Guilherme Costa Antunes

Ginásio Virtual para Casas Inteligentes Conectadas

Virtual Gym for Connected Smart Homes

Universidade de Aveiro
2025

Guilherme Costa Antunes

Ginásio Virtual para Casas Inteligentes Conectadas

Virtual Gym for Connected Smart Homes

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Informática, realizada sob a orientação científica do Doutor Nuno Filipe Correia de Almeida, Investigador auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor António Joaquim da Silva Teixeira, Professor catedrático do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

O presente trabalho foi realizado ao abrigo do Projeto “Connected Home for Healthy Ageing (CHHA)”, financiado pela União Europeia, no âmbito da Open Call #2 do projeto “IMAGINE-B5G” do programa de investigação e inovação “Horizonte Europa”.

o júri / the jury

agradecimentos /
acknowledgements

Em primeiro lugar, gostaria de expressar o meu profundo agradecimento aos meus orientadores, o Professor Nuno Almeida e o Professor António Teixeira, por toda a orientação, disponibilidade, apoio e paciência durante todo o desenvolvimento desta dissertação. Quero agradecer-lhes também pela oportunidade de ter integrado o CHHA e o CasaViva+ e por poder colocar em prática as minhas ideias. Quero agradecer também ao Pedro Carneiro pelo trabalho que desenvolveu em favor deste projeto e por toda a disponibilidade que sempre mostrou para ajudar à sua concretização. Muito obrigado, Pedro.

Quero também agradecer aos meus pais, pois tudo o que vivi e aprendi, tanto agora como toda a minha vida, foi graças a eles, que fazem tudo por mim e pelo meu irmão, a quem também agradeço por me ter aturado todos estes anos. É a eles que dedico este trabalho. Agradeço também a toda a minha restante família pelo apoio incondicional.

Uma palavra também aos professores, treinadores, colegas e pessoas que, de algum modo, fizeram parte da minha vida e que, sem eles, eu certamente não seria quem sou hoje.

Por fim, mas muito importante, agradecer aos meus amigos, porque a verdade é que, sem eles, estes 5 anos não tinham sido o que foram. Perdoem-me por não escrever os nomes de todos, mas vocês também sabem quem são. Agradeço por todo o apoio e por todos os momentos vividos que ficarão para sempre na memória. Sem vocês, nada disto teria tido piada. Muito obrigado por tudo.

palavras-chave

Ginásio Virtual, Estimação de Pose, Avaliação de Exercício, Comunicação 5G, Processamento Remoto, Interação por Voz, Casas Inteligentes, Envelhecimento Ativo

resumo

O envelhecimento da população apresenta desafios crescentes para os sistemas de saúde, particularmente na garantia da autonomia, da motivação e do bem-estar físico entre os adultos mais velhos. Neste contexto, os projetos Connected Home for Healthy Ageing (CHHA) e Casa Viva+ têm como objetivo promover o envelhecimento ativo por meio de ambientes de casas inteligentes, apoiados na conectividade 5G e em serviços digitais. Esta dissertação contribui para estas iniciativas ao propor o desenvolvimento de um sistema de ginásio virtual gamificado para uso doméstico. O objetivo principal foi criar uma solução que incentive a prática regular de atividade física, monitorize a execução dos exercícios e forneça feedback e motivação em tempo real por meio de estratégias de gamificação. A prova de conceito desenvolvida e avaliada suporta interação por voz e resposta gráfica, permitindo também a avaliação remota dos exercícios utilizando tecnologias de estimação de pose alojadas na cloud. O protótipo desenvolvido integra o MediaPipe Pose para análise de movimento em tempo real e foi implementado com sucesso num ambiente de casa inteligente. Os testes realizados no âmbito do CHHA e Casa Viva+ demonstraram a usabilidade, a capacidade de resposta e o potencial do sistema para apoiar o envelhecimento ativo através de rotinas de exercício acessíveis e personalizadas.

keywords

Virtual Gym, Pose Estimation, Exercise Assessment, 5G Communication, Remote Processing, Speech Interaction, Smart Home, Active Ageing

abstract

The ageing of the population presents growing challenges to healthcare systems, particularly in ensuring autonomy, motivation, and physical well-being among older adults. In this context, the Connected Home for Healthy Ageing (CHHA) and Casa Viva+ projects aim to promote active ageing through smart home environments supported by 5G connectivity and digital services. This dissertation contributes to these initiatives by proposing the development of a gamified virtual gym system for home use. The main objective was to create a solution that encourages regular physical activity, monitors exercise execution, and provides real-time feedback and motivation through gamification strategies. The proof-of-concept developed and evaluated supports spoken interaction and graphical output and enables remote evaluation of exercises using cloud-hosted pose estimation technologies. The developed system integrates MediaPipe Pose for real-time movement analysis and was successfully deployed in a smart home environment. Tests conducted in the scope of CHHA and Casa Viva+ demonstrated the system's usability, responsiveness, and potential to support active ageing through accessible, and personalized exercise routines.

acknowledgement of use of AI tools

Recognition of the use of generative Artificial Intelligence technologies and tools, software and other support tools.

I acknowledge the use of Grammarly (Grammarly, https://app.grammarly.com/) to improve the structure of some sentences, grammatical structure, punctuation, and vocabulary.

I acknowledge the use of ChatGPT (Open AI, https://chat.openai.com), Perplexity AI (Perplexity, https://www.perplexity.ai) and Gemini (Google, https://gemini.google.com/app) to help search for, and summarize articles during the background and related work research process.

I acknowledge the use of Visual Studio Code (Microsoft, https://code.visualstudio.com) for writing project code.

I acknowledge the use of Copilot (Github, https://github.com/features/copilot), Claude (Anthropic, https://claude.ai/), ChatGPT (Open AI, https://chat.openai.com), and Gemini (Google, https://gemini.google.com/app) for auto-completion, code generation, and searching for approaches.

I acknowledge the use of Gemini (Google, https://gemini.google.com/app) for image generation.

"A paixão será sempre um motor essencial ao sucesso."

Contents

List of Figures

List of Tables

Glossary

AES	Advanced Encryption Standard
API	Application Programming Interface
AR	Augmented Reality
CHHA	Connected Home for Healthy Ageing
CPU	Central Processing Unit
DTLS	Datagram Transport Layer Security
FCN	Fully Connected Network
FPS	Frames Per Second
GPU	Graphics Processing Unit
HLS	HTTP Live Streaming
HPE	Human Pose Estimation
HTTP	Hypertext Transfer Protocol
ICE	Interactive Connectivity Establishment
IP	Internet Protocol
LiDAR	Light Detection And Ranging
MPP	MediaPipe Pose
NAT	Network Address Translation
NTP	Network Time Protocol
OP	OpenPose
P2P	Peer-to-peer
PEM	Pose Estimation Module
REST	Representational State Transfer
RTMP	Real-Time Messaging Protocol
SDP	Session Description Protocol
SIM	Subscriber Identity Module
SRT	Secure Reliable Transport
SRTP	Secure Real-time Transport Protocol
STT	Speech-to-Text
STUN	Session Traversal Utilities for NAT
SVM	Support-Vector Machine
TCP	Transmission Control Protocol
TTS	Text-to-Speech
TURN	Traversal Using Relay around NAT
UCD	User Centered Design
UDP	User Datagram Protocol
USB	Universal Serial Bus
UX	User Experience
WebRTC	Web Real-time Communication
WEM	Workout Evaluator Module
WHO	World Health Organization
WS	WebSocket

CHAPTER 1

Introduction

This chapter will present the context in which the dissertation is framed and the motivation behind its development. It will also outline some of the challenges that the project aims to address, considering the current context, the specific objectives of the thesis, and the document's structure.

1.1 Context and Motivation

In several countries, including Portugal, the population is ageing at an increasing rate [1]. In general, older people naturally require more healthcare, which, unfortunately, as the proportion of the working-age population decreases, is beginning to lack the human resources necessary to provide the care required for each elderly person and their treatments [2]. Therefore, it is essential to invest in new mechanisms for disease prevention and treatment.

Adopting an active lifestyle is one of the easiest and most effective ways to prevent health issues and maintain overall well-being. To allow people to stay active it is crucial to provide them with the necessary conditions, regardless of their health, financial, or social status. It is also essential to keep them motivated so that they do not give up, thereby benefiting from the long-term results, known to be derived from regular exercise.

Since January 2025, this work has been developed and supervised within the scope of the Connected Home for Healthy Ageing (CHHA) project, which aims to create a next-generation connected home environment focused on promoting health and well-being, particularly for older adults. The project integrates advanced sensors, continuous monitoring, and digital services supported by 5G networks. Among the innovative solutions are a gamified virtual gym, where this work is specifically positioned, and a radar-based monitoring system designed to detect activity patterns within the home.

These technologies, tested in real-life environments, aim to lay the foundation for new strategies in active and personalized ageing, ensuring greater autonomy, safety, and quality of life at home.

The development of technology-mediated mechanisms that allow independence from physical contact with human specialists and treatment locations has significant potential in this field. As a proof-of-concept for this new paradigm, innovative solutions are being implemented for a smart home designed for older adults in a partnership between OLI company [3], the Rovisco Pais Rehabilitation Center, and the University of Aveiro. One of the proof-of-concept being developed is an exercise space within the home that provides support for tutoring, exercise monitoring, and the creation of a collaborative virtual gym, including rewarding users for their efforts.

1.2 Challenges

Exercise is not part of everyone's daily routine. In the 21st century, where physical and mental health are increasingly a concern, it is widely recognized that exercising or simply staying active is a significant step toward better health and improved well-being for all aspects of life without requiring much time.

For individuals, some of the main reasons for not exercising are the lack of time, lack of companionship, inadequate infrastructure, lack of energy, lack of motivation, and not knowing how to do it, among others. Some people even mention that exercising is boring [4] [5].

The lack of exercise has both short-term and long-term effects. For older people, it can lead to cardiovascular, muscular, and bone problems, affecting mobility [6]. Furthermore, the lack of proper treatment for these injuries leads to a rise in sedentary behavior, perpetuating the issue.

From these general challenges, several more operational ones emerge:

- Motivate people to exercise regularly, particularly those who are less active
- Evaluate exercise quality accurately and affordably, tailored to user abilities, while ensuring user-friendliness
- Make home a conducive environment for exercise and physical activity by ensuring accurate monitoring, precise evaluation, and high motivation levels over time.

1.3 Objectives

Aligned with the mentioned challenges, the main objective of this dissertation is to provide house inhabitants with a solution capable of helping them exercise at home while also keeping them motivated in this routine. The system should be capable of

accompany the users during their exercises, monitor exercises execution, and provide feedback on the user, assigning different scores during the training process based on the user's performance in each activity. This aims to improve the execution technique of each task, enhance the efficiency of each workout, and help users achieve their desired results more easily.

As an additional objective, motivated by the relation of this work to ongoing research projects (Casa Viva+ and CHHA), the system must be based on an existing home gym setup and be prepared for deployment in the real house, such as the one being constructed in the scope of project Casa Viva+ and to use 5G network communications for cloud data processing in the scope of project CHHA.

1.4 Method

To achieve the objectives, the following general method was adopted:

1. The Engineering Research method [7] as the basis, with the initial phase of problem selection and definition, followed by the conceptualization and development of a solution, that, in the third phase, is evaluated.
2. User-centered development, including the definition of personas representing the target audience, application scenarios, system requirements analysis, and solution architecture development
3. Iterative development, repeating the phases of development and testing until a solution that meets the project's requirements is achieved and fulfills the objectives
4. Exploration and testing of different tools and techniques, evaluating which ones best fit the project's requirements

1.5 Document Structure

This document is divided into seven chapters.

Chapter 1 is dedicated to introducing the topic of this dissertation, as well as its context and motivation, the challenges it aims to address, and the objectives it seeks to achieve.

Chapter 2 presents the background and related work already available on the subject of this thesis, as well as on the underlying topics, explaining the main concepts that will be addressed and presenting and comparing the work already carried out in the field.

Chapter 3 comprised the requirements gathering process, starting with the creation of personas and usage scenarios.

Chapter 4 presents the proposed system architecture, its various modules, and the communication specifications between system components.

Chapter 5 explains how each part of the system was implemented, the technologies used, and the reasons supporting those choices.

Chapter 6 lists the tests performed and explains the results obtained, as well as the modifications made based on those results.

Chapter 7 presents the conclusions of the dissertation, the work carried out to reach the final prototype, and the possible future developments that can be pursued from the current state of the implementation.

CHAPTER 2

Background and Related Work

This chapter introduces and studies the topics relevant to this dissertation. The Background section contextualizes the project's foundation, offering an understanding of these concepts and how the project can be guided in that direction. The Related Work section evaluates the solutions and strategies already developed to identify potential starting points and areas for improvement, aligning them with the thesis objectives.

2.1 Background

This dissertation proposes developing a solution that leverages existing pose-estimation tools to evaluate exercises and physical activity while applying gamification strategies to encourage continuous use. To understand the available methods, tools and the context in which the solution will be evaluated, background information on Active Ageing, Gamification in Healthcare, and Pose Estimation models is essential for developing and understanding the proposal The next subsections will be dedicated to presenting important aspects of Active Ageing, pose estimation, and gamification in the healthcare context.

2.1.1 Active Ageing

The World Health Organization (WHO) defines the concept of Active Ageing as "the process of optimizing opportunities for health, participation and security in order to enhance quality of life as people age" [8]. The WHO's goal is to address the challenges posed by the ageing of the global population by promoting a shift in how societies view and include older individuals. The WHO's Active Ageing framework is built on three fundamental pillars:

Health: Keeping the risk factors for chronic diseases and functional decline low while increasing protection enables individuals to live longer and with a better quality of life. Maintaining a healthy body allows people to take control of their lives as they age. Nonetheless, everyone who requires healthcare should have access to it whenever needed [8].

Participation: Different sectors of society, such as the labor market, education, and healthcare, should support the full participation of older adults in activities according to their human needs, abilities, and preferences, whether they are socioeconomic, cultural, or spiritual in nature so that they can continue to be productive members of the society in which they live [8].

Security: For older individuals who cannot support and protect themselves, policies and programs should help them and their families secure social and financial support, physical security needs, and rights as they age. The goal is to ensure their protection, dignity, and care should they ever need it [8].

Active ageing offers numerous benefits. Some of the most important include maintaining cognitive function, helping to slow down the loss of brain functions, and enhancing physical health [9], [10]. Promoting a healthy lifestyle helps prevent motor disabilities and chronic issues [10].

Active ageing should be one of the main concerns of modern society. With the increase in average life expectancy, we can expect greater longevity for ourselves and older individuals [11]. Therefore, it is crucial to help older individuals remain active and integrated into society to ensure healthy living of the community and the individuals' survival. In countries where birth rates do not keep pace with mortality rates, this concern is even greater as the proportion of older people continues to rise [1], and human resources are needed to keep the country's services working.

It is essential to provide older adults with conditions that allow them to stay active and stimulate their abilities, for example, through technology. Keeping these individuals within the active population contributes to a country's economy and social welfare and helps combat the "Silent Epidemic" of sedentary behavior [12], [13].

Exercise plays a crucial role in promoting active ageing, it helps maintain and improve functional abilities, bone and muscle health, and cardiovascular health. It also indirectly supports psychological and cognitive well-being, acting as a significant booster of mental health. [14]–[18]. For this reason, the conditions provided to foster active ageing should strongly emphasize physical exercise and strategies that promote its practice.

2.1.2 Pose Estimation

Pose Estimation is a computer vision technique that identifies objects' position and orientation in images and videos by detecting key points on the observed bodies [19]. In general , the human body is identified by detecting specific key points and joints, such as shoulders, elbows, hips, knees, and others, estimating the body's posture based on the configuration of these points [19]. It usually faces and successfully overcomes the challenge of identifying key points over clothing. This technique is applied with great success in 2D images [20]. In addition to estimating body posture (see Figure 2.1), many tools also allow the identification of facial expressions (see Figure 2.2) and the position of fingers on the hands (see Figure 2.3) [21]. This techniques can benefit healthcare, sports analysis, interactive games, entertainment, and gesture-based interaction.

