dHealth 2025
M. Baumgartner et al. (Eds.)

doi:10.3233/SHTI250180

LLM-Driven Adjustments in Serious Games: A Feasibility Analysis

Ivana MOSTACHETTI[a,1], Andrea VITALI[a], Daniele REGAZZONI[a], Caterina RIZZI[a] and Giovanni Pietro SALVI[b]

[a] *Department of Management, Information and Production Engineering, University of Bergamo, via Pasubio, Dalmine (Bergamo), Italy*

[b] *Istituto Clinico Quarenghi, Neurology Unit, San Pellegrino Terme (Bergamo), Italy*

Abstract. Background**:** Serious games (SGs) and telerehabilitation play a key role in the recovery of lost functions in neurological patients, with personalisation and difficulty adjustment being essential features. Objectives**:** This work investigates the feasibility of integrating a large language model (LLM) into an Assessment Serious Game (ASG) to analyse exercise data and recommend personalised rehabilitation programs. Methods**:** Medical knowledge was acquired through meetings with professionals to identify target pathologies and parameters. The ASG was integrated with GroqCloud; the prompt is designed to act as physiotherapist and SG developer to make real-time adjustments and suggest the setting configurations of other SGs. A preliminary test assessed the system's capabilities. Results**:** The LLM effectively recognises real-time adjustments and follows instructions for SGs parameter settings. However, limitations remain in the degree of adjustments and numerical parameter suggestions. Conclusion**:** The analysis demonstrates the feasibility of a designed LLM prompt to adjust SG difficulty and recommend setup parameters, while highlighting areas for improvement in reliability and accuracy.

Keywords. Telerehabilitation, Neurological Rehabilitation, Upper Extremity, Exergaming, Large Language Model

1. Introduction

Recently, serious games (SGs) in rehabilitation have become increasingly important, especially with the rise of telerehabilitation, in the recovery process of patients suffering from neurological disorders (e.g. Parkinson's disease, multiple sclerosis) [1].

Telerehabilitation supports long-term, personalised plans by adapting difficulty to the patient's evolving abilities [2]. Difficulty levels are configured either manually by physiotherapists [3], often using pre-set levels (e.g., low, medium, high) and requiring constant expertise while relying on subjective evaluations, or automatically through deep learning and artificial neural networks (ANN). SGs paired with motion-tracking devices (e.g., motion capture systems) generate extensive data during exercises, enabling ANN techniques [4], [5].

In recent years, Large Language Models (LLMs) have gained ground also in healthcare and rehabilitation. According to Naqvi et al. [6], LLMs provide more reliable

[1] Corresponding Author: Ivana Mostachetti, Department of Management, Information and Production Engineering, University of Bergamo, Dalmine (Bergamo), Italy, E-Mail: ivana.mostachetti@unibg.it

results when the prompt is properly designed. In rehabilitation, there are examples where LLMs have been introduced in chatbot-based applications [7], which often require constant intervention by healthcare professionals. An example where LLMs dynamically adjust the SG parameters through continuous textual prompts provided by the teacher is provided in the educational context by Bonetti et al [8]. Unlike traditional approaches that rely on fixed levels of difficulty, LLMs enhance personalisation by suggesting real-time adjustments to individual parameters based on patient's progress [6, 8], ensuring an automatically tailored and continually evolving rehabilitation experience.

This project explores the feasibility of using a LLM in a SG to enhance rehabilitation by analysing patient data and dynamically adjusting the program to the patient's abilities. The role of the LLM is to automatically suggest which SGs of the platform and the relative configurations the patient should perform according to his/her abilities, with healthcare personnel able to modify the settings only if strictly necessary. The LLM prompt is designed to act as both a physiotherapist and a SG developer. As a physiotherapist, it automatically assesses the patient's performance by analysing the hand trajectory and other relevant parameters collected during the session. As a game developer, it applies its understanding of SG functioning and output generation to determine the best configuration for future sessions.

2. Methods

The section is structured as follows. Section 2.1 explains the construction of the medical knowledge used to design the prompt, section 2.2 describes the tools used and section 2.3 presents the aims of the proposed solution.

