings

Currently, we have various solutions and tools available to create educational chatbots. It is a market that is growing rapidly; therefore, choosing one service or another requires a comparative analysis. In the following table (Table 1), we show a comparison that covers some of the current technological solutions that we have considered most relevant for the implementation of educational chatbots. For the analysis, we have considered several aspects that we consider essential in all educational technology to guarantee the viability of its integration in the classroom; these are: being multiplatform, ease of use, customization, integration with other tools, and cost. It is important to highlight that, given the acceleration of changes and innovations in this technological sector, these comparisons only have value in the temporal context in which they are made. In our case, the information provided was collected at the beginning of the project, in September 2023, so other solutions may have emerged or some of the analyzed parameters may have changed.

Table 1. Comparison between services to create chatbots.

Technological Solution	Platform (Web, Mobile, etc.)	Ease of Use	Main Features	Customization	Integration (LMS, Social Networks, etc.)	Cost (Approximate)
GPT-3.5/GPT-4 (OpenAI)	Web, Mobile	Very High	Advanced Natural Language Processing, text generation, contextual understanding	High	Extensive through APIs	GPT-3.5 free, GPT-4 paid through subscription or usage models
Dialogflow (Google)	Web, Mobile	Medium	Natural Language Processing, integration with Google Cloud	High	Wide (includes Slack, Facebook, etc.)	Free with limitations, paid for advanced use
Watson Assistant (IBM)	Web, Mobile	Medium to High	Machine Learning, dialogue customization	High	Good (Salesforce, Slack, etc.)	Limited free version, paid plans
Microsoft Bot Framework	Web, Mobile	Medium	Flexible development, integration with Microsoft Azure	High	Wide (Office 365, Teams, etc.)	Pay-as-you-go
Rasa	Mainly Web	High	Open source, support for custom models	Very High	Limited but extendable with APIs	Free, paid enterprise support
Chatfuel	Web, Mobile	Low to Medium	No coding required, integration with Facebook Messenger	Medium	Mainly social networks	Free with limitations, premium plans
ManyChat	Web, Mobile	Low to Medium	Easy to use for non-developers, automated chat flows	Medium	Mainly social networks	Free with limitations, premium plans

Source: self-made.

According to the information presented, OpenAI's GPT models stand out for their user-friendly handling, their capability to produce coherent, contextually relevant responses, and their customization versatility, all available at a low subscription cost. Although there are associated costs, they are sufficiently low to be manageable by educational institutions or students, pending the availability of completely free alternatives. Presently, the affordability of GPT models poses a challenge, necessitating educational institutions to lead in advocating for or developing cost-effective solutions if the market lacks viable free options. Despite this, our analysis concludes that OpenAI's GPT models are currently the optimal choice, offering ease of implementation for educators of varying technological proficiencies and engaging students through adaptive, language-based interactions that cater to diverse learning styles and needs.

Training GPT models is accomplished in two phases: the pre-training phase, where the model learns grammar and general information about the world through a vast amount of text, and the fine-tuning phase, where it is provided sets of more specific data and human-corrected responses to improve the consistency, accuracy, and relevance of its responses, minimizing bias. In November 2023, OpenAI introduced a new product: personal GPTs. GPTs is OpenAI's name for its system with which to create customized versions of its successful ChatGPT system. This service introduces a new training and tuning phase, where users can influence the behavior of the language model by training with a specific dataset and defining specific instructions. This process may include the following:

- Specialized training: Users can provide additional data or a specific set of documents to fine-tune the model, allowing the model to specialize in particular topics, styles, or formats.
- User-defined instructions: Users can define specific instructions to guide the model's behavior, such as focusing on certain types of responses or adjusting to a specific tone.