Figure 2.1: MediaPipe Pose Pose output example [22]

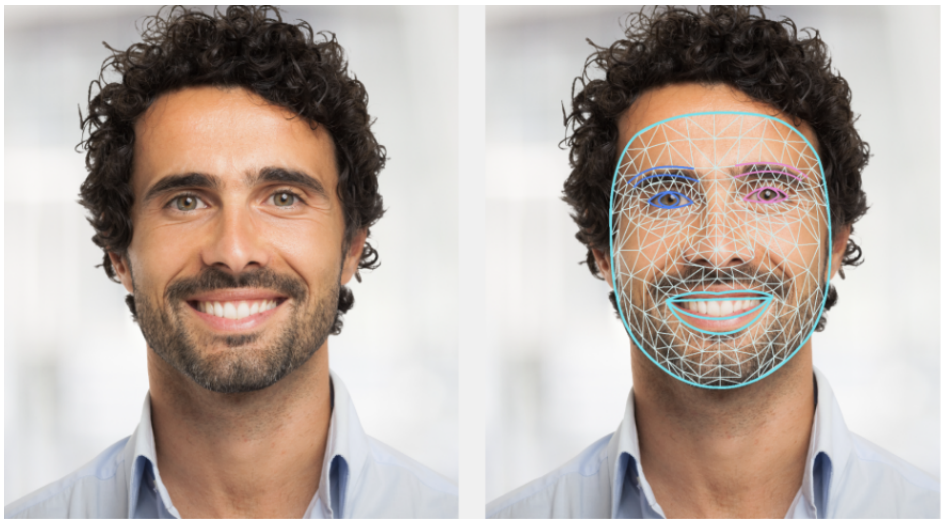

Figure 2.2: Human Pose Estimation applied to Face Landmark Detection [23]

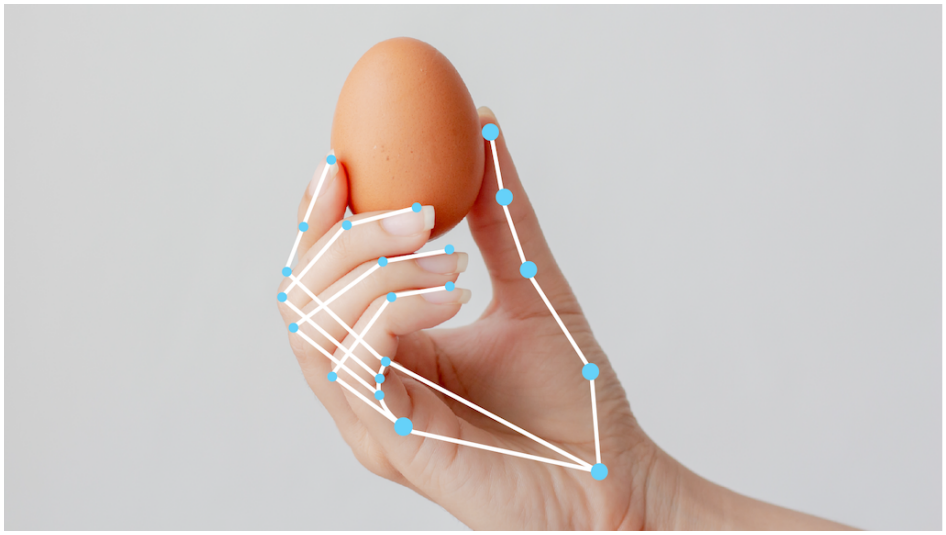

Figure 2.3: Human Pose Estimation applied to Hand Landmark Detection [24]

To detect a person's body position, Human Pose Estimation (HPE) models must be capable of identifying humans in images and detecting the key points for marking the skeletons. There are two approaches to achieve this:

- Top-Down approach, where the model first identifies the person before detecting the key points [25] (see Figure 2.4 (a))
- Bottom-Up approach, where key points of interest in the image are detected and later grouped to form the body parts belonging to each individual [25] (see Figure 2.4 (b))

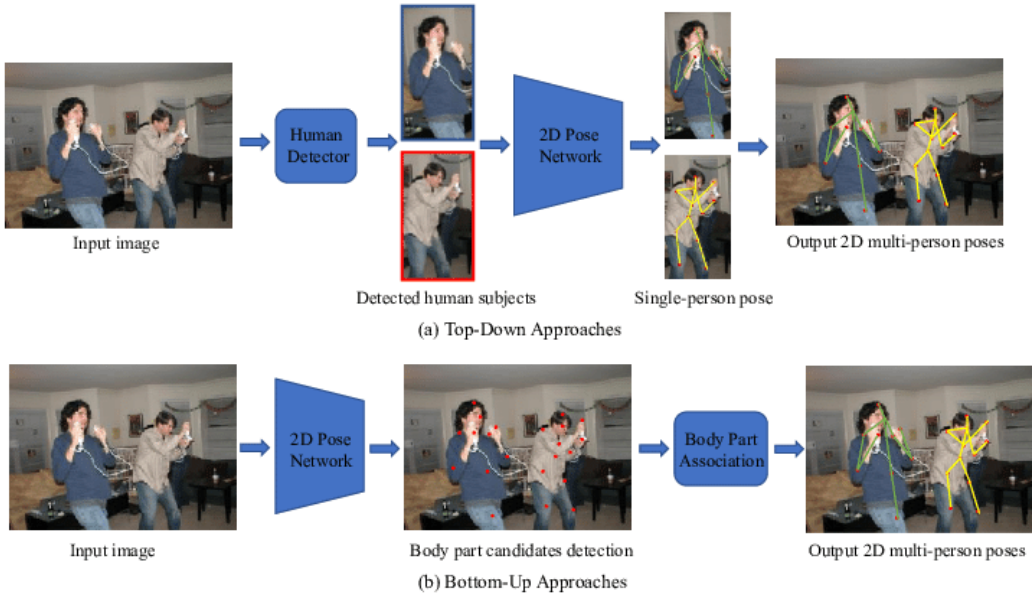

Figure 2.4: Human Pose Estimation possible approaches [25]
(a) Top-Down Approaches, (b) Bottom-Up Approaches

After identifying the body with the appropriate key points, the model uses their coordinates to create 2D or 3D representations.

2D representations map the key points onto a two-dimensional graph, using only the X and Y coordinates for spatial distribution. This type of representation is easier to implement and computationally more efficient. However, limiting the human representation to a 2D image cannot capture the full spectrum of possible movements [20], [26].

This disadvantage can be mitigated using a three-dimensional representation, where a Z-axis is added to the point coordinates. This allows for a greater variety of positions and captures more complex movements and postures [20], [27]. However, this approach is computationally less efficient than the previous one, as it requires more data processing.

The most accurate way to create a three-dimensional representation of the key points monitored on our bodies involves using multi-camera systems or additional sensors, such as Light Detection And Ranging (LiDAR), which uses laser technology to measure the depth of objects from the camera [28]. Depth cameras, such as the Kinect, are also devices capable of creating highly detailed and accurate three-dimensional representations of human bodies [29] [30]. As previously mentioned, these solutions are more expensive and complex than single-camera systems, and they are also less computationally efficient.

OpenPose (OP) [31] and MediaPipe Pose (MPP) [22] are widely used open-source frameworks for Human Pose Estimation that provide practical alternatives to more complex and costly systems. Both primarily rely on simple cameras to return 2D

coordinates of the monitored key points. Both models can process images in real-time and return the coordinates of the captured key points (see Figure 2.1) [22], [32]–[35]. A multi-camera setup can also be natively used with OP to obtain 3D coordinates [36], albeit with some limitations compared to 2D monitoring. Table 2.1 summarizes the key aspects of comparison between the two models.

Table 2.1: Comparison of features between OpenPose and MediaPipe Pose

Feature	**MediaPipe Pose**	**OpenPose**
Performance	Optimised for mobile and web	High performance on desktop
Real-time Capabilities	Designed for real-time use	With Suitable Hardware
Customization	High	Moderate
Pre-trained Models	Various available	Several for different accuracies and speeds
Laguage/Framework	C++, Python, JavaScript	C++, Python
Platform Compability	IOS, Android, Web	Windows, Linux, macOS

For a practical and relatively accessible solution, MPP appears to be a more viable option than OP. Its compatibility with major mobile operating systems and web platforms ensures a broader range of use cases since all modern smartphones have a camera, making it extremely easy to use anywhere. OP, on the other hand, is only compatible with operating systems primarily available on desktop and laptop computers, resulting in a dependency on the location where the system is set up. Furthermore, direct tests with both models showed that MPP achieves significantly faster response times using only Central Processing Unit (CPU) inference, whereas OP has slower response times and heavily relies on Graphics Processing Unit (GPU) usage, increasing hardware investment requirements. Additionally, OP has the disadvantage of producing lower-quality output images.

2.1.3 Gamification in Healthcare

Gamification involves applying typical game elements and mechanics to non-game contexts to increase interest and motivation, and enhance learning [37]–[39]. In the healthcare field, gamification refers to integrating this strategy into the treatment process of patients [40] to improve their engagement and achieve better results with reduced psychological strain. It leverages the motivational nature of games to make tasks associated with treatment more engaging and enjoyable for patients [41].

Gamification has been increasingly used in many healthcare scenarios; for instance, its techniques have been used to help chronic patients manage their health. Mobile applications with game-like features have led to increased adherence to treatment protocols, symptom tracking, and support for maintaining a healthy lifestyle [42]. These applications employ reward systems, avatars, and competition techniques.

Moreover, gamification has also been used for promotion and prevention, particularly among younger people who are more accustomed to these environments [43], [44].

However, challenges must be addressed properly and resolved despite its promising future.

1. Gamification strategies should be tailored to individual preferences and needs rather than assuming a one-size-fits-all solution [45].
2. The entertainment provided by a game or a game-based strategy must be balanced with the therapeutic goals the method aims to achieve [45].
3. It is crucial to foster long-term user engagement and not settle for the initial enthusiasm the proposed solution may generate [42].
4. It is essential to ensure the security and privacy of data in gamified health applications [42].

When properly applied, gamification offers undeniable benefits regarding the type of treatment and its adaptability to the patient or user. This technique is expected to be increasingly used in treatment and prevention contexts, as well as in health and well-being prevention and education. However, such solutions must be developed carefully, considering both the benefits and challenges to maximize their positive impact on patients' health and well-being [46].

2.2 Related Work

To identify and evaluate the gaps and strengths in the existing state of the art, the solutions in the areas targeted by this project were analyzed. These include Home/Virtual Gyms, Exercise Analysis, and Gamification Strategies. Based on this analysis, it will be possible to identify opportunities to develop innovative solutions that overcome the limitations of existing systems and provide a more personalized and effective user experience.

2.2.1 Home Gyms/Virtual Gyms

Technological advancements are transforming traditional gyms into digital platforms. Nowadays, several mobile apps offer the possibility of accessing training plans, such as the Peloton App [47] and Nike Training Club [48]. It is even common to have our training plan provided by a personal trainer available on our smartphone. When working out in a gym, machines are commonly used to perform various exercises. However, training exclusively using one's body weight is also possible. This type of workout is called "calisthenics" and is a common exercise choice for those who choose to train at home, as gym machines can be relatively expensive. Since the COVID-19 pandemic, many people have started adopting the home gym system and viewing their homes as places where they can stay active [49], The interest in exercising at home, combined with technological advancements, creates an interesting opportunity to develop solutions that integrate both.

Maccarone [50] conducted a study over 10 years that demonstrated the effectiveness of a Home-based Full-Body In-bed Gym exercise program. This program consisted of a protocol of 10 exercises, performed three times a week, for a period of two months. The workouts were closely monitored to ensure they were executed correctly, with no adverse effects recorded among the 22 elderly individuals in the study. Although physical well-being did not show a statistically significant change, most participants reported noticeable improvements in their mental well-being. These results demonstrate the potential of incorporating physical activity into the lives of older adults to improve their quality of life through simple exercises, as the one shown in Figure 2.5.

Figure 2.5: Abdominal exercise from Maccarone's study [51]

More modern approaches leverage Augmented Reality (AR) to try to develop a genuinely alternative solution. The study presented in [52] evaluated the possibility of improving performance during training through visual feedback displayed by augmented reality glasses, for example, indicating how far a leg should go during a Mountain Climber exercise. In general, the sample of users who participated in the study found the visual feedback to be useful, motivating, and important for both cognitive and practical tasks. The study's results were also positive; participants adjusted their movements during the workout, allowing them to make the necessary corrections for proper exercise execution. However, it was found that the use of augmented reality feedback increased the cognitive load of the exercise, as there were reports of psychological discomfort caused by the feedback displayed on the virtual body. Additionally, the physical pain caused by the virtual reality headset on the head during the workout was noted as a negative factor.

The Sword Health company created "Phoenix", a product capable of monitoring movement using pose estimation. It consists of an AI-powered care specialist that offers natural conversation with the user, precise correction of movements and postures during

workouts, and clinical analysis to help people feel better from home [53]. Phoenix provides real-time visual and auditory feedback to help users perform exercises more effectively. The company also demonstrated that users in their digital treatment program recover better and faster after surgery than patients receiving in-person treatment [54]. The group of users who used Sword's device and trained on their own initiative ended up having an average treatment time longer than scheduled for the group receiving in-person follow-up, demonstrating increased motivation when using the digital device in this case study.

We can conclude that feedback during workouts helps motivate users and improves their performance during training. Additionally, we can observe that treating people, including active ageing, is possible from home and can be as beneficial, or even more so, than treatment with an in-person specialist.

2.2.2 Exercise Analysis

For summary purposes, "Exercise Analysis" is understood as the aggregation of topics related to posture and movement correction and feedback, the qualification of movement execution, and the counting of repetitions of identical movements (exercises).

Posture is usually defined as the relative alignment of different body parts. Posture correction typically involves keeping the spine straight, maintaining its natural curvature, and addressing deviations from this ideal alignment [55]. Proper posture minimizes strain on the human body, helping distribute weight across all muscles [56]. This, in turn, prevents and treats spinal injuries and deformities that, in older adults, are commonly associated with an increased risk of falls [56]. For this reason, maintaining correct posture during physical exercise is paramount.

The use of pose estimation models for posture analysis during exercises has garnered considerable interest recently, as evidenced by the number of publications and works in this field. While a greater number of projects focused on exercise classifiers, articles also address prototypes for real-time posture correction.

Rahmadani [57] proposed a posture correction prototype that achieved an accuracy of 90.62% using OP as the HPE model and a Support-Vector Machine (SVM) as the classifier. However, the feedback provided by the model is limited to classifying the position of each relevant joint as correct or incorrect, without offering any guidance that could help the user achieve proper execution. This issue persists in other projects [58], [59], where feedback is also limited to requesting the correction of a specific part of the body.

We can observe in the analyzed related work references that there is indeed a significant limitation in the type of feedback provided, which mainly consists of pointing out the error and rarely describes what should be done to correct it [60]. One example

is the Form Check: Exercise Posture Correction Application [61], which uses incorrect examples accompanied by their solutions to classify the type of error being made. Notably, many of the analyzed articles used the COCO [62] and/or MPII [63] datasets to train the feedback models.

Different approaches are used for counting repetitions, varying with the context they aim to evaluate. In one system, whose goal is to classify different types of exercises, the counting is treated as a regression problem using ResNet18 and a Fully Connected Network (FCN) with four layers of neuron sizes 128, 28, 7, and 1, where the activation function is placed at the end of the network [64]. Another approach, applied to the context of Weightlifting, considers the position of the shoulders. To perform a lift, the shoulders need to change their vertical position. The counting is done when the vertical position of a shoulder reaches a single high point after a low point [65]. The reverse is also considered to remove fluctuations at the extreme points.

Different approaches are possible for various functionalities, and it will be necessary to evaluate which ones are most suitable for the development that will be carried out.

In terms of the qualification of the executed movements, no relevant studies assign a quantitative score to the execution of an exercise.

2.2.3 Gamification Strategies

In the context of gyms and physical activity, gamification has been shown in several studies to effectively motivate users to exercise, even encouraging them to take more initiative in engaging with physical activity [66]. Gamification strategies can vary significantly depending on the objectives and target audience.

In Feng Li's research [66], it was concluded that the most commonly used gamification strategies are achievement-based and progress-oriented, involving goals, rewards, and points. It was also found that the second most common strategies are social-oriented, including competition and collaboration among individuals. However, this second approach showed better results when participants, such as family members, were acquainted rather than strangers.

Gamification strategies have also proven effective in combating sedentary behavior [66], with a trend of people becoming more active outside the scope of the studies themselves. Furthermore, it was observed that adopting multiple gamification strategies is more effective than just one and that theoretically guided gamification yields better results.

