International Journal on Cybernetics & Informatics (IJCI) Vol.13, No.4, August 2024

AI Teachers Using a VR/MR Environment for Greater Student Interaction and Immersion

L.M. Escobedo F.

School of Engineering, InterNaciones University,
Guatemala City, Guatemala, June 2024

Abstract. The transmission of knowledge has been a fundamental tool throughout the history of humankind, where each generation imparted and improved the knowledge and skills of the previous one. However, at certain times in history, some individuals attempted to monopolize this knowledge to gain advantages over their peers, thus restricting access to knowledge. As societies grew and diversified, it became evident that collaboration between individuals from different backgrounds and cultures accelerated and enhanced the expansion of knowledge. This process was further enhanced with the arrival of technological revolutions, such as the internet. Currently, artificial intelligence represents a new technological revolution that, if used properly, can produce better prepared individuals equipped with the best tools in history of humanity to expand and improve the knowledge acquired.

Keywords: Education, artificial intelligence, virtual reality, mixed reality.

1 Introduction

One of the most important tools throughout human history is the transmission of knowledge. Historically, one generation would impart its knowledge and skills to the next, and the next generation would improve and expand on that knowledge, thus repeating the cycle. However, some individuals realized that monopolizing this knowledge gave them an advantage over the rest, so they began to reduce the circle of apprentices.

This continued until societies grew exponentially and it became clear that the more individuals from different backgrounds and cultures participated in the expansion of knowledge, the better results could be achieved more quickly and accurately. This phenomenon was intensified by the global connectivity provided by technological revolutions such as the Internet.

With artificial intelligence, we are facing a new technological revolution. If we know how to use it properly, we will be able to create better-prepared individuals with the best tools in history of humanity to expand and improve the knowledge they acquire.

1.1 Shortage of Educators in Guatemala

In Guatemala, there is a serious shortage of teaching staff in schools and other public institutions. For example, there are teachers who have to teach two grades at the same time in the same classroom. A single teacher may have third-grade students and fourth-grade students in the same classroom. In addition, there are cases in which a teacher has up to 60 or more students per classroom.

This shortage of primary and secondary school teachers is most evident in the towns and villages in the interior of the country. The State offers better salaries and conditions to teachers who agree to teach in these localities, especially in the more remote areas.

Some private institutions, such as schools and academies, often hire unqualified personnel as educators, asking only that they have some knowledge of the subject they will teach.

Bibhu Dash et al: CEEE - 2024
pp. 193-201, 2024. IJCI – 2024
DOI:10.5121/ijci.2024.130417

This shortage is not limited to schools alone. At the *University of San Carlos of Guatemala (USAC)*, the National University of Guatemala, there are professors with more than 120 students assigned to each course. This occurs mainly in the first-year courses of the different university courses. It is impossible for a single professor to offer quality education or adequate assistance to so many students. For these reasons, many students do not have the opportunity to ask their professors questions, or even to understand the lesson if they are unlucky enough to find a place only in the seats located at the back of the room.

2 Construction of the System

2.1 VR/MR Technologies

Virtual Reality (VR) is a technology that allows us to enter a virtual world created by a computer using a virtual reality device, such as the Meta Quest 2 or the PICO 4 glasses. This virtual world can contain anything that its creator wants, from a Jurassic world before the extinction of dinosaurs to a classroom like any other. But... What does VR technology have to do with education or AI?

One of the main challenges in remote teaching is capturing the student's attention, and interacting with someone through a screen is not the same as interacting with someone face to face. Not to mention that VR technologies can provide us with a host of tools that educators can use. For example, creating a virtual space where students can learn about the history of pre-colonial Mesoamerica by visiting the reconstructed ancient cities, now in ruins; or learning about geography by seeing a relief map of their country with indicators of all the lakes, rivers and various points of interest for the course.

Another advantage that can be taken advantage of from these technologies is that students can interact with each other, even if they are at great distances, as if they were all in the same place. The importance of communication and social interaction for students was reflected in the confinement due to the COVID 19 pandemic, especially in children in their early years of development. At this stage, they develop social skills and the formation of bonds that will be essential to them throughout their lives [4].

But what if the institution is not interested in remote education? That's where Mixed Reality (MR) technologies come in. MR technologies are a fusion between VR and AR (Augmented Reality) technologies. These technologies try to add a layer of virtuality to the real world and make this layer of virtuality adapt and interact with the real world.

In a classroom, students could use MR devices such as the Meta Quest 3 or Apple Vision Pro glasses. With these devices, students could see how their classroom environment combines with the virtualization layer created by the developer.

