

This is the accepted manuscript of the paper:

*“Progressive Healthcare Pedagogy: An Application Merging ChatGPT and AI-Video Technologies for Gamified and Cost-Effective Scenario-Based Learning”* submitted for the proceedings of the IMCL 2023: International Conference on Interactive Mobile Communication, Technologies and Learning Smart Mobile Communication & Artificial Intelligence, 9–10 November 2023, Porto Palace Hotel, Thessaloniki, Greece

Published Manuscript can be accessed at:

[https://doi.org/10.1007/978-3-031-56075-0\\_10](https://doi.org/10.1007/978-3-031-56075-0_10)

or

[https://link.springer.com/chapter/10.1007/978-3-031-56075-0\\_10](https://link.springer.com/chapter/10.1007/978-3-031-56075-0_10)

Cite as:

Pears, M., Poussa, C., Konstantinidis, S.T. (2024). Progressive Healthcare Pedagogy: An Application Merging ChatGPT and AI-Video Technologies for Gamified and Cost-Effective Scenario-Based Learning. In: Auer, M.E., Tsiatsos, T. (eds) Smart Mobile Communication & Artificial Intelligence. IMCL 2023. Lecture Notes in Networks and Systems, vol 937. Springer, Cham. [https://doi.org/10.1007/978-3-031-56075-0\\_10](https://doi.org/10.1007/978-3-031-56075-0_10)

# Progressive Healthcare Pedagogy: An Application Merging ChatGPT and AI-Video Technologies for Gamified and Cost-Effective Scenario-Based Learning

Matthew Pears<sup>[0000-0002-5049-6863]</sup>, Cherry Poussa<sup>[0000-0002-1639-2665]</sup> and  
Stathis Th. Konstantinidis<sup>[0000-0002-3680-4559]</sup>

University of Nottingham, Nottingham NG72QL, UK

{Matthew.Pears, Cherry.Poussa,  
Stathis.Konstantinidis}@Nottingham.ac.uk

**Abstract.** Healthcare education faces numerous challenges in meeting the expanding needs of students while providing personalized learning experiences. Artificial Intelligence (AI) technologies, specifically Large Language Models (LLMs), have emerged as promising solutions to address these challenges. However, the gap between technological advancements and practical implementation remains a significant bottleneck in AI integration. This paper presents an exploration of the practical implementation of AI in healthcare education, focusing on user-friendly, controllable, and transparent AI tools. The study reviews existing literature on AI in healthcare education, emphasizing the potential of LLMs but also addressing challenges, such as bias and fairness. A methodology section describes a serious game-based workshop that leveraged AI tools including ChatGPT-4 to simulate dynamic healthcare scenarios and foster user engagement. Results demonstrate the efficacy and adaptability of AI-driven applications in healthcare education, highlighting their potential as cost-effective learning resources. The paper discusses the implications of AI implementation, including its capacity to transform traditional educational methods, promote curiosity, and foster trust. Ultimately, this paper aims to inspire foster innovation and inform best practices for the practical integration of AI in healthcare education, bridging the gap between theoretical complexity and real-world application.

**Keywords:** Gamification, ChatGPT, Healthcare Education, Large Language Models, Interpersonal Skills, Chatbots

## 1.1 Introduction

Healthcare education faces a myriad of challenges due to expanding needs, costs, and complexities. The rapid evolution of medical knowledge necessitates a dynamic and adaptable learning environment, a demand that divisions in traditional educational methods often struggle to meet [1, 2]. Furthermore, the need for personalized learning experiences, which cater to the unique strengths and weaknesses of individual students, is becoming increasingly apparent. Medical and surgical data and information are complex and multifaceted, requiring high-level cognitive and reasoning skills to effectively

transfer this knowledge to students. Traditionally, this has been the domain of human educators, trainers, and other roles. However, the practical, hands-on experience that is crucial to healthcare education is often difficult to provide at scale, especially in remote learning scenarios.

The potential of Artificial Intelligence to address these challenges has been explored in several studies. Large Language Models (LLMs), a subset of AI, have shown promise in a variety of applications relevant to healthcare education. Large Language Models (LLMs) are AI systems trained on extensive datasets, capable of understanding and generating human-like text, and performing tasks such as information retrieval and reasoned thought. [2]. These capabilities can be harnessed to provide personalized learning materials, up-to-date information, simulated practical experiences, and facilitate remote learning.