We can observe the conclusions of Feng Li's study reflected in concrete research, such as Estgren's [67], which examines the implementation of gamification strategies in a gym application, focusing specifically on reward and points systems. The aim was to investigate how these gamified features affected users' motivation, engagement,

and attendance. The main gamification strategies employed included progress bars, an experience points system, rewards, and visual animations. The results showed a significant increase in user motivation and engagement when the gamified version of the application was used, compared to the non-gamified version. However, once again, no significant differences were found when these metrics were analyzed individually.

Directed toward an older audience, Kappen and coworkers' study [68] explored the use of gamification strategies over an 8-week period. These strategies included defining objectives in the form of tasks, setting daily challenges to promote a consistent routine, using progress trackers, and implementing points and badges. Participants selected for this study were required to be already active, as the focus was on understanding how gamification could influence the lives of older adults who were not sedentary.

The participant sample was divided into three groups:

- Group 1: used a gamified fitness application.
- Group 2: used a pedometer.
- Group 3: served as the control group.

The study conducted an exhaustive investigation, constantly considering the opinions and preferences of the groups regarding various aspects, such as outdoor activity preferences and individual goals, including appearance and mental well-being.

While the study's results were divided into numerous categories and did not support a single practical conclusion, the primary takeaway was that gamification elements can and should be customized to meet users' specific needs and health conditions. This customization is critical in fostering motivation and encouraging older adults to engage in physical activity more willingly and enjoyably.

2.2.4 Critical Analysis

After analyzing a sample of existing work in the dissertation's focus areas, conclusions can be drawn to develop a competent and efficient solution that meets the established objectives and overcomes the proposed challenges.

Implementing a home gym system based on pose estimation appears to be the most suitable approach, as it does not require significant additional investment, such as that needed for an AR solution, and can provide feedback on exercise execution without causing physical or psychological discomfort, as reported during the use of AR glasses [52]. Furthermore, this pose-estimation-based approach has proven highly viable and efficient [53].

Regarding Exercise Analysis, the current state of the art presents a significant opportunity to invest in a descriptive feedback system that genuinely helps users understand what they are doing incorrectly and how to adjust their body posture to

improve exercise execution. Regarding repetition counting, as evidence, it is essential to employ a method that is effective within the context of this project, given that different approaches can be utilized, necessitating an evaluation to identify the most suitable one.

Current solutions do not assess the quality of physical activity performed during training, requiring the development of a solution from scratch. The closest existing results to something of this nature are, for example, the confidence level with which a classifier identifies the type of exercise being performed, or using a probabilistic binary classifier that returns the confidence in whether the exercise is being executed correctly. Other approaches could consider the relative distance of joint angles from the correct range.

Based on the demonstrated results in previous studies, we can consider adopting gamification strategies, such as progress tracking, badges, challenges, and rewards, as a viable choice. Additionally, the social factor appears to be crucial to the success of the implementation, as it enables motivating multiple people simultaneously without the need to add new strategies to compensate for this lack. An implementation that considers the individual nature of each user and allows some level of customization tailored to their specific needs also seems to be an excellent option, given the impact this approach can have.

CHAPTER 3

From Personas to Requirements

To enable a user-centered requirements-gathering process, several Personas and scenarios were developed during the project's initial phase, and requirements were derived from these scenarios. This section presents the three stages, beginning with the Personas and concluding with the requirements.

3.1 Personas

The most relevant Persona for creating the scenarios and the requirements elicitation process was António Vidal, a 70-year-old retired man who enjoys staying active and wishes to remain so.

Although the development was centered on this Persona, a second, Inês Guimarães, a 27-year-old social media manager, was included in the document to highlight the system's potential to accommodate users of all ages.

Persona 1 - António Vidal

Age: 70
Gender: Male
Occupation: Retired

Background:

António is 70 years old, married, and lives in Aveiro with his wife. He has always been a very active person. When he didn't walk to his job as an architect, he enjoyed going for evening walks with his wife or, when alone, he liked going to the pool for a few laps. With retirement, António began to notice that his body was also asking for more rest, which started to limit the number of times he went out to exercise. After retiring, António realized that he needed to establish a new routine to stay active and avoid the trap of letting his body become inactive. He wanted to maintain a healthy lifestyle so he could play with his grandchildren and occasionally return to the activities he had always enjoyed.

Core Needs:

Like many people his age, António occasionally faces limitations such as knee and back pain. When these pains arise, he needs to adjust his functional exercise routine to focus on recovery and strengthening the affected areas. In addition, António is always careful to ensure he performs the exercises correctly, avoiding the creation of new problems while trying to solve others.

Goals:

António wants to stay physically active, maintain the ability to perform daily activities without assistance, and continue enjoying leisure moments with his friends and family.

Persona 2 - Inês Guimarães

Age: 27
Gender: Female
Occupation: Social Media Manager

Background:
Inês is an independent young woman, passionate about her work. When she started her job as a social media manager, she quickly realized she would spend a lot of time in front of the computer and would need to stay active in some way if she wanted to maintain a healthy lifestyle, so she joined a gym.

Core Needs:
With the arrival of the pandemic, she found herself forced to work from home, with hardly any opportunity to go out. With the temporary closure of her gym, she was left with no choice but to do her exercises at home. Despite her efforts, it was difficult to stay motivated, training alone, stuck at home, and without anyone to guide her.

Goals:
Being alone, Inês often lost motivation and didn't complete her workouts, sometimes even skipping them altogether. To counter this, she started scheduling simultaneous training sessions with her friends so they could encourage one another, creating a sense of group accountability. To add some fun to their workouts, the friends decided it would be interesting to see who could do the most or perform the exercises best from their training plans, turning it into a friendly competition to push each other and improve their personal records.

3.2 Scenarios

Several scenarios were created, including one related to a system-supervised training session and another for gamified aspects. They are presented in the following subsections. For consistency with the document's language, the original content, in Portuguese, was translated into English.

3.2.1 Scenario 1 - Supervised Training Session

Mr. António (Action)
Approaches the exercise area and walks towards the assistant...

Mr. António

I want to do some exercise.

System

Displays a list of available training sessions on the wall.

System

Here are the available training sessions.

Mr. António

I want to do the functional leg workout.

System

Displays two options on the screen/projector: to perform the workout with or without assistance.

System

Would you like to perform the workout with or without assistance?

Mr. António

I want to perform it with assistance.

System

Opens the workout page, displaying all the exercises that comprise it, along with the "Start" and "Go back" options.

Mr. António

Start

System

Displays images captured by the camera in real-time, and over those images, displays a person turning sideways to the camera.

System

Please position yourself sideways to the camera as shown in the image.

Mr. António (Action)

Positions himself as requested by the system.

System

Recognizes the correct posture and changes the image to a person performing the exercise of raising the knee until the thigh is parallel to the ground.

System

Let's get started.

Mr. António (Action)

Tries to perform the displayed exercise.

System

Recognizes that Mr. António attempted the exercise and stops showing the person performing it, displaying a counter with "0/10" in the top right corner.

Mr. António (Action)

Performs the exercise three times.

System

Recognizes that Mr. António performed the exercise three times o exercício de forma má/incorreta and overlays a dynamic illustration of the knee lift he should achieve on top of his real-time knee image.

Mr. António (Action)

Performs the exercise correctly.

System

Changes the counter to "1/10" and plays a success sound.

Mr. António (Action)

Performs the exercise correctly nine more times.

System

Updates the counter after each correctly performed repetition and, at the end, displays a green "completed" symbol over the camera images. Then proceeds to the next exercise, showing a person performing it just as was done for the previous one.

System

Let's start a new exercise.

3.2.2 Scenario 2 - Gamified Unsupervised Training Session

Similar to the scenario described in subsection 3.3.1, Mr. António requests to perform an arms workout, but this time without supervision.

System

Let's get started

System

The system changes the image to a person performing the exercise of raising their arms laterally to shoulder height three times, then stops showing the image and displays a counter with "0/10" in the top right corner.

Mr. António (Action)

Tries to perform the displayed exercise.

System

The system recognizes that Mr. António performed the exercise, and increases the counter to "1/10". Displays a yellow "+5" over the images (from 1 to 10) for 2 seconds, and adds a 5 below the counter.

Mr. António (Action)

Performs the exercise again in the same way.

System

The system recognizes that Mr. António performed the exercise, increases the counter to "2/10", displays a yellow "+5" over the images (from 1 to 10) for 2 seconds, and updates the number below the counter to 10.

Mr. António (Action)

Performs the exercise better than the previous time.

System

The system recognizes that Mr. António performed the exercise, increases the counter to "3/10", displays a green "+8" over the images (from 1 to 10) for 2 seconds, overlays a blue "+1" on top of the "+8" for 1 second, and updates the number below the counter to 19.

Mr. António (Action)

Performs the exercise in the same way.

System

The system recognizes that Mr. António performed the exercise, increases the counter to "4/10," displays a green "+8" over the images (from 1 to 10) for 2 seconds, and updates the number below the counter to 27.

Mr. António (Action)

Performs all the remaining repetitions and exercises until the end of the workout.

System

The system scores each repetition performed by Mr. António according to the quality of his execution. At the end of the workout, it prominently displays "New Record: 356" and plays the audio, "Congratulations, you've improved your personal best for this workout."

3.2.3 Scenario for CHHA testing

This scenario was derived to support the test sessions part of the CHHA project work plan.

Mr. António (Action)

Approaches the exercise area and says to the assistant.

Mr. António

Good morning, assistant. Let's get started with today's session.

System

Good morning, Mr. António. Would you like to connect to the remote server through your 5G network?

Mr. António

Yes, please. I want to connect without them knowing who I am or where I am.

System

Your secure session has started. Would you like to begin today's scheduled workout?

Mr. António
Yes, please.

System
Do you accept the monitoring of this workout to improve your performance?

Mr. António
Yes, I do.

System
Would you like to view network performance metrics?

Mr. António
Yes, I would like to view the server's latency and processing time.

System
All set! Let's start the workout! This is the first exercise.

System
The application shows how to perform the first exercise.

Mr. António
Let's start the exercise.

System
The application displays the images captured by the camera and processed by the server, and in the top right corner, shows the current latency and the server's processing time.

3.3 Requirements Derivation Process

The scenarios produced were analyzed in detail, line by line, to elicit the system's requirements. The objective is to identify, at each stage of the interaction with the system, the functional, non-functional, and interaction requirements that emerge from the reported use cases, thereby ensuring development aligned with the established objectives and expectations.

As an example, the following section presents the requirements elicitation carried out for Scenario 1.

3.3.1 Scenario 1 - Supervised Training Session

Mr. António (Action)

Approaches the exercise area and walks towards the assistant...

Mr. António

I want to do some exercise. [Recognize voice commands] [NLU processing] [Understand Portuguese] [Microphone available]

System

Displays a list of available training sessions on the wall. [Multiple training sessions stored] [Listing of training sessions] [Simple and intuitive interface]

System

Here are the available training sessions. [Voice synthesis] [Voice feedback for interface changes]

Mr. António

I want to do the functional leg workout. [Select training session from list]

System

Displays two options on the screen/projector: to perform the workout with or without assistance. [Ability to perform exercises with and without assistance]

System

Would you like to perform the workout with or without assistance?

Mr. António

I want to perform it with assistance. [Select assisted training option]

System

Opens the workout page, displaying all the exercises that comprise it, [Listing of exercises per training session] along with the "Start" and "Go back" options. [The system allows navigation between options]

Mr. António

Start

System

Displays images captured by the camera in real-time, [Capture video image] [Display captured image in real time] [Camera available] and over those images, displays a person turning sideways to the camera. [Overlay animations on captured images] [Storage of guiding animations]

System

Please position yourself sideways to the camera as shown in the image.

Mr. António (Action)

Positions himself as requested by the system.

System

Recognizes the correct posture [Recognize user's posture] and changes the image to a person performing the exercise of raising the knee until the thigh is parallel to the ground.

System

Let's get started.

Mr. António (Action)

Tries to perform the displayed exercise.

System

Recognizes that Mr. António attempted the exercise [Recognize exercise execution] [Fast recognition of executed exercises] and stops showing the person performing it, displaying a counter with "0/10" in the top right corner. [Overlay a counter on captured images]

Mr. António (Action)

Performs the exercise three times.

System

Recognizes that Mr. António performed the exercise three times [Count number of repetitions] o exercício de forma má/incorreta [Recognize faulty exercise execution] and overlays a dynamic illustration of the knee lift he should achieve on top of his real-time knee image. [Recognize position of body parts] [Generate dynamic lines and animations over body parts]

Mr. António (Action)

Performs the exercise correctly.

System

Changes the counter to "1/10" and plays a success sound. [Recognize correct exercise execution] [Update repetition counter]

Mr. António (Action)

Performs the exercise correctly nine more times.

System

Updates the counter after each correctly performed repetition and, at the end, displays a green "completed" symbol over the camera images. Then proceeds to the next exercise, showing a person performing it just as was done for the previous one. [Switch between exercises]

System

Let's start a new exercise.

3.4 Requirements

The identified requirements were categorized as functional, non-functional, and interaction requirements, and are presented in Tables 3.1, 3.3, and 3.5, respectively. Additionally, since this work is being developed within the scope of the CHHA project, the project-specific requirements have been identified and listed in Table 3.7.

Table 3.1: Functional Requirements (FR) derived from scenarios

ID	Priority	Description
FR01	P2	Listing of training sessions
FR02	P1	Ability to perform exercises with and without assistance
FR03	P1	Listing of exercises per training session
FR04	P1	The system allows navigation between options
FR05	P0	Capture video image
FR06	P0	Display captured image in real time
FR07	P0	Overlay animations on captured images
FR08	P0	Recognize user's posture
FR09	P0	Recognize exercise execution
FR10	P1	Overlay a counter on captured images
FR11	P1	Count number of repetitions
FR12	P1	Recognize faulty exercise execution
FR13	P0	Recognize position of body parts
FR14	P0	Generate dynamic lines and animations over body parts
FR15	P0	Recognize correct exercise execution
FR16	P1	Update repetition counter
FR17	P1	Switch between exercises
FR18	P1	Ability to assess exercise execution
FR19	P2	Ability to assess when to award bonus points
FR20	P2	Recognize achievement of goals

Table 3.3: Non-functional Requirements (NFR) derived from scenarios

ID	Priority	Description
NFR01	P0	Microphone available
NFR02	P1	Multiple training sessions stored
NFR03	P0	Simple and intuitive interface
NFR04	P1	Storage of guiding animations
NFR05	P0	Camera available
NFR06	P0	Fast recognition of executed exercises

Table 3.5: Interaction Requirements (IR) derived from scenarios

ID	Priority	Description
IR01	P0	Recognize voice commands
IR02	P0	Natural Language Understanding (NLU) processing
IR03	P0	Understand Portuguese
IR04	P1	Voice synthesis
IR05	P1	Voice feedback for interface changes
IR06	P2	Select training session from list
IR07	P2	Select assisted training option
IR08	P2	Select unassisted training option

Table 3.7: Requirements derived from CHHA project

ID	Priority	Description
CHHA01	P1	Instant 5G network communication
CHHA02	P0	Concealment of real home and user identity using anonymous identifiers
CHHA03	P0	Encrypted communication protocols with external services
CHHA04	P0	Sending client identifiers to the server to replace real identities
CHHA05	P0	Real-time encrypted transmission protocol
CHHA06	P0	Integration of a service to record exercise execution data
CHHA07	P1	Integration of measurement logging service on both server and client
CHHA08	P1	Calculation of measurements related to network performance
CHHA09	P1	Synchronization of server and client clocks
CHHA10	P1	Logging the timestamp of processing result arrival
CHHA11	P0	Sending exercise execution identifiers to the server
CHHA12	P1	Remote processing

CHAPTER 4

System Architecture

This chapter presents the overall system architecture, its main modules, and their interactions. The modules and key components of the system are also described in greater detail, namely the Local Application and the Remote Services. For each, the role of its submodules is explained, along with how they are expected to interact with and depend on one another. The order in which each module is presented is directly linked to its proximity to the end user.