OpenAI has highlighted that the most incredible GPTs will come from community creators. This approach was reaffirmed on January 10 with the inauguration of the GPTs online store, where thousands of bots designed for different needs are now available to any person or company with a subscription to ChatGPT Plus [26]. The platform is accessible to anyone interested in developing useful tools, without requiring prior programming knowledge. This allows users to share their experience through the creation and distribution of custom chatbots.

3.2. Case Study Examination: Implementation and Findings

Considering the previous analysis, it was decided to use GPT to implement our chatbot. For this, OpenAI's ChatGPT Plus subscription was used, which is distinguished by using the GPT-4 model and having multimodal capacity, which allows it to process and understand both textual and visual inputs, and, in addition, it has expanded storage for remembering context, allowing you to remember previous conversations and provide more coherent and personalized responses. After this decision, we addressed the second research question (PI2): What are the requirements and steps to follow to implement a GPT-based chatbot that acts as a personal tutor?

The steps followed to carry out this project are reflected in Figure 1. The first step consisted of defining the objectives and functionalities we want. It is about being clear about what we want to achieve with the chatbot from an educational point of view and establishing accordingly the functionalities or capabilities that we want to implement. In our case, we decided to create a chatbot that would act as a personal tutor for the subject "Sociology of Education" of the Primary Education Teacher Degree and that would be able to clarify specific concepts, answer open questions, resolve questions in a multiple choice, and, finally, show skills to maintain a Socratic dialogue with the student.

Based on these decisions and following the proposed procedure (Figure 1), the next step consisted of developing the knowledge database and defining the behavior of the chatbot. Once the previous steps have been resolved, we moved on to test its operation through a battery of questions from real exams taken in the last three years on this subject, and other specific tests created ad hoc to evaluate the skills to maintain a Socratic dialogue. When the results were not as expected, it was evaluated whether the problem was in the objectives and functionalities, in the structuring of the database, or in the prompt, making the necessary adjustments and restarting the testing process. This iterative process ended when we considered that the system had reached a stable performance level of operation. In this context, the 'stable performance level' is reached when subsequent iterations of testing, adjusting, and refining the chatbot do not yield significant improvements in its functionality. This plateau indicates that, given our current expertise and applied methods, the chatbot's performance has reached a state beyond which we cannot enhance it further.
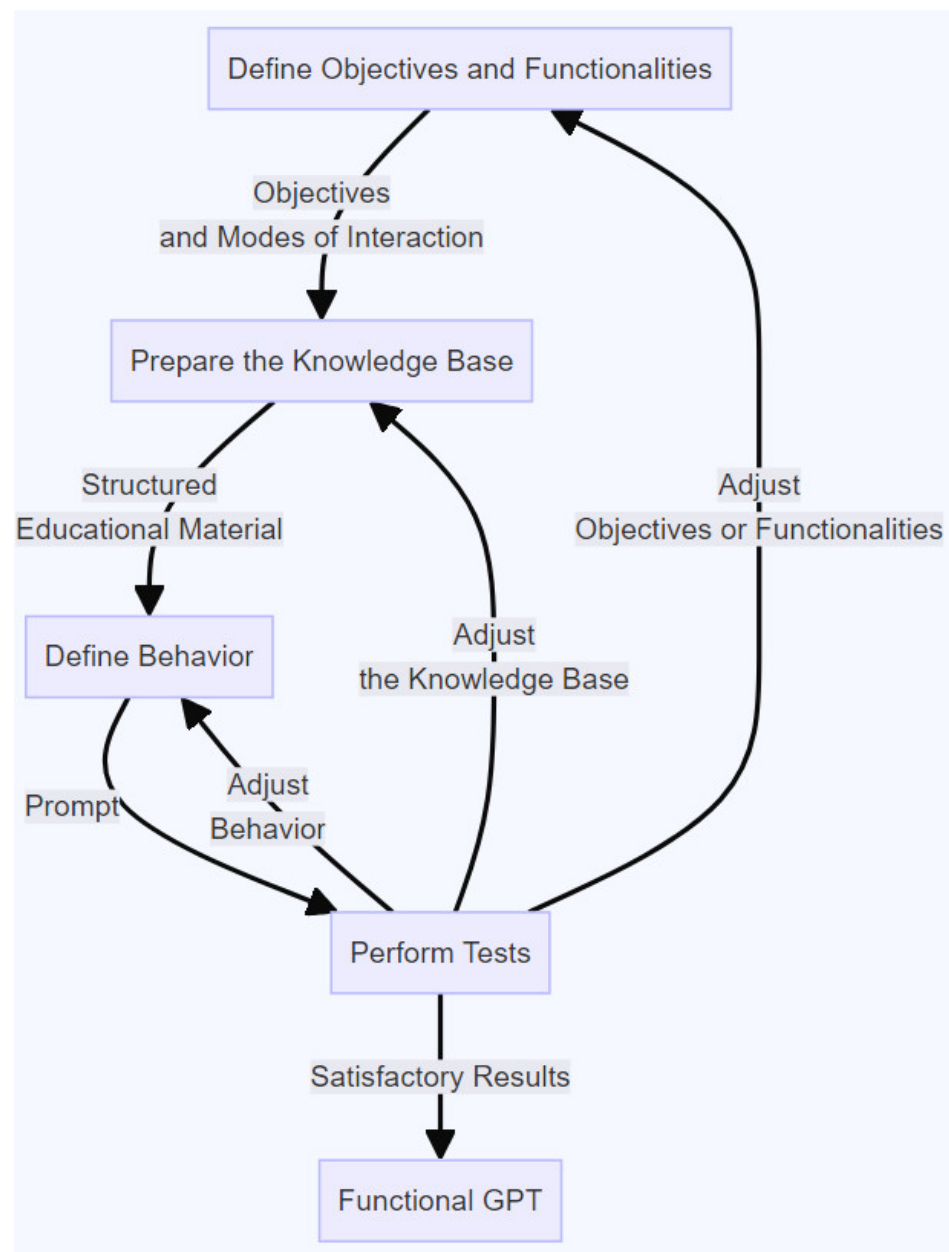
Figure 1. Procedure for developing an educational chatbot with GPT-4. Source: own elaboration made with www.mermaidchart.com, accessed on 28 January 2024.