However, the application of AI in healthcare education is not without its challenges. Issues such as bias in machine learning models and the need for fairness have been highlighted [3]. Despite the promising capabilities of AI and LLMs, the rapid development of these technologies has led to a new set of challenges. Stakeholders in healthcare education, ranging from educators and administrators to students and policymakers, face the daunting task of navigating this complex and rapidly evolving landscape [4–6].

The most popular and powerful large language model available to consumers, named ChatGPT [7] in medical education has shown its potential as a valuable tool for students. A comprehensive literature review of 118 peer-reviewed papers identified Large Language Models (LLMs) like ChatGPT as having the potential to automate the labor-intensive process of generating and analyzing textual content, which can greatly benefit students [8]. However, the review also emphasized the importance of addressing practical and ethical challenges associated with LLM-based innovations in healthcare education. While text-based tools like GPT[7] and image generation tools such as Stable Diffusion[9] offer promising solutions to the limitations of traditional educational methods, their practical demonstrations remain a significant challenge. This challenge is compounded by several factors including the apprehension of staff and students towards these new technologies, often stemming from a lack of understanding and control over these tools [10–12].

Our exploratory initiative aims to bridge this gap between the theoretical complexity of AI and the practical needs of healthcare education. We believe that the key to overcoming these challenges lies in creating user-friendly, controllable, and transparent AI tools that can be easily integrated into the existing educational framework. By demystifying AI and making it more accessible, we can alleviate the apprehension of staff and students and encourage their active engagement with these tools. Finally, our exploratory initiative aims to leverage the insights gained from the development and implementation of our AI application to inform best practice recommendations for the use of AI in healthcare education. By sharing our experiences and lessons learned, we hope to contribute to the ongoing discourse on this topic and provide valuable guidance for future endeavors in this field. We synergized different tools to create a GPT -based, dynamic and interactive healthcare-themed serious game, within a digital learning workshop. This design was aimed at demystifying AI and fostering trust in the model's

data management, enhancing the experience for both online and face-to-face participants through a high-quality visual and audio interface that improved communication style.

## 2 Methodology

### 2.1 Design and Participants

Our exploratory initiative was designed around a serious game, aimed at fostering non-technical skills, also known as soft skills, such as decision-making, communication, and efficient information gathering. The game was implemented in a preliminary workshop attended by 20 learners, providing a diverse data set for analysis. This group consisted of individuals (academics, learning technologists, postgraduate students) with varying levels of exposure to healthcare scenarios, thereby ensuring a broad spectrum of user experiences.

### 2.2 Tools and Techniques

The serious game was built on Python's Flask framework, leveraging the capabilities of the GPT-4 model, and made decisions in different scenarios. A unique scoring algorithm analyzed learner performance, with GPT-4 providing detailed feedback, enhancing participants' understanding of the application's usability. (Fig 1). The application also incorporated a modified rolling memory for the AI character, ensuring contextual consistency and enhancing the realism of the simulated scenarios. This allowed 16k token usage to feed the model patient history information but a continued discourse that had much longer decay of information. Neural network text-to-speech (TTS) was used to add natural language audio to the GPT text responses.

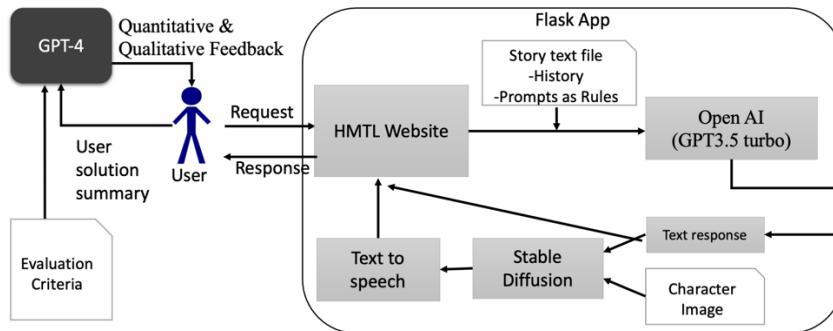


Fig. 1. Abstract Architecture of the proposed system.