4.1 Architecture

The overal architecture is presented in Fig. 4.1

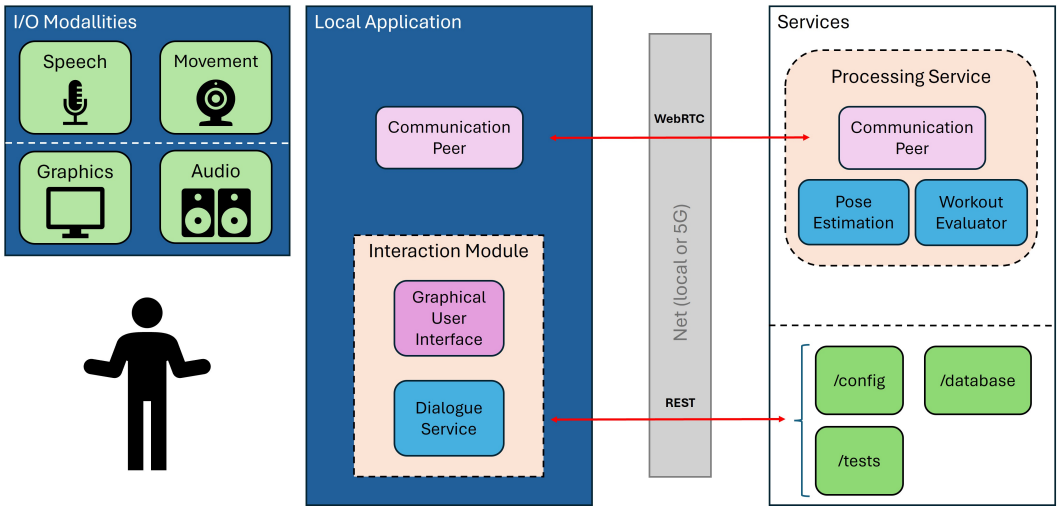

Figure 4.1: Overall system architecture showing its main parts and some of the modules

The proposed architecture for developing this system considered the need to perform tasks that require significantly higher computational power than that of a typical

personal computer. Therefore, these computationally intensive tasks were delegated to a remote environment, a **Processing Service** to which the system's clients connect through a client-server architecture. This service can be deployed either on a local server or within a network facility.

In addition to the **Processing Service** to which the user connects, several modules and services support the overall operation of the system. The modules are integrated within the **Processing Service** itself, such as the Pose Estimation Module (PEM) and Workout Evaluator Module (WEM), and are closely related to the data exchanged directly between the client and the server. There are also other services hosted remotely and accessible from both ends of the system, namely the configuration service, the database service, and the testing service.

The **Local Application** serves as the user's entry point into the system and handles the necessary connections and communications with remote services. It is composed of a communication node, a graphical interface, and a **Dialogue Service**. Communication with the server is carried out both through the communication node and directly by the interaction module itself.

Interaction with the **Local Application** occurs through multiple modalities. The input modalities include speech, used to issue voice commands, and gestures, used to perform exercises. In response to these stimuli, the application provides visual and auditory feedback to inform the user about their actions and requests.

Communication between the application and the **Processing Service** is established through Web Real-time Communication (WebRTC) channels. In contrast, communication with the other remote services is carried out via Representational State Transfer (REST) calls. These communications rely on the standard network infrastructure and are also compatible with 5G connectivity.

To establish WebRTC communications, a third component, the **Signaling Server**, was developed. The **Local Application** and the Processing Server communicate with it via WebSocket (WS), and its sole purpose is to enable direct communication between these two components, as illustrated in the diagram shown in Figure 4.2.

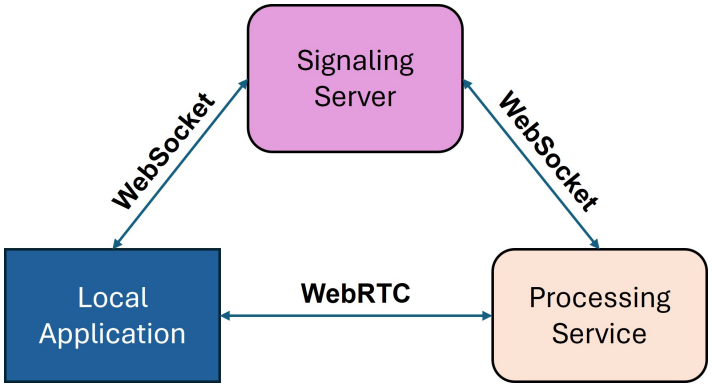

Figure 4.2: Overall signaling architecture.

The following sections provide a more detailed description of each of the two main components of this system, starting with the **Local Application**.

4.2 Local Application

The Local App is presented in more detail in Fig 4.3 and its parts in the following subsections.

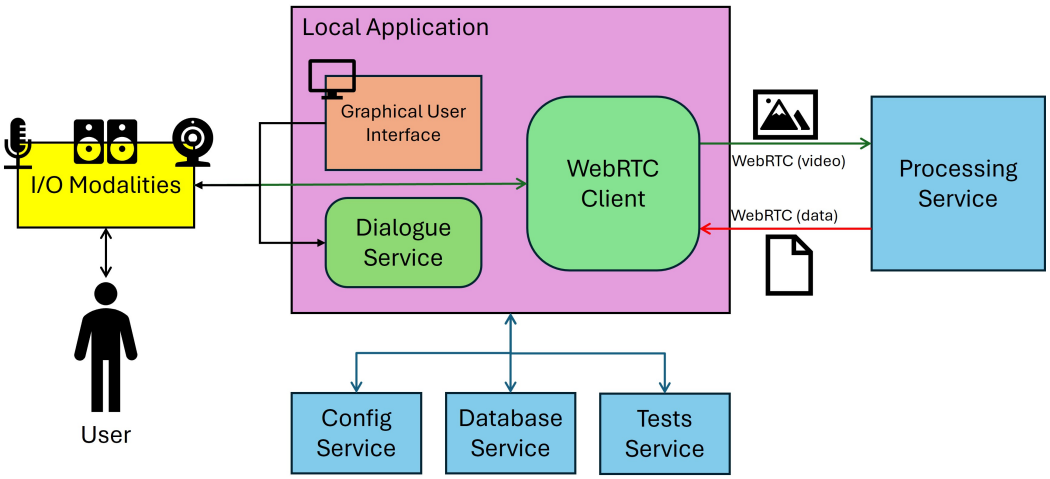

Figure 4.3: Local Application architecture in more detail

4.2.1 System Interaction

Interaction with the system occurs through multiple modalities, with the accepted input modalities being Speech and Movement. The output modalities consist of Audio/Sound and Graphics feedback. To enable full interaction, several hardware devices are employed, allowing comprehensive user interaction through the **Local Application**, which serves as the practical entry point to the system.

These devices are:

- **Microphone:** To capture the user's voice and speech

- **Camera:** To record images of the user for gesture and movement recognition
- **Speakers:** To allow the system to communicate back to the user through the audio modality
- **Display:** To present visual feedback resulting from interaction with the application

Although often unnoticed, interaction with the program is nearly constant, as long as at least one of the four mentioned peripherals is in use.

The Interface is the component of the **Local Application** responsible for all visual feedback generated by the system for the user. However, the Interface does not fully work without the background execution of the **Dialogue Service**, which is responsible for interpreting the user's speech and generating auditory responses and commands based on the user's intentions and actions.

Finally, the Gestures and Movements modality is made possible through the real-time transmission, by the **Local Application**, of the images captured by the system's Camera to the Video **Processing Service**. After processing these images, the service returns a set of data related to them, which is then interpreted according to the current interaction state, enabling the system to react appropriately to the user's body movements.

4.2.2 Video Transmission

The interaction between the **Local Application** and the **Processing Service** is primarily based on the transmission of video and data. The video is captured by the camera connected to the machine running the **Local Application**, and this video stream is sent to the service as a video track. The stream transmission is carried out using the WebRTC protocol, which sends real-time images captured by the camera.

After processing the stream, the **Processing Service** also sends, via WebRTC through a data channel, a set of data related to the analysis of the user's pose. This data allows the interface to render a skeleton over the body detected in the camera images and to evaluate the quality of the exercise execution based on previously transmitted frames, as well as to provide insights into the movement of each body part involved.

4.2.3 Support Services Interaction

The **Local Application** interacts with three additional services besides the **Processing Service**, namely three REST Application Programming Interface (API)s: the Config Service, the Database Service, and the Tests Service.

The Config Service is used during the initial connection phase to the system, exchanging the data required for user authentication and anonymization. The data resulting from this interaction are necessary for subsequent communication with the other REST services.

The Database Service is responsible for storing the user's joint coordinates calculated by the **Processing Service**. Once received, these coordinates are stored in a database along with other relevant data, enabling later analysis of the exercise execution or even the recreation of that execution.

The interaction with the Tests Service by the **Local Application** is limited to sending the timestamps at which a frame is sent to the **Processing Service** and when the results of that processing are returned. Cross-referencing these timestamps with those sent by the **Processing Service** for the same frames allows for an analysis of the network performance and the processing capability of the **Processing Service**.

4.3 Remote Services

As mentioned, the remote services of the architecture are essentially divided into the **Processing Service** and **Support Services**.

4.3.1 Video Processing Service

The details of the **Processing Service** are presented in Figure 4.4.

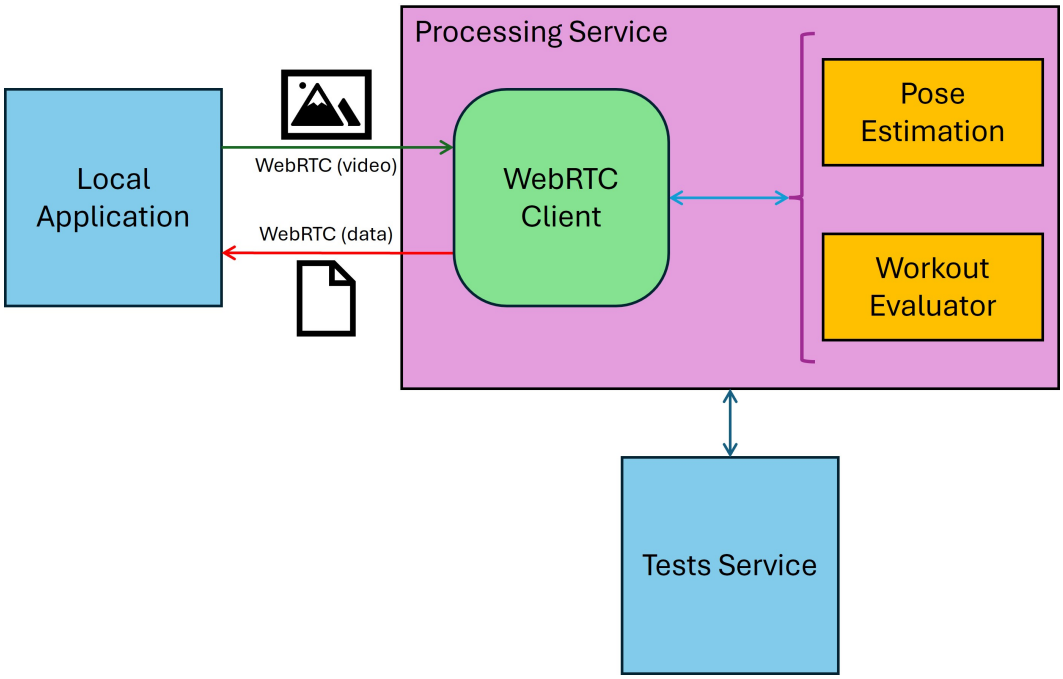

Figure 4.4: Processing Service architecture in more detail

The **Processing Service** operates on a master–worker architecture, where a central access point of the service (the master) connects to the **Signaling Server**, as shown in Figure 4.2. Whenever a new user connects to the system, the master invokes a processing unit (worker). Each invoked unit is responsible for communicating with and exclusively serving the user who requested it.

As previously mentioned in subsection 4.2.2, the task of a processing unit within the **Processing Service** consists of receiving the images captured on the user's side via a WebRTC video track. These images are then automatically directed to the PEM and WEM modules. The PEM is responsible for detecting the skeleton and extracting the coordinates of the user's joints in the analyzed frame, then sending these results to the WEM. The WEM, in turn, uses the previously received results to analyze the progress of the exercise being performed by the user.

Finally, the results obtained by both the PEM and the WEM are sent back to the **Local Application** via a WebRTC data channel, so that the corresponding skeleton and exercise status can be displayed.

In terms of interaction with the **Support Services**, the **Processing Service** communicates with the Tests Service only to report when it receives a frame and when it returns its results, as well as the processing time taken by the PEM.

4.3.2 Support Services

The term **Support Services** refers to the three services in the architecture shown in Figure 4.1 highlighted in green, namely the Config, Database, and Tests Service. The integration of these services arose from the need to satisfy the CHHA Requirements, related to the CHHA project, specifically requirements **CHHA02**, **CHHA04**, **CHHA06**, **CHHA07**, **CHHA08**, **CHHA10**, and **CHHA11** from Table 3.7 The system operates independently of these services and can continue functioning correctly in the event of a failure in any of them, with the only limitation being that the monitoring of metrics and data relevant to CHHA will no longer be possible.

The Config Service, as the name suggests, is the configuration service that allows registering its location and characteristics and returns a unique ID that the **Local Application** must use during its interaction with the other services, so that the metrics can be associated with it without compromising sensitive data.

The Database Service is used to store data related to the execution of exercises, ranging from the coordinates of the user's joints in each frame to additional relevant information, such as the number of repetitions at a given moment, the momentum, age, weight, height, and so on. This data is helpful in that the execution can be recreated through the sequence of coordinates recorded for each frame in the database, allowing, for example, a professional in human movement sciences to assess the user's movements in some manner. Storing the coordinates instead of the video itself not only preserves the user's privacy in case of third-party access but also reduces storage requirements. From a system evolution perspective, this data can also be used to train models for evaluating each exercise.

Finally, the Tests Service was introduced to monitor network performance, taking

into account the CHHA requirements. Temporal data regarding the sending and receiving times of each frame between the **Local Application** and the **Processing Service**, as well as the sending and receiving of the corresponding processing data, is sent to this service. This data is then stored in the service for subsequent analysis.

CHAPTER 5

Developed System

This chapter provides details on the solutions considered for each part of the system and describes how the modules were subsequently implemented. The order in which each component is presented follows the chronological sequence of its implementation, to ensure that no element is introduced before those on which it functionally depends.

5.1 Video Transmission Protocol Selection

The initial focus of the development was, necessarily, the video transmission between the **Local Application** and the **Processing Service**, since all posture and exercise analysis data are obtained by the service that processes the images captured by the **Local Application**. Therefore, in order to meet the requirements of displaying the captured images in real time, along with the processed data, **FR06** and **FR07** from Table 3.1, it was necessary to select a video transmission protocol that prioritizes low latency.

Several protocols were considered for implementing real-time video transmission, namely:

Real-Time Messaging Protocol (RTMP) is a protocol developed by Adobe [69] for the real-time transmission of audio, video, and data over the Internet, maintaining low latency over a Transmission Control Protocol (TCP) connection between client and server [70].

Secure Reliable Transport (SRT) is an open-source streaming protocol engineered for secure and reliable low-latency transmission of audio and video over unpredictable networks, like the public internet. It uses User Datagram Protocol (UDP) as its transport layer and incorporates mechanisms for packet loss recovery,

Advanced Encryption Standard (AES) encryption, and dynamic latency control [71].

HTTP Live Streaming (HLS) is an adaptive bitrate streaming protocol developed by Apple [72] that leverages standard Hypertext Transfer Protocol (HTTP) connections. The video is divided into smaller media segments, enabling scalability and dynamic quality adjustment [73].

WebSocket (WS) is a technology that enables the establishment of a bidirectional communication channel over a single TCP connection. It allows real-time exchange of generic data and keeps the connection open, enabling both endpoints to send and receive messages at any time [74].

Web Real-time Communication (WebRTC) is a real-time communication protocol for the Web that supports the transmission of video, audio, and data between peers. It supports adaptive streaming, Network Address Translation (NAT) traversal, and secure communication through Datagram Transport Layer Security (DTLS) and Secure Real-time Transport Protocol (SRTP) [75].

Several qualitative aspects of different solutions were tested and summarized in Table 5.1.

Table 5.1: Results of the qualitative assessment of the analyzed protocols

	RTMP	**SRT**	**HLS**	**WebSockets**	**WebRTC**
Implementation difficulty	low	low	low	medium	hard
Minimum Delay	> 3 s	> 2 s	> 5 s	> 100 ms	> 150 ms
Peer-to-peer (P2P)	No	Yes	No	No	Yes
Additional server	Yes	No[1]	Yes	Yes	Yes[2]

After this study and testing phase, it was concluded that video transmission would have to rely on either WS or WebRTC. Although WS were neither designed nor optimized for media streaming, their inclusion among the alternatives emerged as a potential way to circumvent the complexity of implementing WebRTC. However, the tested transmission approach consisted of converting each frame into a string and then sending it to the destination. This process proved inefficient due to the need to decode each string upon reception and the requirement for an intermediary server, which, in practice, adds another network hop that can increase transmission delay, revealing the lack of scalability of this solution. Both the frame rate and the image resolution can easily lead to an increased computational demand for the **Local Application**'s machine, which must be capable of transmitting all frames with minimal delay.