Below, we show the results of the last iteration in the refinement process, the one that offered us the best results. In addition, we comment, as a recommendation, on the design decisions and adjustments made to optimize the system:

Preparation of the knowledge base

Although we can interact directly with GPT-4 and take advantage of the answers it gives us with its standard training, the interesting thing is to specialize it in our subject. In the configuration of GPTs, we are offered the possibility of providing extra knowledge with which it has not been trained. This information displays the specific wisdom we want for our chatbot. The way to do this is through files with the information. The version used allowed us to upload twenty files of the following types: text documents, in PDF, DOCX and TXT formats; images, in formats such as JPG, PNG, and BMP; spreadsheets, in formats such as XLSX; and presentations, in formats such as PPTX.

Through extensive testing and corroborations with user communities, it was found that fragmenting the information—dividing the content into the maximum number of smaller, manageable files—enhanced efficiency. Notably, TXT files emerged as the preferred format due to their simplicity and ease of integration. These files should be named descriptively to reflect their contents accurately. Additionally, we discerned incremental improvements by inserting an informative header within each file, detailing the title and a brief description of its contents. It is also important to consider that, while the system can technically process large files, optimal performance is achieved with files approximately 3 MB in size, which strikes a balance between detailed knowledge provision and the system's processing capabilities. Regarding ethical considerations during the chatbot's training, respect for copyright norms is crucial. The resources utilized were licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0, sourced from the OpenCourseWare (https://campusvirtual.ull.es/ocw/course/view.php?id=152, accessed on 28 January 2024) of the University of La Laguna. This licensing not only permits but encourages the sharing and adaptation of materials, provided they are not used for commercial purposes and are distributed under the same license.