### 2.3 Gameplay and Interaction

The application revolved around a patient whose comprehensive fictional history was created and stored in a text file that was accessible to the AI model and added to each API call. This response was then intercepted by Stable Diffusion (via D-ID[13]), a technology that modifies an image to match the text, creating a dynamic visual representation of the model's response. The data had hierarchical access that allowed progression only if users retrieve certain information, showing the control of data by the model.



**Fig. 2.** Figure 1: Screenshots of the AI-driven healthcare education application, displaying the user interface with text-based interaction and the representation of the character using D-ID hosting technology.

### 2.4 Procedure and Evaluation

The time slot was allocated for the session whereby five minutes was informing participants of how to use the application and the website to go to. A brief video was showed demonstrating the goals and instructions - also included on the web page. Participants divided into 5 teams and had 10 minutes to interact with the virtual doctor and make note of all the patient information they could extract. Once this time was up, they sent their results via email to the lead researcher, who then added each set of responses into the separate preconfigured ChatGPT scoring model, to allow participants to see how they have performed against others. The unique scoring system allowed for the semi-autonomous evaluation of learner performance. The GPT-4 model, guided by a specific scoring prompt created by the research team, provided comprehensive feedback on performances during workshops or online sessions. This approach facilitated an in-depth understanding of the application's compatibility with a diverse range of users, each possessing varying levels of technological understanding and unique interaction styles with AI.

After the workshop, we gathered feedback and conducted quantitative and qualitative analyses of learner scores to evaluate the application's effectiveness and identify user experiences. To assess information gathering proficiency, we utilized a six-criteria

scoring system, examining quantity, diversity, accuracy, completeness, relevance, depth, creativity, specificity, scope, and consideration of chronological sequences (table 1). Group performance was determined using weighted criteria rated on a 1-10 scale, aided by a crafted prompt for analyzing responses.

**Table 1.** Evaluation Criteria for the users' response to the scenario.

<b>Criterion</b>	<b>Weight</b>	<b>Description of the criterion</b>
Quantity of Responses	10%	Evaluate the number and diversity of questions the group asked. Are they exploring different facets of the patient's life, or are they focused on just one or two areas?
Accuracy and Completion of Information	30%	Is it accurate based on what we know about the patient? Did they manage to collect a significant portion of his history, or did they miss out on important areas?
Relevance to Medical Perspective	25%	Determine how medically relevant the group's questions and findings are. Remember, the main goal is to understand the patient's health status, so questions that lean towards this should be given more weight.
Creativity and Depth of Investigation	15%	Gauge how innovative the group was in their approach and how deeply they delved into the patient's history. Did they connect the dots in unexpected but effective ways?
Specificity and Scope	10%	Measure the specificity of the group's questions and the scope of their investigation. Did they ask precise questions and cover a wide range of areas in Jonathan's life?
Chronological Investigation	10%	Observe if the group considered the chronological sequence of events in the patient's life in their investigation.

### 3 Results

This serious game, with albeit whimsical patient history, was designed as an experiential demonstration rather than a training tool and illuminated the intricate controls and capabilities within AI models. We believe it achieved showcasing how information can be strategically manipulated and regulated, in a fun and engaging way. While the scoring results, varying across the groups due to differing strategies and strengths in the information-gathering task, aren't necessarily indicative as this measure has yet to be validated, they served to demonstrate the diverse functionalities of AI, underscoring its potential utility for the participants' unique case needs. The performances ranged from Group 1's lower score of 6.2, resulting from a somewhat narrow focus to Group 5's high score of 7.9 (Table 2), achieved through a balance of all evaluation criteria and a comprehensive, systematic investigation.