[1] A relay server can be used

[2] Signaling server only needed to establish the connection at the beginning

Solutions such as RTMP, SRT, and HLS, despite their ease of implementation, are clearly unsuitable for real-time video transmission. Even the lowest delay recorded during testing was not sufficient to prevent users from clearly perceiving a mismatch between their actions and what they saw on screen. These solutions would only be viable in contexts where a certain amount of delay is acceptable, such as in a surveillance system.

Therefore, it was concluded that WebRTC would be the most suitable protocol for this scenario. Despite its implementation complexity and the need to develop a signaling server, its focus on real-time video transmission and low latency make it the obvious choice to meet the system's requirements. Its ability to adapt transmission quality to network conditions, built-in encryption, peer-to-peer communication, and widespread use in commercial video calling solutions provides further justification for this choice.

5.2 Client-Server Communication

To implement the client-server architecture proposed in Figure 4.1 and establish WebRTC communications between the system components, a signaling server was implemented to meet a previously identified requirement. The purpose of this server is to enable communication between two endpoints that, initially, do not know each other's location but intend to establish a connection, given that the server has a known and accessible Internet Protocol (IP) address. This **Signaling Server** was built on top of a FastAPI WebSocket [76] API with three available communication channels:

- Clients
- Processing Services
- Processing Units

As the name suggests, the client's channel is available for system clients to connect and request a connection to a **Processing Service** that has been previously registered on the **Signaling Server**. Each machine running a **Processing Service** connects to the **Signaling Server** through a dedicated service channel. Typically, there is only one service per machine, as additional instances are unnecessary; however, multiple machines or services can connect to the server, which balances the flow of clients across the registered processing services.

Once registered with the server, a **Processing Service** begins receiving service requests from newly connected clients. However, due to the P2P nature of WebRTC, the process running the service cannot establish multiple simultaneous connections with different clients.

It was precisely to address this limitation that a master–worker architecture was adopted for the processing services. In this setup, the main process, fully implemented in Python, is responsible for communicating with the **Signaling server** and for spawning processing units (processes), which are also entirely implemented in Python. The main service process never directly communicates with the client. When the **Signaling Server** notifies that a processing unit is required for a client, the service spawns a new process that immediately registers itself on the server through the dedicated WS channel for processing units. This process then establishes the WebRTC connection with the client's **Local Application**.

The lifecycle of each Processing Unit process is limited to the interaction with the client that requested it. It is created when the client connects and, under normal conditions, terminates immediately after the client disconnects. The WS communication between the **Signaling Server** and the **Processing Service** is maintained as long as both are running, ensuring that they remain mutually updated regarding the demand for processing units and the termination of those units.

Figure 5.1 presents the sequence diagram illustrating the communication flow between the **Local Application**, the **Signaling Server**, and the **Processing Service**.

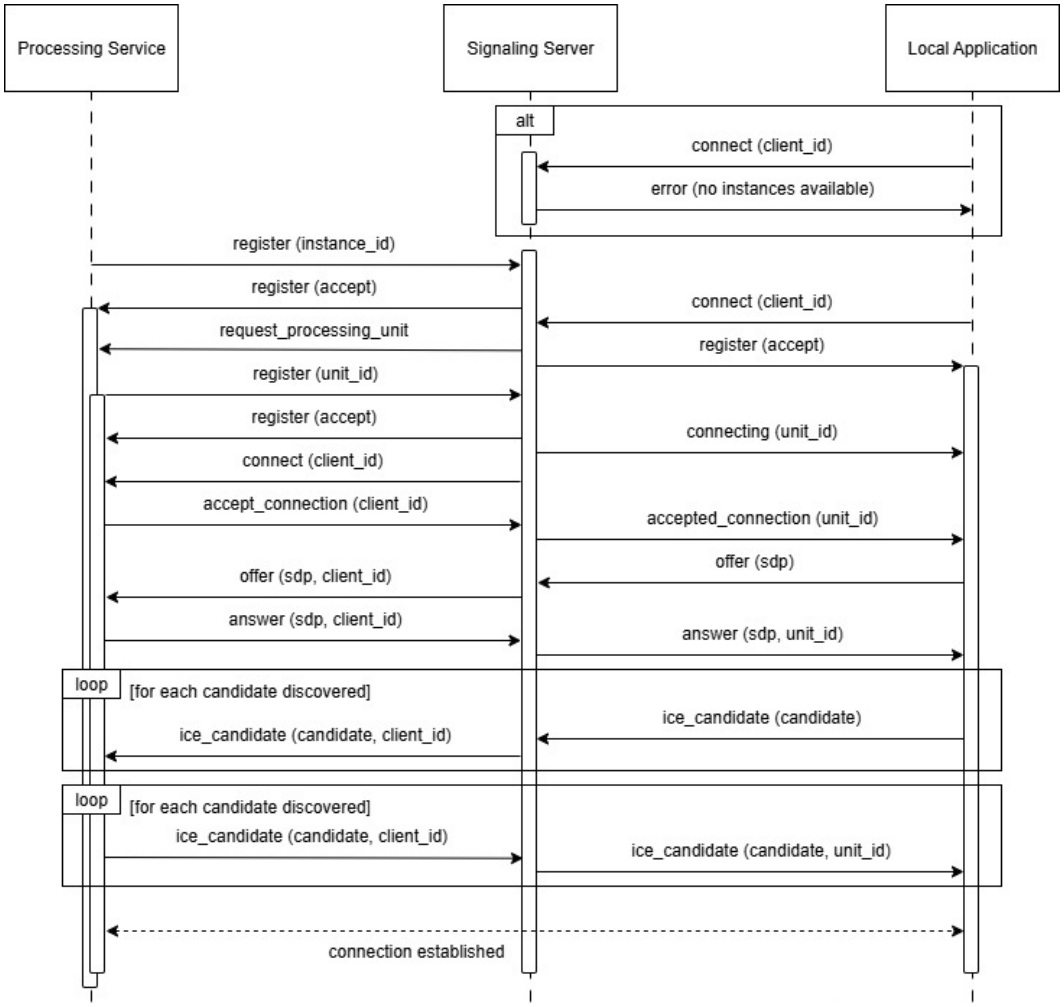

Figure 5.1: Sequence diagram of the communications between the Local Application, Processing Service, and Signaling Server

After the unit is registered, it receives an instruction to accept the connection request, while the client waits for this response. Once the confirmation is received, the client creates a WebRTC connection (offer) Session Description Protocol (SDP), which includes connection details such as identifiers, codecs, media types, and other parameters. The unit then responds with another SDP (answer), specifying the accepted conditions for the connection.

To establish WebRTC communications, the use of Session Traversal Utilities for NAT (STUN) and Traversal Using Relay around NAT (TURN) servers is required. These servers assist the endpoints of a WebRTC connection in discovering their respective public addresses and enable communication by bypassing barriers imposed by NAT and firewalls. Similar to the **Signaling Server**, the STUN/TURN server must be hosted at a known IP address to facilitate access. For this system, a coTURN server was used due to its ease of installation on a Linux machine.

After the offer–answer exchange, each peer relies on these servers to discover

Interactive Connectivity Establishment (ICE) candidates, which are then shared with the other peer until a valid communication path between them is found. Once the exchange of ICE candidates is complete, communication no longer occurs through an intermediary (in this case, the **Signaling Server**), and direct P2P communication channels are established between the two peers. The connection with the **Signaling Server** is maintained only for the exchange of critical messages, such as error notifications from either peer.

Figure 5.2 presents the sequence diagram corresponding to these direct communications between the **Local Application** and a Processing Unit.

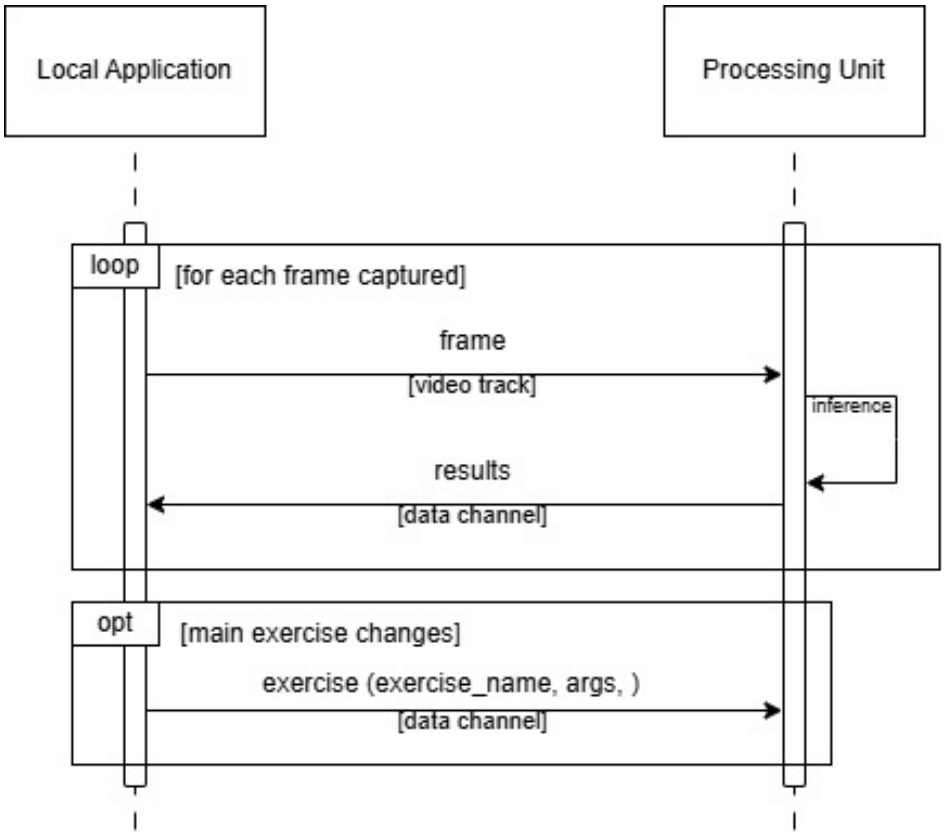

Figure 5.2: Sequence diagram of the direct communications between the Local Application and Processing Service

As indicated in the architecture shown in Figure 4.1, two communication channels are established between the **Local Application** and the Processing Unit, one for video and another for data. The video channel is used to transmit the real-time video stream captured by the camera to the processing unit, while the data channel is used to send back the results of the image processing. Although the video resolution is reduced before transmission due to limitations related to the PEM, the entire video stream dynamically adapts to network conditions to ensure that frames continue to be processed, even if that requires lowering the Frames Per Second (FPS) rate or the frame resolution. Video

transmission through this protocol also ensures user privacy, as all communications are encrypted by default.

5.3 Instantiation of the Pose Estimation Module [Remote]

After the video stream reaches the **Processing Service**, it is immediately redirected to the PEM. As identified in subsection 2.1.2, the most suitable model for pose estimation, and the one implemented in this system, is the MPP.

In the initial development phase, a simple processing strategy was implemented in which the model analyzed each frame as an independent image, always processing only the most recently received frame and ignoring any previous unprocessed frames. This approach prevented the creation of a frame queue and avoided a gradual increase in response delay. However, this solution proved to be highly limiting, both due to the large number of frames that the server could not process in time and the inaccuracy and lack of smoothness in the user's skeleton detection, as the recognition was performed on independent images. Moreover, only the most basic version of the MPP model, the Lite version, could be used due to the even longer inference time from the heavier models. Considering that the delay experienced by the user includes the time required for data transmission to the **Processing Service**, the image processing itself, and the return of the results, it becomes clear that this delay is significantly higher than what would be observed in a typical video call. Therefore, it is crucial to minimize the duration of each stage of this transmission as much as possible.

Given this, the implementation complexity of the solution increased significantly, as concurrency mechanisms had to be introduced to enable the use of Mediapipe Tasks, a more advanced and efficient version of Mediapipe. This version not only allows video stream processing that takes into account the results of previous frames when detecting the skeleton in subsequent ones, but also supports GPU-based processing, thereby reducing pose estimation time. In cases where the heavy model is used with GPU acceleration, which requires greater computational capacity, the inference time difference compared to the lighter model is practically negligible, not exceeding a few tens of milliseconds on mid-range GPUs. When compared to CPU-based inference, using the same system with a GPU results in processing speeds roughly five times faster than those achieved with the CPU. This transition from frame-by-frame to stream-based processing, combined with GPU utilization, not only reduces inference time on the **Processing Service** but also enables the use of heavier and more accurate model variants (Full and Heavy) without perceptible differences in processing time compared to the Lite version, thereby achieving significantly more precise joint estimation.

The mechanism ensuring that only the most recently received frame is processed was

maintained as a safety measure, in case of poor network quality or other unexpected issues. However, in practical terms, it was observed that nearly all received frames are processed, since the inference time per frame is shorter than the typical value of $\frac{1000\,\text{ms}}{\text{FPS RATE}}$ for the transmission.

5.4 Exercise Assessment [Remote]

For the implementation of exercise evaluation, a clear preference was identified for using machine learning models to carry out the tasks proposed in this functionality.

As investigated in subsection 2.2.2, the initial goal of this proposal was to employ machine learning models for movement correction and feedback, execution assessment, to be used by the gamification module, and exercise repetition counting.

However, the lack of pre-trained models for the selected group of exercises, the need to build entirely new datasets, as no existing ones met the project needs, and the necessity to explore different model types to identify those best suited for each task, combined with the shift in focus toward remote processing and Human–Computer Interaction, made this exploration, as well as the implementation of the gamification component itself, unfeasible within the available timeframe.

5.4.1 Implementation of an algorithmic approach

Thus, the solution adopted for implementing the WEM relies on an algorithmic/-geometric approach to the results produced by the PEM. The PEM outputs a list of quintuple values for each frame. Each quintuple consists of the X, Y, and Z coordinates, along with visibility and presence indicators. Each quintuple represents a specific joint, identified by its index in the list. The connections between joints are also already predefined, for example, the thigh corresponds to the connection between the hip and knee quintuples on the same side.

In this approach, exercise evaluation is primarily performed through the analysis of joint angles, the lengths of the connections between them, and their relative positions. Since the MPP can estimate 3D coordinates for the various body joints, these metrics would, in theory, be sufficient to define a set of relatively simple rules to assess, with reasonable accuracy, whether an exercise is being performed correctly, and, if not, to identify which geometric constraints of the movement are being violated. However, when this theory was put into practice, it became evident that the MPP's ability to accurately detect the Z coordinate (depth) from 2D camera input fell far short of expectations, producing depth values that were entirely unsuitable for any meaningful calculation.

The solution devised to partially mitigate the limitation of working solely with 2D coordinates involved a greater incorporation of distances between joints, body parts,

as well as their proportions relative to one another. For instance, when the camera is positioned roughly in front of the user, it can be observed that, in a seated position, the apparent length of the thighs is significantly shorter than that of the lower legs. However, this approach also presents limitations, particularly regarding camera height, the user's angle relative to the camera, and individual body proportions.

5.4.2 Exercises sample

Considering all these constraints, along with the need to define relatively simple exercises suited to the target audience, taking into account their limitations and goals, the exercises selected were as follows:

1. Lateral arm raises, Figure 5.3, 1
2. Alternating front leg raise in a seated position, Figure 5.3, 2
3. Static walking, Figure 5.3, 3

Figure 5.3: Representation of the evaluated exercises (partially generated with Gemini [77])

For the executions of Exercise 1, the arms and torso are evaluated. To be considered correct, an execution must begin from a relaxed position, with the arms perpendicular to the ground plane, and the evaluation starts when an upward movement of the arms is detected. For the repetition to be counted as correct, the arms must be fully extended and parallel to the ground, positioned laterally to the body. The torso must remain upright, which is verified by checking whether the shoulder line is parallel to the hip line and whether the hip–shoulder lines on each side of the body mirror each other. If either the arms or the torso fail to meet the required conditions at the end of a repetition,

that repetition is not counted, and feedback is provided regarding the specific body parts that are incorrect.