And where do Artificial Intelligence technologies fit into this educational model? AI technologies will help us create the adaptable, scalable and almost perfect teacher or tutor for each student. For this purpose, we would first need a combination of several artificial intelligence algorithms.

2.2 NLP

First, in order for the student to be able to interact with the teacher, we would need a Natural Language Processor (NLP). This is a Machine Learning technology that trains computers to interpret and understand human language.

NLP combines computational linguistics, machine learning, and deep learning to understand human language.

Computational Linguistics This science creates language models with computers, using syntactic and semantic analysis. It is used in tools such as language translators, text-to-speech synthesizers and voice recognition software.

Machine Learning This technology trains computers with sample data to improve their efficiency in recognizing features of human language, such as sarcasm and metaphors.

Deep Learning A subfield of machine learning, it teaches computers to recognize complex patterns in data using neural networks that mimic the human brain.

To implement an NLP system, text or speech data must first be collected and prepared. Preprocessing involves tokenization, stemming, lemmatization, and stopword removal. The preprocessed data is then used to train NLP models with machine learning, thereby improving their accuracy. Finally, the model is deployed to predict specific outcomes in a production environment [1].

2.3 Speech-to-text (STT)

In order for the NLP system to interpret what the student says, we first have to translate it into text. To do this, we will use a speech recognition algorithm or *Speech-to-text*.

Speech-to-text algorithms convert audio fragments into word sequences using automatic speech recognition (ASR) technology. This technology facilitates the creation of subtitles and transcripts, improving accessibility and management of multimedia records.

Common examples of this class of algorithms are:

Connectionist temporal classification algorithm (CTC) Train systems to convert audio to text, even if the durations don't match perfectly.

Hidden Markov Models (HMM) These determine the most likely sequence of words based on the sounds in an audio sample.

These algorithms improve the accuracy and responsiveness of speech recognition through extensive training, allowing computers to better handle the nuances of human speech and improve their performance over time.

The use of deep learning techniques significantly improves speech recognition algorithms, making audio-to-text conversion more accurate and efficient. Neural networks, which mimic the structure of the human brain, are popular in this field due to their ability to handle complex variables such as poor audio quality or unusual speech patterns.

Neural networks are applied in several aspects of speech recognition:

Neural networks are applied in several aspects of speech recognition. With *feature extraction*, they break down audio to isolate key components such as pitch and frequency. With *acoustic modeling*, they relate these acoustic features to linguistic elements such as letters and words. In addition, *language modeling* uses these networks to help form coherent phrases and sentences. With *end-to-end speech recognition*, a single network can manage the entire process, simplifying system architecture. Finally, with *speaker recognition*, neural networks identify specific speakers based on their vocal characteristics.

These applications enable computers to better understand human speech and improve their performance on speech recognition tasks [10].

2.4 Text-to-speech (TTS)

Since communication is a means of input and output, we need to translate the NLP response into audio for the student. To do this, we will need an algorithm that will convert the text into spoken language. For this, we use a Text-to-speech type algorithm.

Text-to-Speech (TTS) technology converts text into audible speech, acting as an interface between written and auditory content. The process involves analyzing the text, breaking it down phonetically, and synthesizing it into spoken words.

Artificial Intelligence (AI) is revolutionizing text-to-speech, creating voices that convey the warmth and nuance of human speech. Unlike traditional TTS systems, AI uses advanced techniques, including machine learning algorithms, to continually adapt and improve voice output.

- Función del aprendizaje automático:Machine learning algorithms analyze large volumes of speech data to reproduce the complexities of human speech, such as variations in pitch and rhythm.
- Spectrograms and waveforms: Spectrograms and waveforms provide visual representations of sound frequency and intensity, essential for AI to understand and replicate human speech patterns.
- AI Model Training: AI models are trained on extensive datasets to capture the nuances of speech, including dialects and accents.
- Overcoming speech challenges: AI handles challenges like intonation, stress, and emotions, making TTS more expressive and engaging.

Voice Cloning TTS has advanced significantly, moving from mechanical voices to smooth, lifelike voices thanks to advanced AI models. Companies such as ElevenLabs, Amazon Polly, and Deepgram Aura are leading these developments. Current TTS models generate speech from scratch, mimicking the human brain to produce a more natural sound.

These technologies include multilingual support and allow for the creation of custom voices. Improvements in speech rate and prosody make the synthesized voice clearer and more expressive.

Deep learning enables TTS to capture the nuances of human speech, delivering a more natural experience.