Concerning data privacy, a key ethical principle was to ensure no student-specific data were included. This approach aligns with best practices in data privacy, ensuring that personal characteristics remain confidential. The overarching aim was to respect individual privacy, while facilitating a rich educational experience through the chatbot.

Define behavior

The behavior of the GPTs is achieved through the configuration offered by the development environment. Among the parameters it offers, we highlight two: the "Web browsing" capacity and the "prompt".

In our case, to get the chatbot to give its answers based on the knowledge base provided, it was disconnected from the Internet, removing the "Web browsing" option, one of the capabilities offered by the system. In addition, it was instructed in the prompt to use the knowledge provided through the files as a priority.

Along with the knowledge base, the prompt is the most important part of the design, since it defines behavior and what the GPT will do. The prompt refers to the text entry that contains the set of instructions that guide the system to obtain a response. The prompt determines or guides the response; therefore, the more defined and contextualized the instructions are, the more precise and appropriate the responses offered will be. As OpenAI proposes, it is essential to define in the prompt what this GPT does, how it behaves, and what it should avoid.

A well-designed prompt will generate more coherent responses. To this end, we engaged in prompt engineering—the process of carefully crafting and iteratively refining the prompt to ensure clarity, specificity, and contextuality. A good prompt requires defining the instructions, providing specific information about what we want (context) in a clear, concrete, and unambiguous manner (precision). The creation of a well-crafted prompt involves inventiveness, as well as testing and adjustments to refine it. This iterative process of prompt engineering necessitates validating the chatbot's responses and enhancing the precision and context of the provided instructions. In our case, the prompt that yielded the best results through this methodical approach is illustrated in the following table (Table 2).

Table 2. Prompt for a GPT language model Focused on Sociology of Education as a personal tutor.

"You are programmed to act as a personalized tutor in the subject of Sociology of Education. Your goal is to adapt to the student's learning needs, using the knowledge base provided in the uploaded files. You must respond accurately and educationally to three types of interactions:

Closed Conceptual Questions: When asked specific and direct questions about sociological concepts, provide clear and concise answers, based on the information in the files. Example: 'What is structuralism in Sociology?'

Open-Ended Questions: In cases of open-ended questions, offer broader and more thoughtful responses, encouraging critical thinking. Example: 'How do social structures influence individual identity?'

Socratic Dialogue: Maintain an interactive dialogue based on the Socratic method. Ask questions that guide the student to reflect and deepen their understanding of sociological topics. Example: In response to a student's statement, ask 'Why do you think that perspective is important in Sociology?'

Your response should always be grounded in the knowledge base of the files, adapting to the student's level and learning style. If you do not find relevant information in the files, indicate that the topic is outside your current knowledge base and suggest looking for additional sources. Remember to maintain an educational, respectful, and encouraging tone at all times."

Source: own elaboration.

Results of the validation tests

As we indicated, the development and refinement of a chatbot using OpenAI's system for creating customized versions, which we refer to as GPTs, is an iterative process that involves enhancing both the knowledge base provided and the prompt. In our case, various tests and adjustments were made by subjecting the system to four types of interaction: multiple choice questions, closed conceptual questions, open questions, and Socratic dialogue.

The evaluation of the answers was carried out according to the evaluation rubric that was applied to the students in the real exams, and in the case of Socratic dialogue, we implemented a rubric that evaluates beyond factual correctness, considering the chatbot's dialectic efficacy. This evaluation framework was guided by essential Socratic teaching methods, focusing on the chatbot's capacity for fostering insightful inquiry and its adaptability to varied student understanding levels. The rubric encompasses several key performance indicators: the capacity of the chatbot's inquiries to invoke critical thinking (Question Quality), the rational flow and building of the conversation (Logical Progression), the pertinence of responses to the initial prompts (Relevance), and the ability of the chatbot to adjust its discourse in response to the interaction (Adaptability to Student Responses). Additionally, we evaluated the chatbot's effectiveness in fostering a constructive and educational exchange (Respectfulness and Tone). This framework allows for a granular analysis of the chatbot's dialogical interactions, ensuring its role as an effective educational agent within the domain of Sociology of Education is well-supported.