In Exercise 2, only the legs are evaluated. This exercise requires the use of a chair so that the user can sit down, and the evaluation begins only when the seated position is detected. This detection is based on the angles formed by the shoulder, hip, and knee triplets, as well as the ratio between the thigh and lower leg lengths. After the initial position is identified, the user must lift one leg until the lower limb is fully extended forward. This verification is performed by analyzing the angle formed at the knee joint and the change in the apparent length of the lower leg. Due to the MPP's depth limitations, the user is instructed to perform the exercise with a slight separation between the legs, allowing for a sufficiently clear lateral view of the leg for accurate analysis.

Finally, Exercise 3 evaluates both the legs and the arms. In this exercise, the user is asked to march in place, lifting one leg at a time while coordinating the movement of the arm opposite to the leg being lifted. Starting from a normal standing position, the evaluation process is designed to allow the user to maintain a steady cadence without requiring exaggerated or unnatural pauses between steps. The arm movement is considered correct if some elbow flexion is detected and the hand of that arm passes along the line of the shoulder on the same side. For the movement of the opposite leg to be considered correct, both the knee and ankle joints must be elevated relative to the corresponding joints of the supporting leg, simultaneously with the execution of the arm movement described above. Whenever any of these conditions are not met, feedback is provided regarding the erroneous movement of the corresponding body part.

In all exercises, the analysis is performed only if the joints required for evaluation are visible at the start of each repetition. If a repetition is deemed correct, in addition to providing feedback on the body parts being assessed, the WEM also returns an indication that the repetition has been successfully counted.

5.5 User Interaction support

By analyzing the scenarios and the system requirements, it is possible to clearly identify the forms of interaction that the user should be able to perform. The following interaction modalities were therefore identified.

- **Input:** Speech
- **Output:** Speech and Graphics

The user should be able to communicate verbally with the **Local Application** and perform body movements to execute each exercise. In response to these stimuli, the system must provide both visual feedback, which the user can see, and auditory

feedback, specifically in the form of speech. In general, the system's feedback, regardless of its type or modality, should be clear and easy to interpret. Thus, for interaction through the **Local Application**, two main components were developed, which must operate in a unified and well-integrated manner: the interface and the agent.

As visual feedback is the most important aspect of the system, its implementation is addressed first, followed by the agent's.

5.5.1 Visual Feedback

The most critical feedback provided by the system is clearly the images captured by the camera and the skeleton that can be drawn on top of those images in real time using the results of the PEM. Initially, a Python OpenCV-based window was implemented to allow visualization of these data (see Figure 5.4). At this stage of the prototype, it was already possible to visualize performance feedback for each body part. Each part appeared in white when detected in a resting position, in green when performing the correct movement, and in red when the movement was erratic.

However, the need to develop a user-friendly interface, suitable for use by non-technical users, quickly led to the replacement of this solution with a more complete interface developed in Next.js (see Figure 5.5).

Figure 5.4: Old Python Open-CV based interface

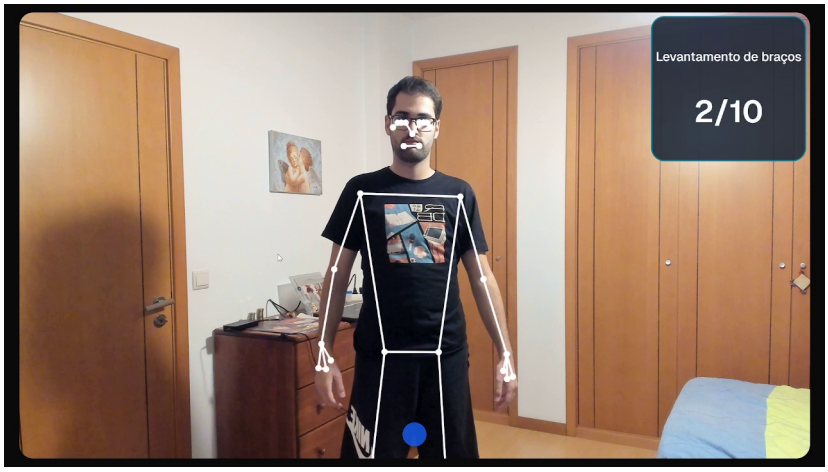

Figure 5.5: Actual Next.js interface

The new interface not only allowed the visualization of the HPE results but also displayed any information deemed necessary for the user's interaction with the system. Such information ranged from guidance on how to fully interact with the application to details about the current state of the exercises, the number of completed repetitions, the target number of repetitions, graphical representations of the performance of different body parts involved in each exercise, demonstration videos, and indicators of the current state of voice interaction, among others.

The ultimate goal of this interface, in addition to displaying the graphical results of the interaction, was also to teach users how to interact with it autonomously, eliminating the need for prior training to understand what to do. The application itself provides this training through practical exercises. Figures 5.6 and 5.7 illustrate examples of this training, showing how the agent's feedback is displayed and how the agent should be invoked, respectively.

Figure 5.6: Landing page with agent's feedback information displayed

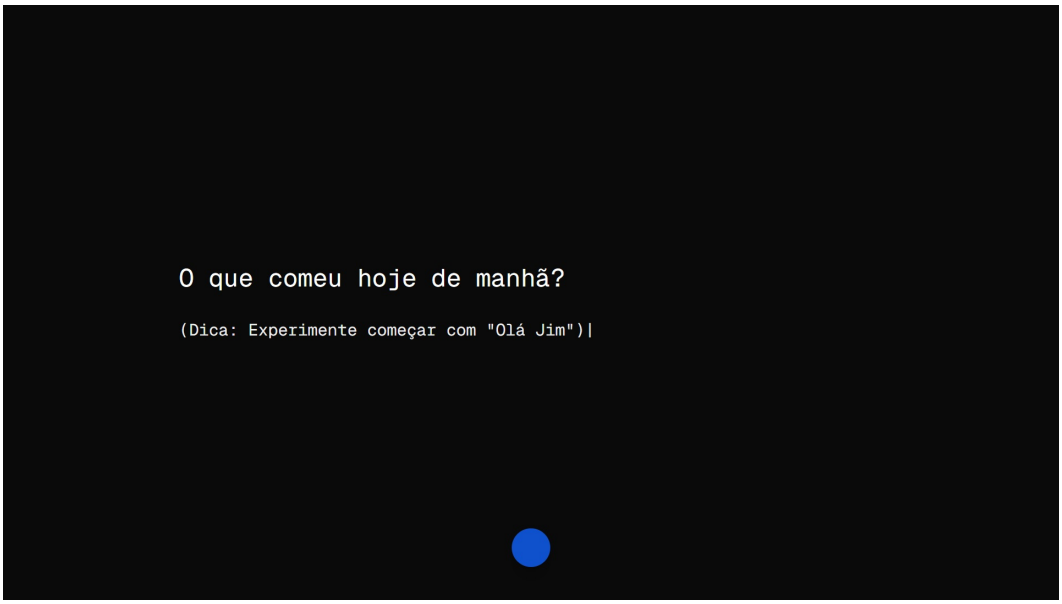

Figure 5.7: Page with a tip for summoning the agent

5.5.2 Spoken Interaction

The agent is, in practice, the entity that interacts with the user through voice; it is responsible for listening to and transcribing the input speech, and responding verbally to the stimuli it receives, whether they originate from speech or from the interface itself.

As illustrated in Figure 4.1, a microphone and speakers are required for the speech modalities. With these hardware devices connected to the system and functioning correctly, interaction with the agent always begins with the wake word *"Olá Jim"*

in Portuguese. This wake word, implemented using the Porcupine [78] engine, when detected, opens the "voice channel" and listens to what the user has to say.

Usually, it is always necessary to say the wake word to interact with the agent, but there are situations in which a request to the agent may require an immediate follow-up response. In these cases, where a quick "Yes" or "No" is typically expected, the voice channel opens immediately after the agent finishes speaking to allow for a more fluid, intuitive, and natural dialogue.

The Speech-to-Text (STT) technology used to transcribe speech into text is the Web Speech API [79], chosen for its ease of integration with the Next.js project, its free usage, and the relatively good results observed. The transcribed speech is sent via WS communication to a locally running **Dialogue Service**, which is essential for the correct operation of the entire application.

This service runs in the background on the machine hosting the **Local Application** and contains a Rasa NLU [80] model, trained with a predefined set of intents. These intents are:

- greet
- goodbye
- affirm
- deny
- start_traing_exercise
- next_exercise
- help
- help_exercise
- presentation

Since the name of each intent is self-explanatory, whenever a user's utterance matches one of these intents, the agent responds, and, depending on the current screen, the graphical interface may also generate a corresponding response.

When the agent is in a state where it wants to listen to the user, a circle located at the center of the screen, usually blue, changes to red and expands, indicating that it is actively listening.

As mentioned, the agent can also respond, in addition to interpreting speech. One of the objectives of the application is for interaction with the agent to feel as natural as possible, as if a human were speaking to the user, so that it is more easily accepted than an agent that appears robotic.

The responses are predefined and typically relate to a user's intent or an interface transition. Some intents have more than one possible reply, with one being selected at random to enhance the User Experience (UX). A brief search was therefore conducted

for Text-to-Speech (TTS) solutions in Portuguese, European dialect, that were free to use and as natural-sounding as possible. This approach enables the generation of synthetic speech from written text almost instantly, rather than pre-recording audio for all predefined responses using a human voice.

Similar to STT, the Web Speech API also provides a free-to-use TTS tool. However, the generated voice gives an extremely artificial impression, making it unsuitable for the desired level of naturalness. More modern solutions, such as the TTS tools from OpenAI [81] and Gemini [82] , proved, after some testing, to be suitable for the intended purpose. The final choice fell on OpenAI's solution, as it allows extended free usage of the service and provides a very good response time under stable network conditions. Once the response audio is generated, the **Dialogue Service** streams it via WS, and it is immediately played back on the interface side.

With a view to evolving towards a Natural Language Generation (NLG) process, the use of a TTS tool also proved to be the better choice, since it would not be possible to pre-record responses with this generation system.

Similar to speech input, when the agent speaks to the user, the same circle changes to green and vibrates according to the frequency of the audio being played.

CHAPTER 6

Results

In this chapter, three types of results from testing and implementation of the gym system are presented. The goal was to improve the system based on the evaluation of each test's outcomes. The presented results are:

- *Tests of the proof-of-concept with 5G (in the scope of CHHA project)*
- *First evaluation with users*
- *Final prototype deployment at Casa Viva+*

6.1 Tests of the proof-of-concept with 5G (CHHA)

The first experiment was conducted on May 9, 2025, at the Instituto de Telecomunicações (University of Aveiro), within the scope of the CHHA project. Its objective was to test the conditions of the available 5G infrastructure and to evaluate the performance of the first development phase of the system — the communication between the **Local Application** and the **Processing Service** — using the simpler initial implementations within the same infrastructure. This infrastructure included a 5G antenna with a 20 MHz bandwidth, which served as the access point to the network hosting the **Processing Service**. The computer running the **Local Application** was equipped with a 5G modem connected via Universal Serial Bus (USB) 3.0 and a Subscriber Identity Module (SIM) card, enabling direct communication between the machine and the antenna, thus avoiding intermediaries.

The webcam used in this experiment, which captured the images transmitted to the **Processing Service**, had a resolution of 1280 horizontal pixels by 720 vertical pixels and a frame rate of 30 FPS. However, due to the previously mentioned limitations of the MPP model of the PEM, and in an attempt to achieve higher transmission speed, the frames were resized to 640 by 480 before being transmitted to the service.

The photo in Figure 6.1 shows the setup used for the **Local Application** in this test. The machine used was a laptop connected via USB 3.0 to the 5G modem, visible to the left of the computer with a pinkish light. The webcam mentioned earlier was the one embedded in the laptop's display frame.

Figure 6.1: Setup of tests performed with 5G network

To perform network performance measurements, six key points were defined, identified in the diagram shown in Figure 6.2.

- **Point A:** Measurement taken immediately before sending the frame, in the **Local Application**
- **Point B:** Measurement taken immediately after receiving a frame, in the **Processing Service**
- **Point C:** Measurement taken immediately before starting the MPP inference process on the received frame
- **Point D:** Measurement taken after obtaining the inference results
- **Point E:** Measurement taken immediately before sending the inferred results back to the **Local Application**
- **Point F:** Measurement taken immediately after receiving those results, in the **Local Application**

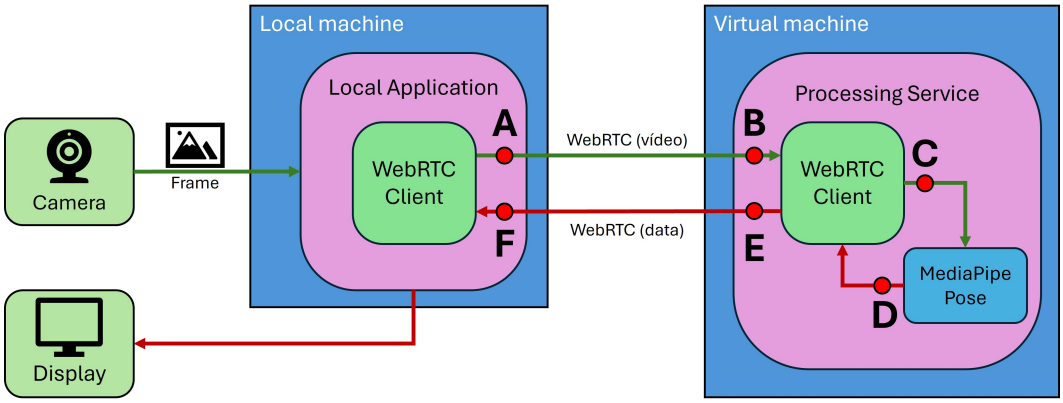

Figure 6.2: Diagram showing the measurement points in the system pipeline

Thanks to the measurements taken at these points, it was possible to determine four critical time intervals.

- **End-to-end Latency**: Calculated as the difference between points **A** and **F**, it represents the total system latency, from the moment the **Local Application** sends a frame until its results are received back in it. In practice, this corresponds to the time delay the user perceives between performing a movement and seeing it reflected in the skeleton overlay on the image.
- **Frame Time**: Measured between points **A** and **B**, it represents the time taken for a frame to travel from the **Local Application** to the **Processing Service**.
- **Inference Time**: Measured between points **C** and **D**, it reflects the inference time required by the MPP model for that frame.
- **Results Time**: Measured between points **E** and **F**, it represents the network time required to send the inference results back to the **Local Application**.

The results of these measurements are shown in Figure 6.3, while the mean, median, and standard deviation values of these four metrics can be found in Table 6.1.

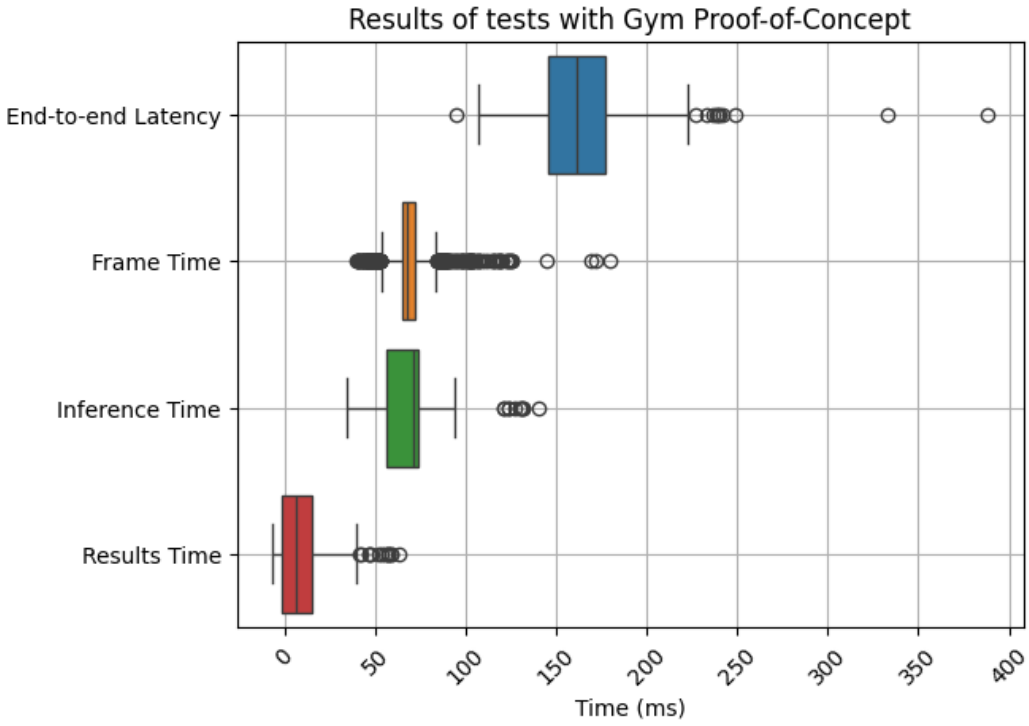

Figure 6.3: Results of the 5G network performance tests for the Gym proof of concept

Table 6.1: Performance metrics for the proof-of-concept pipeline in 5G networks

	End-to-end Latency	**Frame Time**	**Inference Time**	**Results Time**
Mean	163.56	70.27	67.85	7.81
Median	161.06	67.90	70.78	5.94
Standard Deviation	19.15	8.59	8.96	9.24
Minimum	94.93	39.55	34.38	-6.78
Maximum	388.21	179.33	140.47	63.03

During these tests, synchronization issues were observed, with negative time values appearing in the measurements of the **Results Time** metric. These issues are expected, as the setup involved two independent machines running on their own clocks. Unfortunately, the infrastructure lacked hardware synchronization, so synchronization was achieved using Network Time Protocol (NTP) servers via the Python `ntplib` library. This library provides the offset between the **Local Application** machine's clock and the NTP server, which was applied to each recorded timestamp. Although this approach does not achieve the desired level of precision, it was sufficient to obtain reasonably valid values given the measurement order of magnitude. Moreover, the End-to-end Latency and Processing measurements each use the same clock for their start and end timestamps, ensuring these values accurately reflect the actual durations they represent.