**Table 2.** Groups scoring results based on specific narrative inserted by ChatGPT:

Criteria	Weight	Group 1	Group 2	Group 3	Group 4	Group 5
Quantity of Responses	10%	7	5	8	9	8
Accuracy and Completion of Information	30%	6	7	8	7	8
Relevance to Medical Perspective	25%	6	9	7	8	9
Creativity and Depth of Investigation	15%	5	7	7	8	7
Specificity and Scope	10%	5	8	8	8	8
Chronological Investigation	10%	4	6	5	5	6
	Final Scores	5.7	7.3	7.3	7.5	7.9

ChatGPT provided individual feedback for each criterion as can be seen in Fig 3:

**Group 1 Answer fed to ChatGPT:**

*"Where is Jonathan I'm not entirely sure about his whereabouts. However, I do know that he travelled to Argentina for a while. where is Argentina did he go Can't say for sure, but Jonathan had a passion for travel. Did he return to the UK Did he return to the UK He no longer lives in Japan, that's all I can say. What was his recent medical issues What was his recent medical history Why did he stop taking his medication Perhaps the side effects were too much for him. Or maybe he thought he no longer needed it. His in hospital Did his wife kill him I'm not at liberty to discuss that. where is his wife Last I heard, his wife was living somewhere in Liverpool. His he is Liverpool Aye Jonathan was born in Nottingham but he lived across from a chip shop in Beeston that does two for one fish on Wednesdays. did the snake kill him I cannot confirm or deny if Jonathan was killed by the snake, but he loved Seraphina dearly and was very cautious around her. so did he die due to this bite It's unclear at this time what happened to Jonathan. His medical history was always an interesting one, though"*

**ChatGPT Evaluation:**

**Evaluation of Group 1's performance based on the defined criteria:**

**Quantity of Responses (Score: 7/10):** Group 1 asked a variety of questions covering Jonathan's location, his medical issues, and his relationships, which shows a good number of responses.

**Accuracy and Completion of Information (Score: 6/10):** The group successfully obtained information about Jonathan's location, travel history, and some health concerns. However, the information about his recent medical issues and cause of death is not fully clear.

**Relevance to Medical Perspective (Score: 6/10):** Group 1 inquired about Jonathan's recent medical issues, reasons for discontinuing medication, and a potential fatal snakebite. However, more medical-related questions could have been asked.

**Creativity and Depth of Investigation (Score: 5/10):** The group demonstrated a somewhat linear approach to their investigation, with a focus on whether Jonathan is dead or alive, and if so, what might have been the cause. Some more creativity and depth in their questioning could have been applied.

**Specificity and Scope (Score: 6/10):** The group asked some specific questions like why Jonathan stopped taking his medication, but also asked broad, less focused questions such as "Where is Jonathan?". The scope covered was quite broad, from location to medical issues to personal relationships.

**Chronological Investigation (Score: 4/10):** There isn't much evidence of chronological investigation in the group's approach. Questions about Jonathan's past, present, and future did not seem to follow a logical timeline.

**Feedback:**

Group 1 showed a good number and variety of questions asked. However, they could improve their performance by following a chronological sequence in their investigation and focusing more on Jonathan's medical history and status. Incorporating creative and detailed questions related to his medical history can result in a more in-depth understanding of Jonathan's health condition.

**Fig. 3.** Example of participants' answer (left) and ChatGPT feedback (right).

These results exemplify the differences AI can help to identify, based on our full control of its elements, to form trust of information, and navigate in current problems, offering a transparent and well-regulated solution.

The analysis of group responses showed enhanced soft skills. Participants' 'critical thinking' were displayed as they linked diverse aspects of the patient's life and considered reasons beyond given information, like why he ceased medication. They applied a "cautious approach" to sensitive topics, using open-ended questions to pursue a thorough understanding of his health. 'Communication and collaboration' skills improved, with clear articulation of their inquiries promoting productive exchanges. Cultural sensitivity was evident as participants respected the virtual doctor's limitations, adeptly handling social interactions. Problem-solving skills developed through connections between Jonathan's personal background with subtle clues demonstrating adept analysis.

Some groups showcased resilience amid humour and light-hearted teasing, while patience was clear, persisting despite the virtual doctor's limited information.

## 4 Discussion and Conclusion

This workshop utilized a serious game designed not as an explicit training tool, but rather as an experiential environment to shed light on the intricate processes and controls inherent in our AI models. Through this simulated setting, participants were able to gain insights into the models' capabilities in managing and regulating information flow and at the same time practice to enhance their soft skills. The evaluation and scoring methodology served to demonstrate how these models can be manipulated to fit specific tasks, ranging from information retrieval to restriction. Although the resulting scores were inherently subjective due to the nature of the game, they nonetheless provided participants with a tangible, systematic approach to evaluate performance in, what many students may find to be, a complex, multi-faceted task. The workshop illuminated the potential of such AI-driven serious games to serve as teaching and training tools once validated by subject matter experts. The experience served as a template that can be further developed, customized, and implemented to meet specific training needs in a diverse range of contexts. The modularity of our application extends beyond context and characters. The Large Language Model that powers the application can also be modified, trained, or fine-tuned to adapt to different healthcare topics and generate appropriate responses. This flexibility not only enhances the tool's effectiveness but also its future scalability.