2.1. Medical Knowledge Acquisition

As a first step in the project, acquiring the necessary medical knowledge was essential to design the LLM prompt effectively. This was achieved through meetings with a team of health professionals, including three physicians, two physiotherapists and an occupational therapist, in order to identify the target pathologies and the parameters characterising the patient's abilities. The pathologies of interest are: (i) Parkinson's disease, (ii) cerebral stroke, (iii) multiple sclerosis, (iv) traumatic brain injury, (v) spinal cord injury and (vi) cerebellar disorders.

To ensure a comprehensive yet focused assessment, some parameters provide a more accurate representation of the patient's clinical condition than others and should therefore be prioritised in the analysis. Hence, the medical team was asked to identify and rank the most relevant parameters for each pathology.

Figure 1 summarizes the results of this process, showing an overview of the parameters used to assess the patient's abilities and ranking their importance for each pathology, starting from the 1st position, where a lower number reflects higher relevance. In particular, the left column lists the parameters, which include both kinematic and cognitive measures; each parameter is briefly described to clarify its clinical relevance. The right column shows the ranking of importance for each pathology. On the far right, a legend assigns a specific colour to each pathology, allowing quick identification and interpretation of the data. For example, if a patient suffers from spinal cord injury, the most important parameter in defining his or her abilities is “pinch distance”, followed by "precision", and, finally, equally important, "time" and "covered area"; all other parameters are taken into account, although to a lesser extent.

Acquired Parameters	Importance					
	Parkinson's Disease	Cerebral Stroke	Multiple sclerosis	Traumatic Brain Injury	Spinal Cord Injury	Cerebellar Disorders
VELOCITY: Velocity of the gesture execution	1	4				
DISPLACEMENT: Covered displacement during the exercise		4		3		
TIME: Time taken to complete the exercise	4	3		3	3	2
PINCH DISTANCE: Distance between the tip of the index finger and the thumb when pinching			2		1	3
ERRORS: Number of incorrect targets made during the exercise			3	1		
PATH ERROR: Ratio of the optimal distance the patient should run moving from one point to the next to the distance actually run	2	2				
PRECISION: How close to the centre the target is hit			1	3	2	1
COVERED AREA: Index of the range of movement; ratio of the area covered by the patient during the exercise to the available one	3	1			3	
TREMORS: Presence of tremor during the performance of the exercise	4		2			2
COLLABORATION: Assessment of the patient's ability to follow the instructions given, considering precision and continuity of movement				2		
COMPREHENSION: Ability to understand the instructions and objectives of the exercises, considering errors and reaction time		4		2		

Figure 1. The diseases considered and the parameters prioritised to define the patient's condition

2.2. Tools

2.2.1. Assessment Serious Game (ASG) and N-RehLab Platform

In a previous work, we developed the Assessment Serious Game (ASG) [9], combined with a Leap Motion Controller [10] and aimed to evaluate the patient's abilities. The ASG consists of two levels, "Connect the dots" and "Pop the balloons", each composed of three steps. The parameters acquired in the ASG, shown in Table 1, are used to automatically set and update the difficulty levels of the SGs available in the N-RehLab platform, such as *Whac-a-Mole*, *Moka Coffee* and *Paint On Canvas*. This web interface allows medical staff to monitor performance and confirm suggested ASG configurations.

2.2.2. LLMs and Prompt

GroqCloud [11] has been implemented in the ASG, using the *mixtral-8x7b-32768* model of Groq, since it provided the most accurate responses for our purposes, particularly in comparing numerical values derived from acquired kinematic parameters [12]. The integration between the ASG and GroqCloud is via Rest APIs: the ASG sends patient performance data to GroqCloud, which processes it and provides feedback to refine the rehabilitation program. The communication between the ASG and the LLM is direct, automatic, and continuous, enabled by a C# module that interacts with the LLM. This module continuously analyse the data, i.e. hand's trajectory and other parameters in Figure 1, recorded during the rehabilitation session and dynamically modifies the parameters settings, ensuring an adaptive rehabilitation process tailored to the patient's evolving needs.

GroqCloud is supposed to have three "actors":

- "**System**", to which we describe the context, the medical knowledge, which role the LLM prompt is designed for, how the ASG works, ect.;
- "**User**", in this case the ASG, which provides the information and the data about the rehabilitation session of the patient;
- "**Assistant**", the response of the LLM, which must follow the instructions provided to the system.