To better interpret these results, two preliminary tests were conducted to establish communication between the **Local Application** and the **Processing Service**, with the latter hosted on the same machine in both cases, but using different network infrastructures. These infrastructures were home fiber broadband and commercial 5G in the city of Aveiro, with the relevant results presented in Table 6.2 and Table 6.3, respectively.

Table 6.2: Performance metrics for the proof-of-concept pipeline in fiber networks

	End-to-end Latency	**Frame Time**	**Inference Time**	**Results Time**
Mean	162.94	49.17	72.86	25.62
Median	161.91	47.11	72.72	25.40
Standard Deviation	14.67	5.95	4.11	9.80
Minimum	114.47	35.94	43.26	7.38
Maximum	308.63	83.47	128.05	68.81

Table 6.3: Performance metrics for the proof-of-concept pipeline in commercial 5G networks

	End-to-end Latency	**Frame Time**	**Inference Time**	**Results Time**
Mean	196.69	89.41	56.99	30.78
Median	193.05	86.04	54.75	27.35
Standard Deviation	21.24	13.97	4.83	11.55
Minimum	122.52	52.87	32.68	11.61
Maximum	393.93	215.12	125.49	126.12

Since the machine hosting the **Processing Service** remained the same throughout all experiments, a noticeable variation in **Inference Time** times can be observed when comparing Tables 6.1, 6.2, and 6.3. This variation can be explained by the different testing environments in which the images were captured: the tests in Table 6.1 were conducted in a laboratory, those in Table 6.2 in a home's room, and those in Table 6.3 outdoors. Differences in environment, camera angle, and lighting conditions fully justify the observed variation in **Inference Time** times, which do not represent perceptible differences for the user.

Excluding the **Frame Time** and **Results Time** values due to the previously mentioned synchronization issues, we can compare the **End-to-end Latency** to evaluate system performance across the different networks. It can be observed that the tested 5G infrastructure and the fiber network achieved very similar average latency values. In contrast, the commercial 5G network was slightly slower than the other two. The tested 5G infrastructure also exhibited greater latency variability, as evidenced by its higher **Standard Deviation** and a wider range between maximum and minimum

latency values. This can be attributed to the different propagation media — air for 5G versus optical fiber — with the latter being more consistent. However, this variability did not compromise the system's usability or perceived performance.

In conclusion, both theoretically and practically, the tests demonstrated that the evaluated 5G infrastructure met the minimum requirements for smooth system operation and is suitable as the communication backbone between the system's different components.

For future experiments aiming to reduce end-to-end delay, the priority is to implement a GPU-accelerated version of the MPP model capable of processing a continuous video stream instead of independent frames. This implementation would reduce the model's inference time, which, in practice, is the only component where total response time can still be optimized. The choice of WebRTC has already been justified by its superior responsiveness compared to alternative protocols, so its inherent delays cannot be further reduced.

At the user interaction level, it was also prioritized to develop a more user-friendly interface, enabling a more natural and intuitive interaction with the system.

6.2 First evaluation with users

The demonstration of the CHHA project took place on September 19, 2025, once again at the Instituto de Telecomunicações. The goal of this demonstration was to test the implemented system with real users, who were later presented with a questionnaire to gather feedback on their experience with the system.

The setup shown in Figure 6.4 aimed to replicate a home environment in which the user would have, in their living room, a projection of the **Local Application**, a camera with a microphone to capture both the user's movements and their voice commands, and a speaker system allowing the system to provide audible responses to the stimuli it receives.

Figure 6.4: Setup during the first user's test

A group of users of different genders, age groups, and professional backgrounds was selected to test the developed system. At the end of each test, participants were asked to complete a questionnaire (Table 6.4) consisting of eight Likert-scale questions and one open-ended question, where users could provide additional comments or suggestions about the system.

Questionnaire				
1. How would you rate the responsiveness of the image relative to the real movements?				
1. Very Slow	2. Slow	3. Moderate	4. Fast	5. Very Fast
2. How would you rate the smoothness of the images displayed on the screen?				
1. Very Poor	2. Poor	3. Fair	4. Good	5. Excellent
3. How would you rate the quality of the monitoring of your skeleton?				
1. Very Poor	2. Poor	3. Fair	4. Good	5. Excellent
4. How would you describe the average difficulty of the exercises presented?				
1. Very Easy	2. Easy	3. Moderate	4. Hard	5. Very Hard
5. Do you consider the system to be intuitive?				
1. Not Intuitive	2. Slightly Intuitive	3. Moderately Intuitive	4. Intuitive	5. Very Intuitive
6. Do you think the system helped improve your performance in each exercise?				
1. Strongly Disagree	2. Disagree	3. Neutral	4. Agree	5. Strongly Agree
7. Do you find the information displayed during an exercise easy to understand?				
1. Strongly Disagree	2. Disagree	3. Neutral	4. Agree	5. Strongly Agree
8. In a conversation with someone, how likely are you to mention a positive experience with this system?				
1. Very Unlikely	2. Unlikely	3. Possible	4. Likely	5. Very Likely

Table 6.4: Prototype evaluation questionnaire for the first user's test

The results for each questionnaire item, including the mean and standard deviation, are presented in Figure 6.5.

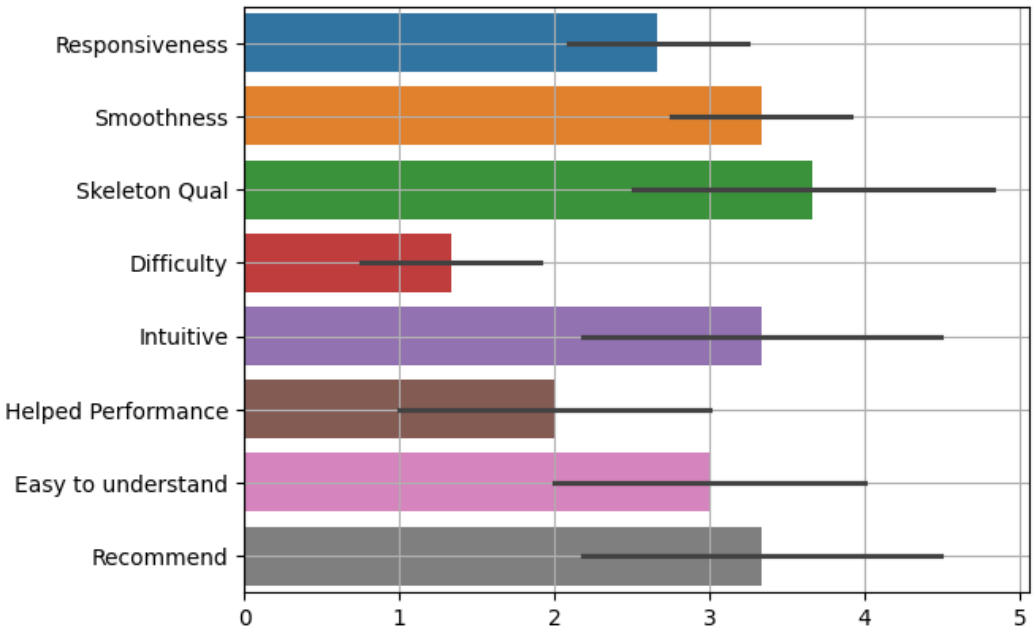

Figure 6.5: Results from the user's testing questionnaire

The first two questions were directly related to the network's performance and its impact on the user experience. It can be observed that the general opinion regarding the image responsiveness, directly related to the end-to-end delay measured in the May 9 test, was that it was neither too slow nor too fast. However, image smoothness yielded better results, showing that despite a slightly noticeable delay in the image, the displayed video remained fluid. One possible explanation for the moderate satisfaction with response time was the evident network degradation throughout the day at the testing facilities, which required reducing the captured FPS rate to avoid overloading the network, whose responsiveness had dropped significantly below normal levels.

The ratings for the quality of skeleton tracking indicate that the choice of MPP models for the PEM was a sound decision. The ratings of 1 and 2 for the question about the difficulty of the exercises also suggest that this choice was appropriate, at least from a difficulty standpoint, as the exercises were intended to be easy to reduce resistance to participation among older age groups.

Despite the generally positive results in the questions about system intuitiveness, ease of understanding instructions, and likelihood of recommending the system to others, there were some more negative responses in these areas, with the usefulness of feedback receiving the lowest rating in the questionnaire. The suggestions provided in the open-ended question, combined with direct observations during the tests, revealed that the interaction model with the system needed to be reviewed and improved.

To interact with the system via voice, users had to say "Olá Jim" to activate the assistant and prompt it to start listening to their commands. At this stage, the **Local**

Application only instructed the user to say "Olá Jim" once to begin the training session, as shown on the screen in Figure 6.6, which did not allow users to properly learn or get used to interacting with the assistant. Additionally, there was no graphical feedback indicating when the assistant was ready to listen, so visual elements only appeared after the user began speaking, that is, after invoking the assistant. It was also observed that, for simpler yes/no questions, users tended to respond immediately after the question was asked, ignoring the wake word "Olá Jim."

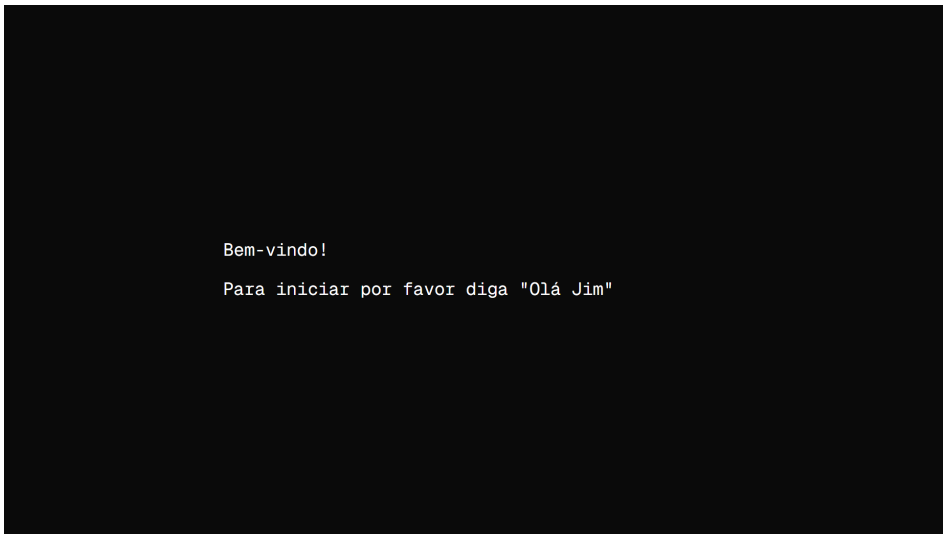

Figure 6.6: Old prototype landing page without tutorial

Following these initial user tests, resolving interaction issues and creating a short tutorial to teach and help users get accustomed to speaking with the assistant before starting the training sessions were identified as top priorities.

6.3 Final prototype deployment at Casa Viva+

As defined in the project objectives, the solution was developed with a real-world deployment in mind, within the scope of the Casa Viva+ project.

6.3.1 Final Prototype

The prototype tested by users, described in the previous section 6.2, was improved with the addition of agent state indicators, an initial interaction tutorial to facilitate learning how to interact with the system, demonstration videos played while the agent explains an exercise, and minor adjustments to the exercise evaluation algorithms.

Before user testing, it was not clear when the agent was listening or whether it was attending to the user continuously. To address these issues, a colored circle representing the assistant and its current state was added at the bottom-center of the screen. These states are illustrated in Figure 6.7.

When the assistant is idle, waiting for interaction or for the wake word “Olá Jim”, the circle remains static and blue. When it is called or asks a question that allows an immediate response, its state changes to red and expands according to the audio frequency levels it is receiving, indicating that it is listening. When it starts speaking, the circle turns green and exhibits the same expanding behavior, but this time based on the audio being played, indicating that it is speaking to the user. This explanation was also added to the system’s landing page, as shown in Figure 6.8.

As previously mentioned, in the tested version, the wake word had to be said every time for the assistant to start listening. This was slightly modified to a mechanism where, immediately after the assistant asks a question, it starts listening automatically, allowing the user to respond right away and making the interaction more natural.

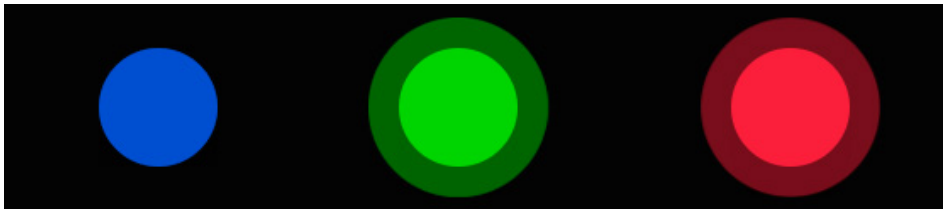

Figure 6.7: Visual representation of the assistant’s possible states

In terms of learning how to interact with the system, the reality was that there was no formal instruction on how to communicate with the assistant, apart from the previously shown old landing page in Figure 6.6, which was relatively brief. Therefore, several pages dedicated exclusively to interaction were added before the start of the training.

To the landing page, information about the possible states of the assistant was added, as shown in Figure 6.8, where each colored circle can be read, from left to right, as “Speaking,” “Waiting,” and “Listening.”

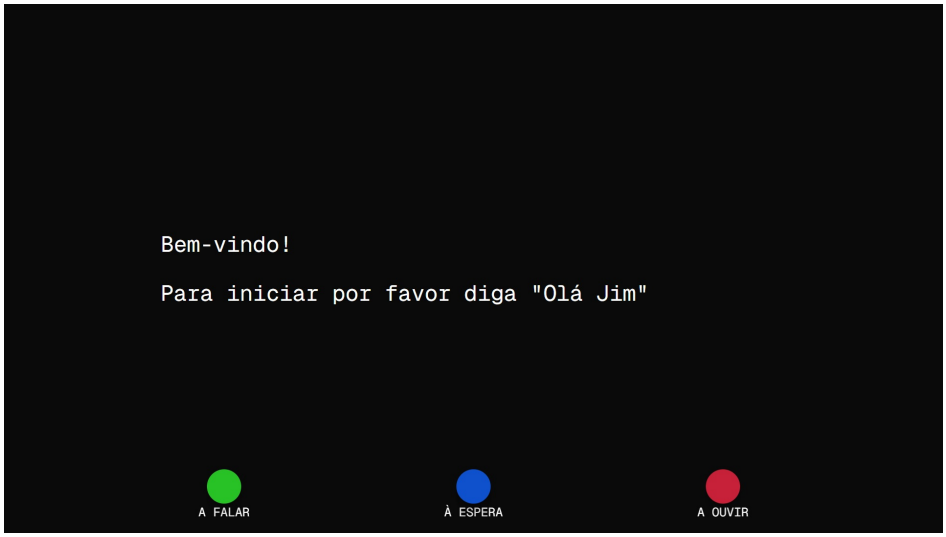

Figure 6.8: Actual prototype landing page with assistant information

After moving to the next screen, triggered by using the wake word, a new message appears that complements the instructions the assistant provides via audio, prompting the user to repeat "Olá Jim", as shown in Figure 6.9. This encourages the user to use the wake word again, helping them become familiar with the process.