Voice cloning is a significant advancement in TTS technology, allowing for the creation of digital replicas of human voices. It relies on sophisticated algorithms that analyze and reproduce the unique tonal characteristics of a human voice. Using deep learning techniques, the system is able to capture the nuances of the voice, producing results that are strikingly similar to the original voice.

The process involves taking samples of a person's voice and then using TTS engines to generate speech that mimics the original voice patterns.

Potential applications include personalizing digital experiences, creating posthumous messages, and generating authentic content for advertising campaigns.

However, concerns are being raised about the potential misuse of this technology, such as identity theft and unauthorized use of a person's voice. It is therefore essential to establish clear legal and consent frameworks to protect individual vocal identities.

Recent technological advances have made this technology more accessible to the general public, thanks to innovations in machine learning and data processing that have improved the quality of cloned voices.

It is essential to prioritize consent and adapt legal frameworks to address the ethical and security challenges associated with voice cloning [5].

2.5 Learning Curve

The learning curve measures the success of learning over time. It is represented by two axes:

1. Horizontal Axis: Measure effort in units such as days, sessions, or attempts.
2. Vertical Axis(Performance)): Measure success in the activity, such as correct answers, grades, or speed.

To chart the curve, the task must be repetitive and measurable. For example, a runner training to reduce his time in a 15km race over 90 days. The horizontal axis shows the days and the vertical axis the minutes needed to complete the 15km each day.

A downward curve would indicate an improvement in performance. Factors such as motivation, environment and teaching methodology can influence the learning curve [9].

All of these factors could be measured, customized, and optimized individually for each student in the AI education system proposed in this article. If the students performance tests show a lack of understanding of the subject matter, the system could adapt and improve using genetic algorithms [3] and change its approach until it finds the education paradigm —previously trained in the system— suitable for a particular student.

3 Possible Use Cases

Now that we have established the technologies, we only need to ask ourselves: How do we bring all these concepts together and create a functional application for students?

First, we would need to enlist the help of one or more educators who teach the course we want to design, and the assistance of the institution is also important. For practical reasons, a series of hypothetical use cases for this technology will be proposed.

3.1 Use Case 1: History Course

A student in his history course is learning about World War II, so he puts on his VR headset and looks at his AI teacher dressed as a USSR general. The teacher *"flies"* him to Stalingrad and explains the importance of the Battle of Stalingrad, how the Soviets won, and how this battle was decisive in changing the course of the war. All while student-friendly reenactments and documentary material from the period are shown. If the student(s) had a question, they could interrupt the teacher at any time to have it answered.

3.2 Use Case 2: Physics Course

Professor AI invites a VR representation of Isaac Newton into the classroom. He tells them the story of how he came up with his three laws and proceeds to explain them with practical examples, while Professor AI explains the formulas on a virtual whiteboard. The teacher continues the class with theoretical exercises adapted to each student's abilities. These exercises will increase in difficulty as the student improves, based on the learning curve of each individual student.

Thanks to VR/MR technologies, any object could be easily represented in Virtual space, so demonstrating theories with practical examples would be a relatively simple task.

3.3 Use Case 3: Natural Sciences Course

Students are learning about white blood cells, red blood cells, and how the immune system works in general. For this reason, the teacher takes them on a journey inside the human body where students can see how white blood cells defend the body from viruses. Teacher IA then explains the importance of good nutrition, hygiene, and other concepts useful to the students.

3.4 Use Case 4: Painting Course

For this particular case, it is assumed that the student already knows the principles of drawing and only wants to learn how to paint. A painting simulator is created where the teacher assigns a series of exercises for the student to learn the difference between materials, textures. Depending on the student's learning curve, the AI instructor will analyze the student's progress and assign exercises appropriate to their level and ability.

3.5 System Development

As can be seen in the four use cases above, it can be observed that the AI teacher will actually use the NLP algorithms until it is intervened or at the end of a phase of the lesson. Therefore, the course has to be pre-designed, as previously stipulated. For this, we would use education professionals to design the courses, animation and 3D design professionals for the graphic representations.

For the AI teacher's movements, *Mocaps* (Motion Capture) animation prefabs could be created that the system would activate through a Neural Network trained to infer facial movements depending on the output audio using the tonality and words for the *lip sync* as data.

Similarly, for body movements, a Neural Network could be trained to analyze videos of classes with real teacher so that it can infer the intentionality in the movements. This intentionality would be transferred to instructions, where according to the NLP responses, the system would be told which Mocap animation should be activated int the AI teacher at each moment, so that their movements look natural and do no cause *"Uncanny Valley"* in users.

4 Economic Costs

To consider the profitability of such a system, one must consider the costs of equipment and labor against wages and everything involved in hiring a human teacher..