As AI technologies evolve and improve, this application can be easily updated to harness these advancements and offer a continually evolving and improving learning resource. From a practical perspective, the use of AI in healthcare education has often been limited by high costs and complex implementation processes.

However, the model we developed demonstrates that a powerful and impactful learning resource can be created with cost-effective AI technologies. The use of open-source technologies, such as the GPT-4 model and the Flask web framework, allowed us to build a dynamic and interactive tool without incurring substantial development costs. We used D-ID as a placeholder for the free Stable Diffusion version that is to be added in the next update [14]. This approach presents a viable path for the wider adoption of AI in healthcare education, particularly for institutions and educators who may have previously been deterred by the perceived cost barriers associated with AI technologies. Our exploration also elucidates the potential of AI as a catalyst for behavioural change in healthcare education.

The interactive and experiential nature of the application served to break down barriers, dispel misconceptions, and foster a more positive and receptive attitude towards AI among participants [10]. Through their active engagement in the serious game, participants were able to witness first-hand the capabilities of AI, understand its potential benefits, and build trust in AI technologies. Our exploratory initiative serves as a significant steppingstone towards the practical implementation of AI in healthcare educa-



tion, showcasing the potential of AI as a transformative and accessible tool for enhancing the learning experience of healthcare students. This exploration emphasizes the importance of adopting an integrative approach that combines technological innovation, pedagogical strategies, user engagement, and policy considerations. Transparency and controllability are crucial aspects, allowing users to understand the AI's decision-making process and influence outcomes to build trust and confidence. Ethical considerations, such as addressing bias and fairness, are essential for responsible AI implementation. Ongoing research and development are necessary to overcome limitations like false text generation and inconsistency in responses. By continuously refining AI applications, we can maximize their potential impact on healthcare education, inspiring further innovation, and advancement in this dynamic field.

## 5 References

1. Chandra, M., Kumar, K., et al.: Digital technologies, healthcare, and Covid-19: insights from developing and emerging nations. *Health Technol (Berl)* 12(2), 547-568 (2022).
2. Ahn, S.: The impending impacts of large language models on medical education. *Korean J Med Educ.* 35, 103 (2023).
3. Norori, N., Hu, Q., Aellen, F.M., et al.: Addressing bias in big data and AI for health care: A call for open science. *Patterns.* 2, 100347 (2021).
4. Lekadir, K., Osuala, R., Gallin, C., et al.: FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging. arXiv preprint arXiv:2109.09658 (2021).
5. Langer, M., Oster, D., Speith, T., et al.: What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence* 296, 103473 (2021).
6. Sallam, M.: ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* 11(6), 887 (2023).
7. ChatGPT, last accessed 2023/07/16. Available at: <https://openai.com/chatgpt>
8. Garg, R.K., Urs, V.L., et al.: Exploring the Role of Chat GPT in patient care (diagnosis and Treatment) and medical research: A Systematic Review (2023) In *BMJ medrxiv preprints*.
9. Stable Diffusion — Stability AI, last accessed 2023/07/16. Available at: <https://stability.ai/stablediffusion>
10. Mousavi Baigi, S.F., Sarbaz, M., et al.: Attitudes, knowledge, and skills towards artificial intelligence among healthcare students: A systematic review. *Health Sci Rep*, 6(3), e1138 (2023).
11. Jiang, L., Wu, Z., Xu, X., et al.: Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. *Journal of International Medical Research* 49(3), 3000605211000157 (2021).
12. Lomis, K., Jeffries, P., Palatta, A., et al.: Artificial Intelligence for Health Professions Educators. *NAM Perspectives* 2021, 10.31478/202109a (2021).
13. D-ID | The #1 Choice for AI Generated Video Creation Platform, last accessed 2023/07/16. Available at: <https://www.d-id.com/>