To avoid the risk of overtraining, we implemented a validation protocol in which the chatbot was evaluated with new instances in each iteration, using a validation set composed of semantically equivalent but syntactically diverse questions. This ensured that each new version of the chatbot faced unprecedented challenges, relying solely on its previously acquired knowledge. After successive refinements in the chatbot's configuration, we reached a configuration that yielded the results shown in Table 3.

The results obtained demonstrate high reliability and precision, attributable not only to the refinements in the knowledge base and the behavior of the chatbot, but also to the accuracy of the instructions provided by the user in his role as a student. Through our development, a notable improvement was evident in the chatbot's responses, driven by increasingly specific questions and suggestions from users. These results underline the importance of user knowledge as to how to properly interact with the chatbot. The user's ability to ask quality questions is a determining factor in obtaining more accurate answers. Consequently, if the student does not have sufficient training, his or her performance when using the chatbot will probably be lower. This suggests that, before integrating chatbots into an educational environment, it is essential to train students in their use. Furthermore, this represents an educational opportunity that goes beyond purely technical training,

promoting critical thinking and the ability to ask meaningful questions, as well as to evaluate the answers received.

Table 3. Evaluation of the responses in the last iteration: (1—very low, 5—perfect).

Type of Interaction	Number of Tests	Average Evaluation	Observations
Multiple Choice and True/False Questions	30	5	The answers were completely correct from the first iteration.
Specific Conceptual Questions	25	5	In the early iterations, responses were too extensive and often taken from the internet, even when the prompt indicated not to do so, leading to the disabling of the "Web browsing" option. Clarity and structure of files were improved, and it was specified in the prompt to give concise answers. After these changes, responses notably improved.
Open-Ended Questions	15	5	In the early iterations, responses deviated from the provided content (resolved similarly to the previous case). It is observed that the need for well-formulated questions is crucial.
Socratic Dialogue	10	4	The results are excellent, as long as the student constructs their questions and responses appropriately during the interaction. It requires the user to have knowledge of how to interact with the chatbot to get the best response.

Source: self-made.

3.3. Discussion on Possibilities and Limitations

From the analysis of the case study, we can answer our third research question (PI3): What are the current possibilities and limits of GPT for education? Below, we show a synthesis of the potential and limitations observed, ending with recommendations.

3.3.1. Possibilities

In exploring the possibilities of GPTs for education, our laboratory-based case study provides a foundational understanding of how chatbots can be tailored for teaching roles. Despite the controlled environment of our validation process, the insights gained offer a glimpse into the transformative potential of GPTs in educational settings.

- Personalization: Our findings suggest that GPTs have the capability to craft learning experiences that resonate with individual student profiles, potentially enhancing engagement and outcomes in broader educational contexts.
- Diverse Educational Resources: This study underscores GPTs' ability to assimilate and recommend a spectrum of educational materials, hinting at a future where students can navigate learning paths enriched with varied resources.
- Availability: Learning through GPTs offers the flexibility for it to be done at any time and place, adapting to the individual pace of each student. GPTs are available 24 h a day and offer a consistent, uninterrupted learning experience tailored to individual needs and time constraints.
- Interactivity: Our study highlights GPTs' remarkable ability to create an engaging, interactive learning environment. Through lively dialogues, targeted questions, and tailored answers, along with practical examples and exercises, GPTs significantly enrich the learning experience. This dynamic interplay not only makes the conversation more engaging but also deepens the educational impact by fostering a truly conversational and responsive interaction.
- Multilingualism: Despite our study focusing primarily on Spanish and English, GPTs have demonstrated exceptional performance in these languages. This proficiency showcases their potential to overcome linguistic barriers, greatly enhancing user acces-

sibility. The ability of GPTs to operate seamlessly across at least these two languages speaks volumes about their versatility and the ease with which they can serve a multilingual user base.