On the one hand, the cost of a pre-trained NLP such as Open AI's GPT-4o would have to be evaluated. Its costs range from $5 per million input tokens to $15 per million output tokens [8]. It must be considered that each student would represent a user and depending on the student, he or she could use more or less interactions with the system..

On the other hand, the *"Open-Assistant"* NLP could be used, which is an already trained algorithm, not as powerful as GPT-4o, but which could be scaled to improve it and, being Open Source, it can be modified to the taste and need of the educational institution [7]. To operate this model, a strong investment in R&D&I would have to be made and have hardware capable of running the algorithm in real time, this would depend on the requirements of the institution that requires it. For this, a Hardware with *LPU (Language Processing Unit)* would be needed, such as the one created by Groq [6]. This is a technology still in development, so there are no official prices yet.

Another option would be to use the NLP created by the Meta company, *"Llama-2 70B"*, which has already been tested on Groq LPUs.

The traditional alternative to LPUs is to use GPU-based technologies, such as computers with Nvidia RTX 3090 graphics cards, priced between approximately $1200 and $1500, or with an Nvidia RTX 4090 card (for greater processing power) that ranges between $1900 and $4000, to which you would have to add the costs of electricity. By purchasing the hardware in bulk, costs could be reduced.

One possible option for small and medium-sized institutions would be to use decentralized distributed LLM networks. These consist of creating networks of computers around the world and *"sharing"* the computational processing costs, like P2P technologies. The Open Source platform *"Petals"* runs on *Llama-2 70B*, *Falcon 40B+* and *BLOOM 176B* for this purpose [2].

For the *"Speech-to-text"* system, there are a variety of audio models that could be used. Such as OpenAI's *Whisper*, which costs $0.006 per minute [8]. This, in the author's personal opinion, would be an unnecessary expense since there are a variety of other locally run *"Speech-to-text"* systems that could be used, or one could create one's own system optimized for the application.

For the *"Text-to-speech"* system, the OpenIA TTS (Text-to-speech) algorithm could be used, with a price of $15 per million characters [8]. And, like the Whisper system, this would also represent an unnecessary waste of money for the same reasons. The alternative would be to directly use Voice Cloning and reach a financial agreement with the person who *"donates"* their voice or look for a free voice on the Internet.

If VR or MR equipment is used, it would be necessary to consider whether the user would be required to purchase their own equipment or if the institution would provide it. In any case, VR/MR equipment such as the *Meta Quest 3* is around $500 and the *Apple Vision Pro* glasses are around $3500. While the *PICO 4* VR glasses range between $390 and $460 depending on the model. The price of the target equipment is an expense to consider both for the development and for the final implementation of the system.

In addition to all this, we must add the salaries of the programmers who create the system, those who train it and those who maintain it, which depending on the final purpose and the size of the institution could be more or less high. In addition, we must also consider the salaries of the designers and animators who create the visual part and the salaries of the educators who create, supervise and approve the various educational content according to each subject.

However, it must be taken into account that these expenses would only occur in the R&D&I stage or in future expansions. Therefore, all human personnel—not including those in charge of system maintenance—would be *"saved"* after the final implementation of the system.

Because of this, we can conclude that spending on human resources would represent a large investment by the institution at the beginning and very little over the years. A single team of programmers and technicians could maintain a system that hosts hundreds of thousands and perhaps millions of users, depending on the size of the institution.

5 Education and Class Struggle

Education and Class Struggle is a book based on the lectures that Aníbal Ponce gave in 1934. The book traces the history of education, from the primitive community to capitalism, showing how education reflects and perpetuates class divisions. In classless societies, education was spontaneous and collective. With the emergence of private property and

social classes, education became an instrument of the dominant classes to oppress the subordinate ones.

According to Ponce, education in capitalism forms individuals who support the system of exploitation. Ponce criticizes the idea that bourgeois education can be reformed to be fair and argues that reforms are only possible after a revolution. He advocates a new socialist education that serves to fully develop individuals and build a classless society [11].

Leaving aside Ponce's Marxist overtones, it is a fact that historically the *"upper"* classes in societies tend to have access to the best education. However, with an education provided by an AI, this inequality of opportunities in education would disappear and this educational revolution, not necessarily socialist, that Ponce spoke of could be achieved.

6 Continue Development

Now that we have both human and material resources, we have to consider what will come after completing the development.

Due to the evolutionary nature of this system, it is necessary to draw up a plan for the future and not limit yourself to a single use case. It is important to think from the beginning about developing a modular system that can be reused in subsequent applications. For this reason, the work methodology that would best adapt would be a combination of agile methodologies, especially prioritizing the SCRUM methodology.