The GroqCloud prompt has been designed to simulate a physiotherapist who assesses the key aspects of the rehabilitation and configures the SGs accordingly. Using the medical knowledge described in section 2.1, the LLM evaluates parameters from rehabilitation sessions, adjusts the ASG settings in real-time and suggests which SGs in N-RehLab are most suitable for the patient's abilities and their the configuration settings.

2.3. Preliminary Test of the Solution Technological Feasibility

The preliminary test aims to assess (i) the system's ability to make real-time adjustments to the ASG difficulty level based on the patient's performance and (ii) the technological feasibility and reliability of the LLM in defining the SGs configuration parameters in N-RehLab. This is done according to medical advice, while checking for potential hallucinations, e.g. incorrect configurations, incorrect analysis or comparison of data.

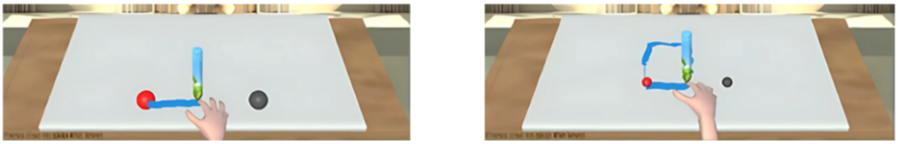
To evaluate the technological feasibility, standardisation, and potential hallucinations related to the second aim, we submitted several requests to the LLM using the same input data text files: Session A (first session, worse performance) and Session B (second session, better performance). In the preliminary test, the prompt was designed to analyse and compare a cerebral stroke patient's most recent performance (Session B) with a previous one (Session A). The instructions included parameters ranking shown in Figure 1, an explanation of the ASG, and how the parameters acquired by the ASG influence the configuration of the SGs in N-RehLab.

3. Results

3.1. Real-Time Adjustments

Real-time adjustments of the level of difficulty correctly detects when the conditions are right to increase the difficulty of the exercise, or when it needs to remain unchanged due to suboptimal patient performance. In particular, the C# module that invokes the Rest API with LLM is automatically executed by code at the end of each completed step.

For instance, Figure 2 illustrates the first (a) and the second (b) steps in "Connect the dots". As there was a continuous and significant improvement in the achievement of all targets in the first phase, the size of the targets was reduced in the second step.



(a) (b)

Figure 2. Reduction in target size between the step 1 (a) and the step 2 (b) in the "Connect the dots" level.

3.2. Suggestion of Configurations

The results show that the LLM accurately evaluates the parameters obtained during the exercises. Specifically, the C# module that calls the LLM's Rest API is automatically executed by the code at the end of the exercise, comparing the current rehabilitation session with the previous one stored in N-RehLab. Furthermore, the configuration settings it proposes for the N-RehLab SGs are in line with the instructions given during the prompt design. During the preliminary test, we noticed that the LLM sometimes provides descriptive outputs instead of numerical values. For example, the parameter 'Mole change time' in Whac-a-Mole was described twice as follows: *'Reduce the 'Mole change time' parameter from X seconds to Y seconds (where Y < X)'*.

To address this, we have re-designed the prompt by adding instructions with greater complexity, completeness and constraints, such as requiring integer values for certain parameters. These adjustments improved the reliability of responses compared to previous results. Table 1 shows the results of 10 iterations: the first column lists the SGs, the second column lists the parameter and its possible settings, and the third the results.

Table 1. Results of the configuration parameters proposed by LLM starting from the comparison of the same two performances multiple times

SGs	Parameter (setting values)	Results of 10 Iterations
Whac-a-Mole	Changing moles time (integer)	2-4 seconds
	Score to reach (integer)	15-25
Mokka Coffee	Dimension (Big or Normal)	Normal
	Sequence hints (Yes or No)	No
	Gesture (Pinch or Touch)	Pinch
	Prono-Supination (Yes or No)	No
	Object Location (Ordered or Casual)	Ordered
Paint On Canvas	Figure to draw (Easy, Medium, Hard)	Medium
	Figure to fill (Easy, Medium, Hard)	Medium
	Required Precision (Low, Medium, High)	Medium
	Gesture (Pinch or Touch)	Pinch

4. Discussion and Conclusions

The proposed work is a preliminary analysis to verify its technological feasibility, mostly to understand the presence of hallucinations and the ability of the system to compare data.