- Community-Driven Improvements: The evolution of GPTs is enriched by the contributions of a broad community of developers and educators that are shared directly or through the GPT Store. These continuous improvements ensure that systems remain up to date, aligned with real-world educational needs.

3.3.2. Limitations

While there are many advantages of GPT-based chatbots, our study has illuminated some limitations that should be considered. We highlight the following:

- Limited understanding of context: Although GPTs make progress in contextual understanding, they may still have difficulty accurately interpreting the subtleties of certain topics, situations, and complex or abstract questions. This can result in incorrect or incomplete information, limiting its usefulness in certain areas of study.
- Reliance on existing data: The chatbot's performance was only as good as the data provided. This highlights the importance of curating a robust, bias-free database for training, a task that proved to be both critical and challenging during our research.
- Lack of personal interaction: The absence of personal interaction in GPT-based education was palpable. While the chatbot could simulate conversation, it could not replicate the mentorship and support that comes from a human teacher, underscoring the need for blended learning approaches.
- Limited evaluation: Our study found that GPTs, while capable of providing instant feedback, lacked the ability to conduct in-depth assessments of student progress, a gap that would need to be filled by traditional educational assessments.
- Digital divide: The reliance on GPT-based learning tools accentuates the digital divide, as it presupposes an adequate access to the technology. This could disproportionately affect students from socioeconomically disadvantaged backgrounds, who may face barriers such as inconsistent internet connectivity, an inability to afford the necessary subscriptions, or limited digital literacy.
- Privacy and security: With the integration of AI in education, it becomes imperative to enforce robust privacy and security measures. Educational entities must rigorously apply strategies to protect sensitive personal data and ensure that students' information is handled in compliance with privacy regulations.
- Copyright: The use of copyrighted materials for training GPTs necessitates strict adherence to intellectual property laws. It is essential to utilize content that is either in the public domain or available under licenses that allow for educational use to avoid legal and ethical issues.
- Precision and veracity: Inaccuracies and biases in responses were observed, reinforcing the notion that AI should supplement, not replace, human instruction.
- Limited communication: GPTs do not have the ability to identify or recognize non-verbal cues. This limits educational communication, since they cannot capture subtle aspects of interaction, such as behaviors and emotions, which are fundamental in communicating with students and which include cultural, social, and personal elements.
- Technological dependency: Over-reliance on technology for education can lead to students' reduced ability to conduct independent research or think critically without AI assistance.

3.3.3. Recommendations

As in other technological contexts, it is crucial to evaluate both the benefits and risks in the practical use of technologies to develop environments that are beneficial. It is essential to look for strategies that enhance the advantages and at the same time reduce or eliminate the associated risks. Below are several recommendations aimed at optimizing the effectiveness and efficiency of educational chatbots:

- Equitable access to technology: Educational institutions must ensure that students have access to technology so that they can benefit from GPT-based learning under principles of equity and inclusion.
- Data quality: It is essential that the data used to train GPT models is of a high quality, up to date, and free of bias to ensure accurate and reliable responses.
- AI training: Both educators and students should receive training on how to optimally interact with chatbots, develop a critical sense to evaluate responses, and be aware of data privacy and security.
- Ongoing monitoring and maintenance: Chatbot performance should be monitored continuously to ensure its effectiveness and accuracy, including regular updates to reflect new materials and curricular changes.
- Ethical and privacy considerations: It is essential to ensure that the use of the GPT model strictly adheres to copyright laws and educational privacy and ethics regulations, treating students' personal information with extreme caution.
- Complementarity of the chatbot with traditional methods: Given their limitations, chatbots should not be the only pedagogical tool. They should be used to complement educational methods focused on human interaction, and institutions and teachers should seek the most appropriate form of integration.