7 Ethical Implications

We must consider all the ethical implications that a technology like this could generate. To do this, we need to ask ourselves a series of questions:

Is it ethical to take away the jobs of millions of teachers for the benefits outlined above?

What if an authoritarian government, current or future, takes advantage of this technology to impose its political and social ideology on its entire population? Since there are no human teachers, AIs would not have the rational capacity and free will to *"deviate"* from the official discourse to plant the seed of doubt in their students and encourage them to think beyond what is stipulated. Even without considering an extreme case such as an autocracy, since AIs do not have abstract thinking, they could not encourage the student to seek, nor accept, solutions that go too far outside the parameters of their training, which could cause stagnation and frustration in the most eager students.

What would happen if AI security is violated and it is deliberately given incorrect data to misinform students? In such a case, would the institution be able to detect the failure quickly and accurately, if it is detected at all?

Would it be right to give AI unsupervised free rein for the earliest education of children? And what about technical knowledge or professional studies?

What good would it do humans to acquire new knowledge and skills if AI could replace us in all areas of life, almost effortlessly? This kind of questioning by students could lead to a stagnation of technology and innovation.

These are just some of the many hypothetical cases that could arise in the use of AI for education. It will be the joint task of governments and societies to consider these and many more scenarios and ethical dilemmas that may arise and set a limit on how far it makes sense to use AI in the educational field.

8 Conclusion

In conclusion, implementing this system requires a substantial initial investment for the R&D&I phase, followed by lower maintenance costs at start-up. Additionally, it is crucial to consider all ethical implications and potential future expansions for various use cases before commencing the project.

References

1. AWS (2024), "¿Qué es el Procesamiento de lenguaje natural (NLP)?". Obtained from AWS: https://aws.amazon.com/es/what-is/nlp/#: :text=tareas%20de%20NLP%3F-,¿Qué%20es%20la%20NLP%3F,y%20comprender%20el%20lenguaje%20humano
2. Borzunov, Alexander; Baranchuk, Dmitry; Dettmers, Tim; Ryabinin, Max; Belkada, Younes; Chumachenko, Artem; Samygin, Pavel; Raffel, Colin. (2022) "Text-to-Speech Models". Obtained from GitHub: https://github.com/bigscience-workshop/petals
3. Caparrini, F. S. (2024). "Algoritmos Genéticos". Obtained from CS US: https://www.cs.us.es/~fsancho/Blog/posts/Algoritmos_Geneticos.md.html
4. CFH. (September 14 2022). "Comunidad escolar: ¿cómo favorece el desarrollo académico?". Obtained from CFH: https://www.cfh.edu.mx/blog/comunidad-escolar-beneficios-academicos#:~:text=La%20comunidad%20ayuda%20a%20estudiantes,que%20favorece%20el%20 rendimiento%20académico
5. Deepgram. (January 9 2024). "Text-to-Speech Models". Obtained from Deepgram: https://deepgram.com/ai-glossary/text-to-speech-models
6. Klinken, E. v. (March 5 2024). "Will Groq depose Nvidia from its AI throne with the LPU?". Obtained from Techzine: https://www.techzine.eu/blogs/infrastructure/117260/will-groq-depose-nvidia-from-its-ai-throne-with-the-lpu/
7. LAION AI. (2024). "Open-Assistant". Obtained from GitHub: https://github.com/LAION-AI/Open-Assistant
8. OpenAI. (2024). "Pricing". Obtained from OpenAI: https://openai.com/api/pricing/
9. Osorio, O. (April 5 2024). "Curva de aprendizaje: Cómo medir y potenciar tu aprendizaje". Obtained from TinyRockets: https://www.tinyrockets.com/blog/curva-de-aprendizaje
10. Verbit Editorial. (2024). "Unveiling the power of speech-to-text algorithms". Obtained from Verbit: https://verbit.ai/ai-technology/speech-to-text-algorithms/#:~:text=Speech-to-Text%20Algorithms%3A%20The%20Basics&text=Essentially%2C%20speech-to-text,speech%20recognition%20technology%20or%20ASR
11. Wanschelbaum, C. (2015). "Educación y lucha de clases". Obtained from SciELO: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982015000300014

Author

L Escobedo received studies in Science and System Engineering from University of San Carlos of Guatemala, and he did Bachelor degree in 3D Production and Animation. Currently, he is pursuing his Master degree in Artificial Intelligence from the InterNaciones University. His research interests include, Mayan Cultures, Extended Reality, and Artificial Intelligence.