The solution shows that the data collected by the ASG are correctly analysed by GroqCloud, which automatically adjusts the level of difficulty in real-time by means of Rest API. However, the degree of changes, such as the reduction of the size of dots or balloons, described in section 3.1, is sometimes exponential rather than proportional to the session data. This suggests the need for more refined instructions in the prompt design to ensure proportional adjustments.

Secondly, the results show that the solution is able to identify and consistently suggest the parameter configuration values of the other SGs in N-RehLab based on the prompt design. In our initial results, twice the suggested configurations were expressed in words rather than the expected integers. The reliability of the responses improves with more detailed and restrictive prompt design. As shown in Table 1, the proposed research standardises configuration parameters for ordinal categorical or boolean values, e.g., gesture or object location. However, numerical parameters still represent a challenge for standardisation, although the resulting values fall within ranges deemed acceptable by the medical personnel involved in the project, underlining the absence of hallucinations.

In conclusion, the proposed solution represents a preliminary technological investigation into the feasibility of using LLM prompt in a SG. The results demonstrate the ability of LLM to adjust the difficulty level of the SGs according to the prompt design and to coherently suggest SGs configurations, without hallucinations. However, the reliability and accuracy of the adaptations can be improved, as current limitations arise from insufficient system information.

Future work will focus on improving the prompt design, incorporating Retrieval-Augmented Generation (RAG), and extending the knowledge provided. Large-scale tests, first with healthy subjects and then with neurological patients, are also planned to evaluate the validation of the proposed work.

Acknowledgements

The authors would like to thank the medical personnel of *Istituto Clinico Quarenghi* (San Pellegrino Terme, Bergamo) for their precious support in defining parameters to be measured. This project of Mostachetti I. is funded by PNRR (DM 118/2023 for research topics related to public administration). This work is partially funded by the National Plan for NRRP Complementary Investments - project n. PNC0000003 – AdvaNced Technologies for Human-centrEd Medicine (ANTHEM).

References

[1] C. S. González-González et al., Serious games for rehabilitation: Gestural interaction in personalized gamified exercises through a recommender system, *Journal of Biomedical Informatics* **97** (2019), 103266

[2] N. Belotti et al., A Tele-Rehabilitation Platform for Shoulder Motor Function Recovery Using Serious Games and an Azure Kinect Device, *Studies in health technology and informatics* **293** (2022), 145–152

[3] H.-T. Jung et al., Rehabilitation Games in Real-World Clinical Settings: Practices, Challenges, and Opportunities, *ACM Trans. Comput.-Hum. Interact.* **27**(6) (2020),

[4] K. Chen et al., Generative AI Based Difficulty Level Design of Serious Games for Stroke Rehabilitation, *IEEE Internet of Things Journal* (2024),

[5] N. Nasri et al., An semg-controlled 3d game for rehabilitation therapies: Real-time time hand gesture recognition using deep learning techniques, *Sensors (Switzerland)*, **20**(22). MDPI AG, 1–12, (2020),.

[6] W. M. Naqvi et al., Large language models in physical therapy: time to adapt and adept, *Frontiers in Public Health* **12** (2024),

[7] R. Almeida et al., Physio: An LLM-Based Physiotherapy Advisor, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **14612 LNCS** (2024), 189–193

[8] F. Bonetti et al., Using LLMs to Adapt Serious Games with Educators in the Loop, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **15348 LNCS** (2025), 68–77

[9] I. Mostachetti et al., Automatic Configuration of a Telerehabilitation Platform for Neurological Patients, *Design Tools and Methods in Industrial Engineering IV* (2025), 117–124

[10] Ultraleap, https://www.ultraleap.com/leap-motion-controller-overview/ (accessed May 13, 2024).

[11] Playground - GroqCloud, https://console.groq.com/playground (accessed Jan. 15, 2025).

[12] mixtral, https://ollama.com/library/mixtral (accessed Mar. 11, 2025).