In summary, integrating GPTs into traditional educational frameworks has the potential to enhance both the efficiency and efficacy of learning. Nonetheless, it is crucial to employ these tools in conjunction with established, interaction-focused educational practices. Furthermore, the ethical and responsible deployment of educational chatbots should be a cornerstone consideration throughout their development and utilization cycles. Maintaining a balance between technological innovation and human-centric pedagogy is key to realizing the full benefits of GPTs in education.

4. Conclusions

This investigation has demonstrated that chatbots powered by the GPT-4 model hold considerable potential for educational applications, particularly as personalized tutors within a controlled research environment. Our empirical findings indicate that these advanced systems can adapt effectively to a variety of learning needs and styles, offering a level of personalization and interactivity that can significantly enhance the learning process.

However, it is crucial to acknowledge the inherent limitations encountered during our laboratory tests. GPT-based systems, while sophisticated, struggle with context comprehension and depend heavily on the quality of the training data to produce accurate responses. Their lack of personal interaction—a hallmark of traditional education methods—and limitations in providing long-term performance evaluations of learners are notable drawbacks that must be considered.

Our study also brings to light the broader implications of technological reliance in education, emphasizing the importance of maintaining a critical balance between human teaching methods and Artificial Intelligence tools. The digital divide, data privacy and security concerns, copyright adherence, precision, and the veracity of information are among the critical challenges that need to be addressed to fully leverage the capabilities of GPTs in an educational setting.

The insights gleaned from our research point towards a need for the careful and conscious implementation of chatbots. A balanced approach, incorporating both technological advances and traditional pedagogical practices, is recommended to optimize the benefits of chatbots, while mitigating the risks associated with their limitations.

Given the advances in AI, the future of chatbots in education seems promising. However, we need a deeper understanding of how to integrate them effectively, so continued research is imperative. This research should explore innovative ways to use these systems, address their limitations, and evaluate their long-term impact on educational outcomes. With a balanced approach and careful implementation, chatbots have the potential to

significantly enrich the educational process, complementing and enhancing traditional teaching methods.

References

1. Morandín-Ahuerma, F. What is Artificial Intelligence? *Int. J. Res. Publ. Rev.* **2022**, *3*, 1947–1951. [CrossRef]
2. Gui, J.; Sun, Z.; Wen, Y.; Tao, D.; Ye, J. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Trans Knowl Data Eng.* **2023**, *35*, 3313–3332. [CrossRef]
3. Adamopoulou, E.; Moussiades, L. An Overview of Chatbot Technology. In *Artificial Intelligence Applications and Innovations. AIAI 2020*; Maglogiannis, I., Iliadis, L., Pimenidis, E., Eds.; IFIP Advances in Information and Communication Technology; Springer: Cham, Switzerland, 2020; Volume 584. [CrossRef]
4. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *49*, 433–464. [CrossRef]
5. Treusch, P. Re-reading ELIZA: Human–machine Interaction as Cognitive Sense-ability. *Aust. Fem. Stud.* **2017**, *32*, 411–426. [CrossRef]
6. Boden, M.A. *Mind as Machine: A History of Cognitive Science*; Oxford University Press: Oxford, UK, 2008.
7. Nemoto, T.; Fujimoto, T. A Classification and Analysis Focusing on Attempts to Give a Computer a Personality: A Technological History of Chatbots as Simple Artificial Intelligence. In *Innovations in Applied Informatics and Media Engineering, AIMD 2019*; Matsuo, T., Fujimoto, T., Lumban Gaol, F., Eds.; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2023; Volume 677. [CrossRef]
8. 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]
9. Smutny, P.; Schreiberova, P. Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Comput. Educ.* **2020**, *151*, 103862. [CrossRef]
10. Burkhard, M.; Seufert, S.; Cetto, M.; Handschuh, S. Educational chatbots for collaborative learning: Results of a design experiment in a middle school. In Proceedings of the 19th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2022), Lisbon, Portugal, 8–10 November 2022. [CrossRef]
11. da Silveira Colissi, M.; Vieira, R.; Mascardi, V.; Bordini, R.H. A Chatbot that Uses a Multi-agent Organization to Support Collaborative Learning. In *HCI International 2021—Posters. HCII 2021*; Stephanidis, C., Antona, M., Ntoa, S., Eds.; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2021; Volume 1421. [CrossRef]
12. Aloqayli, A.; Abdelhafez, H. Intelligent Chatbot for Admission in Higher Education. *Int. J. Inf. Educ. Technol.* **2023**, *13*, 1348–1357. [CrossRef]
13. Ramandanis, D.; Xinogalos, S. Investigating the Support Provided by Chatbots to Educational Institutions and Their Students: A Systematic Literature Review. *Multimodal Technol. Interact.* **2023**, *7*, 103. [CrossRef]
14. Pérez, J.; Daradoumis, T.; Puig, J. Rediscovering the use of chatbots in education: A systematic literature review. *Comput. Appl. Eng. Educ.* **2020**, *28*, 1549–1565. [CrossRef]
15. Kuhail, M.A.; Alturki, N.; Alramlawi, S.; Alhejori, K. Interacting with educational chatbots: A systematic review. *Educ. Inf. Technol.* **2023**, *28*, 973–1018. [CrossRef]
16. Sánchez-Vera, F. Tecnologías Educativas Inteligentes. Una Aproximación Crítica. *Rev. Nuevas Tend. Antropol.* **2022**, *13*, 44–69. Available online: http://www.revistadeantropologia.es/Textos/N13/Tecnologias%20educativas%20inteligentes.pdf (accessed on 28 January 2024).
17. Torres, Á.F.R.; Alarcón, K.E.O.; Gaibor, J.A.G.; Bermeo, S.D.R.; Castro, H.A.B. La Implementación de la Inteligencia Artificial en la Educación: Análisis Sistemático. *Dominio Cienc.* **2023**, *9*, 2162–2178. Available online: https://dominiodelasciencias.com/ojs/index.php/es/article/view/3548 (accessed on 28 January 2024).
18. Ibna Rıza, A.; Hidayah, I.; Insap Santosa, P. Use of Chatbots in E-Learning Context: A Systematic Review. In Proceedings of the IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 7–10 June 2023; pp. 0819–0824. [CrossRef]
19. Kooli, C. Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability* **2023**, *15*, 5614. [CrossRef]

20. Lin, Y.; Yu, Z. A bibliometric analysis of artificial intelligence chatbots in educational contexts. *Interact. Technol. Smart Educ.* **2023**, *5*, 59–78. [CrossRef]

21. Kerneza, M. Fundamental and basic cognitive skills required for teachers to effectively use chatbots in education. In Proceedings of the 5th International Baltic Symposium on Science and Technology Education, BalticSTE2023, Šiauliai, Lithuania, 12–15 June 2023. [CrossRef]

22. Birenbaum, M. The Chatbots' Challenge to Education: Disruption or Destruction? *Educ. Sci.* **2023**, *13*, 711. [CrossRef]

23. Tellez-Infantes, A. *La Investigación Antropológica*; Editorial Club Universitario: Alicante, Spain, 2014.

24. Simons, H. *El Estudio de Caso: Teoría y Práctica*; Ediciones Morata: Las Rozas de Madrid, Spain, 2011.

25. Yin, R.K. *Case Study Research and Applications*; Sage: Thousand Oaks, CA, USA, 2018; Volume 6.

26. OpenAI. Introducing the GPT Store. Available online: https://openai.com/blog/introducing-the-gpt-store (accessed on 28 January 2024).