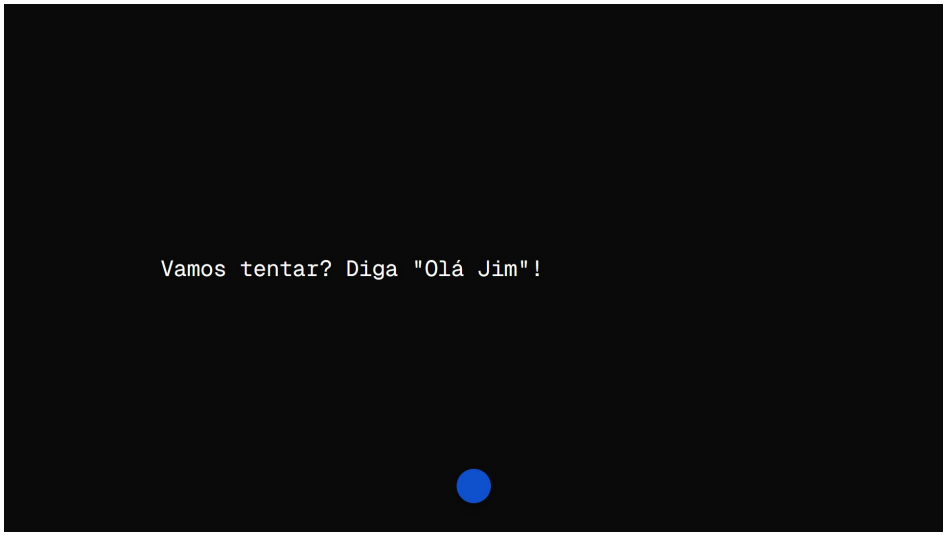

Figure 6.9: New screen to encourage the use of the wake word

Following this screen, an additional level of interaction is introduced. The subsequent screens, presented in Figure 6.10, appear sequentially and present open-ended questions, requiring the user to provide a response. At this stage, however, the assistant is configured to listen again only after the wake word has been detected. The intention is for the user to start with "Olá Jim" and then, upon seeing the assistant switch to a listening state, immediately provide their response. If the user forgets to use the wake

word, a reminder message appears on the screen after 30 seconds without hearing it, as shown in Figure 5.7.

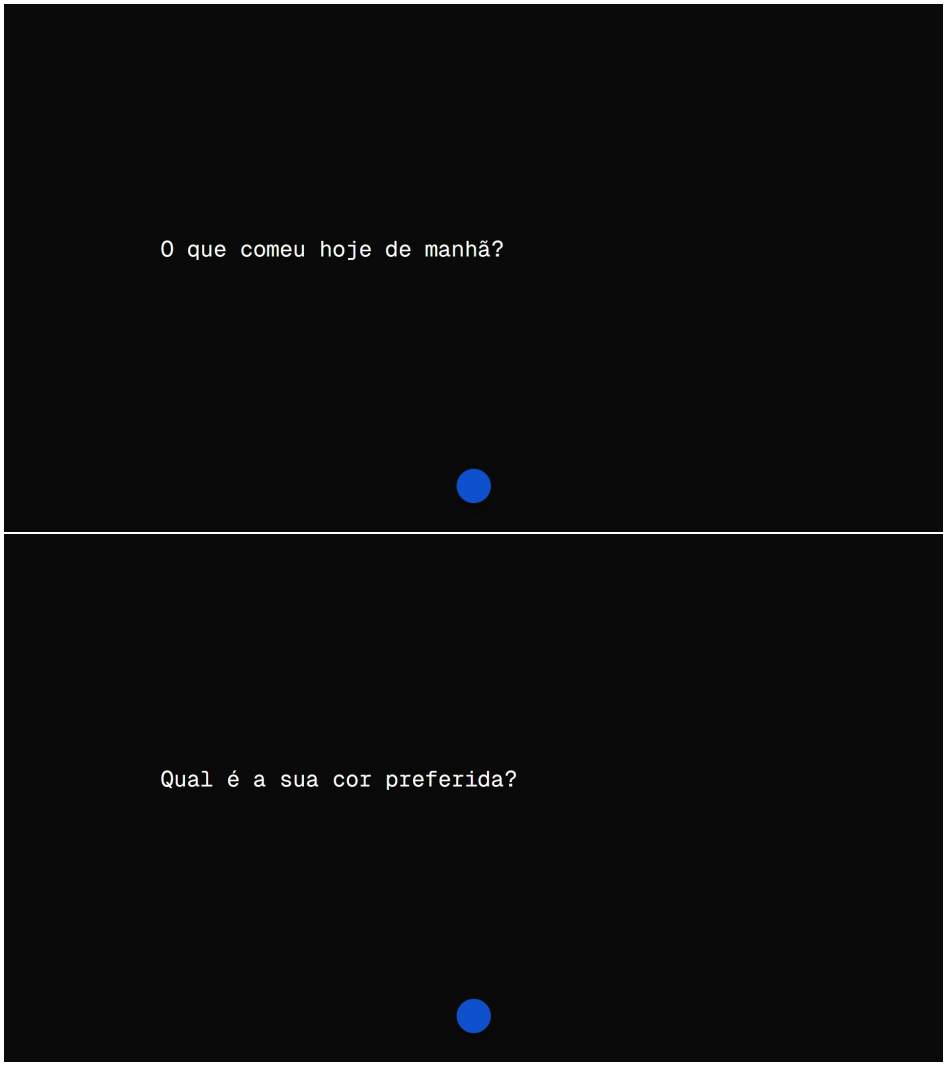

Figure 6.10: Screens designed to guide the user through a complete interaction

To conclude the tutorial, a penultimate screen appears, instructing the user to notify the assistant when they are ready to start the training session, as shown in Figure 6.11.

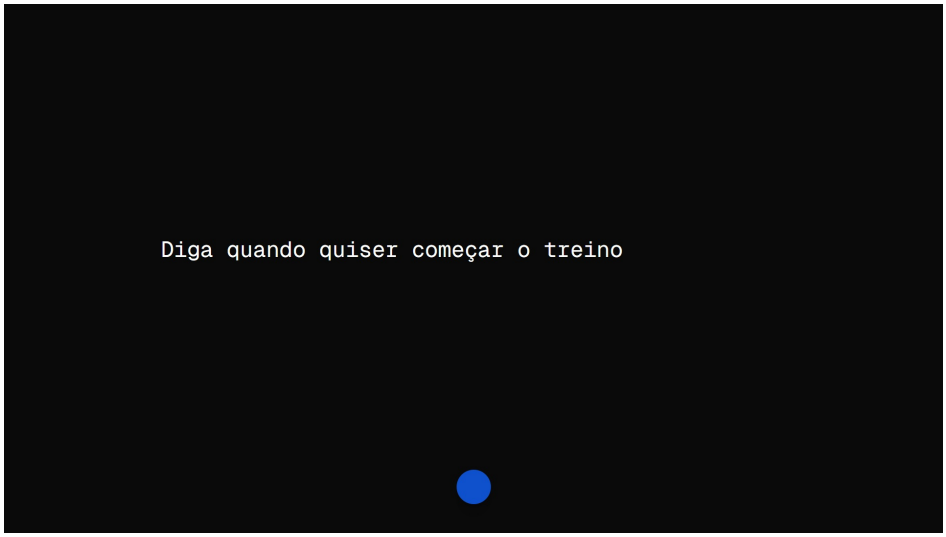

Figure 6.11: Screen asking to start the training session to induce an immediate response

The purpose of this screen is complemented by the following one, shown in Figure 6.12. When indicating that they want to start the training session, both the screen and the assistant's speech present a Yes/No question asking whether the user is sure they want to begin. As previously explained, in situations like this, the assistant is programmed to start listening immediately. The goal of this final phase of the tutorial is precisely to demonstrate to the user that they can respond without using the wake word.

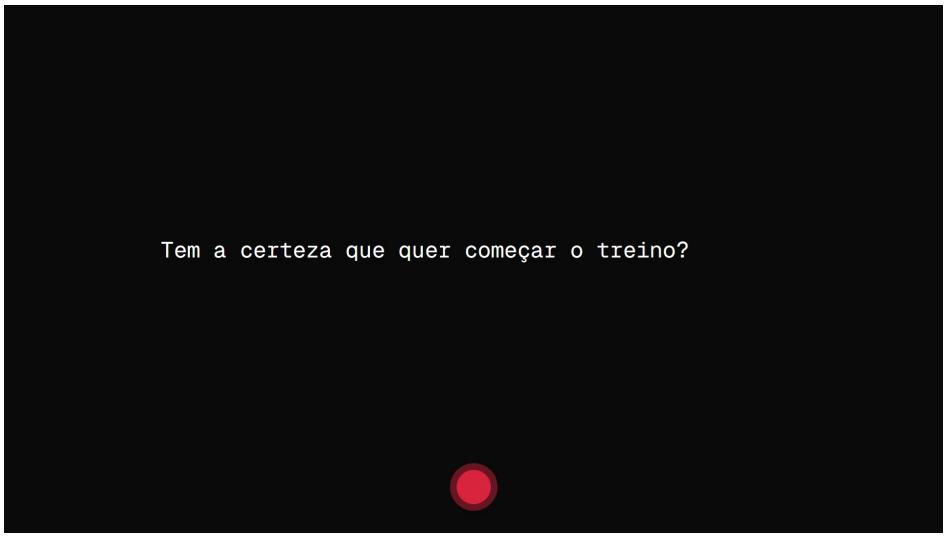

Figure 6.12: Screen conducive to immediate response

Finally, immediately after the session begins, the assistant verbally explains the exercise that the user is expected to perform. Understandably, a purely verbal explanation can be confusing, especially for more complex exercises. Therefore, to complement the audio explanation, a demonstration video appears on the screen, as shown in Figure 6.13,

and remains visible only until the user gives the command actually to start performing the exercise.

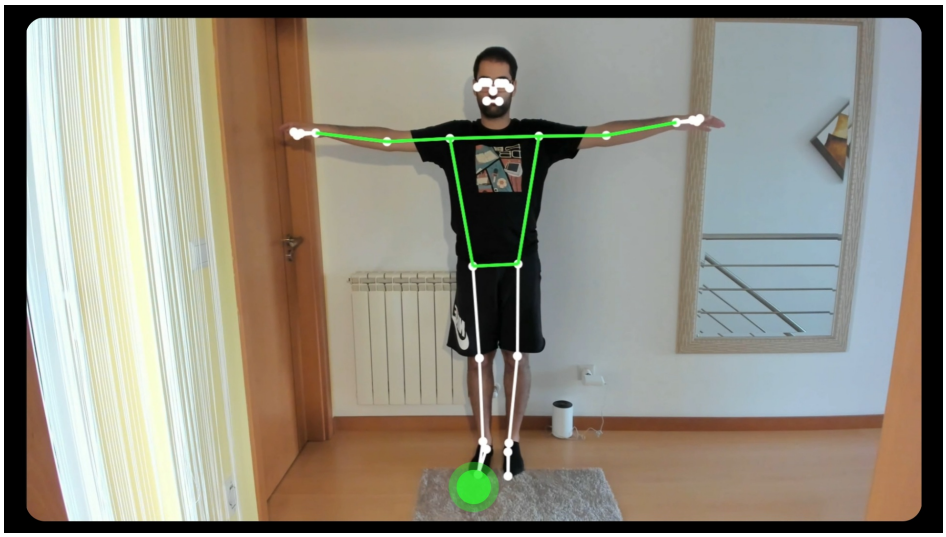

Figure 6.13: Demonstration video shown during the audio explanation

6.3.2 Deployment Process

This deployment began in mid-June and was completed during October at the house built in the scope of project Casa Viva+ at the Centro de Medicina de Reabilitação da Região Centro-Rovisco Pais, at Tocha (Figure 6.14). During the deployment process, several visits were conducted to install a physical server to host the project's remote services, including the **Processing Service**, the **Signaling Server**, and the **Support Services**; perform network tests; install the gym setup; determine the optimal positioning of the projector and camera; and carry out tests for video transmission and remote processing.

Figure 6.14: Photo of the CasaViva+ project house in the Centro Rovisco Pais

To run the **Local Application**, a Mini-PC was used, connected to the same projector and camera employed during the user testing described in section 6.2.

The final setup can be seen in the photo in Figure 6.15, where the following components are identified: number 1 identifies the camera responsible for image capture, an Insta360 Link 2, capable of recording in 4K@30 FPS and equipped with an integrated microphone used for voice interaction; number 2 corresponds to the projector that displays the graphical interface, a Hisense PL2, also with 4K resolution, and built-in speakers used to play the assistant's responses; number 3 indicates the Mini-PC, a Dell Pro Micro QCM1250 with an Intel(R) Core(TM) Ultra 7 265T processor, 32 GB (2×16 GB) DDR5 RAM, 1 TB SSD, and Wi-Fi 6E AX211, where both the **Local Application** and the **Dialogue Service** are executed.

This Mini-PC is visible in the image but can be hidden for a more minimalist visual setup. Finally, number 4 represents the projection area, where the projector displays the graphical interface on a projection screen.

Communication between the two machines was made possible through the setup of a private internet network, established via a 5G router with access to the global internet.

Figure 6.15: Local application setup at the CasaViva+ project site

With this installation, it was possible to test the operation of the new system version in the environment for which it was designed. An example of such tests is shown in figure Figure 6.16.

Figure 6.16: Prototype testing at the CasaViva+ project site

CHAPTER 7

Conclusion

This chapter concludes the document by summarizing the main phases of the work, highlighting the key results achieved, and providing guidance for the continuation of the work.

7.1 Summary of the Dissertation Support Work

The work comprised, in this order, the research, design, development, and evaluation phases, structured according to the Engineering Research method and the User Centered Design (UCD) methodology. The main tasks that supported the development of the final prototype were:

Related work search – To understand the existing work in the areas covered by this dissertation, research was conducted on three of the main topics initially planned: home/virtual gym systems, exercise analysis, and gamification strategies. This research revealed that the combination of these three areas has not been thoroughly explored, thereby opening up a wide range of possibilities for further investigation, particularly in the use of Pose Estimation.

Personas and Scenarios – Based on the project objectives and the findings from the background and related work review, two Personas were developed to represent the different types of users who might be interested in such a solution. To describe real user interactions with the system, a set of scenarios was created to help identify the requirements to be fulfilled.

Initial Requirements – From the scenario analysis, it was possible to derive the functional, non-functional, interaction, and CHHA-project-specific requirements, as well as determine the priority of each in the selection and implementation process. This list of requirements enabled the identification of the initial development phases and, consequently, also influenced the design of the system architecture.

Architecture definition – Although initially guided by the identified requirements, the prototype's architecture evolved continuously, mainly due to the needs and limitations of the technologies encountered during implementation. Defining an architecture enabled the development to be divided into well-defined modules, which were essential for implementing the first, more basic prototype.

Initial Development and First Tests – The first development phase focused on implementing an initial version of the Local Application and the Processing Service, which serve as the foundation for the final system. This version enabled us to implement the communication mechanism and conduct the first tests over a 5G network, allowing us to compare and analyze the system's performance across both the 5G networks and the fiber network.

Refinement after Initial Tests – After the first round of tests, it became clear that the image transmission and inference process needed to be made as efficient as possible. Significant changes were made to how the models were used and to the hardware performing these tasks, transitioning to a GPU inference.

Speech interaction and new interface – To provide a logical, easily understandable experience for the average user, a Dialogue Service was developed to handle all speech-based interaction between the application and the user. Additionally, a more user-friendly interface was created, enabling users to understand everything happening within the system. It was also during this phase that the training plan to be offered was defined, taking into account the limitations identified with the 2D cameras.

First tests with users – An initial user test was conducted to assess the prototype's maturity, identify implementation gaps, and, above all, demonstrate that deploying a virtual gym over a 5G network was not only possible but also feasible. User feedback was positive regarding network-related aspects; however, it was less favorable concerning the quality of voice interaction.

Refinement after users feedback – After user testing and considering their feedback, it was deemed necessary to make the Dialogue Service more robust and intuitive, with clear indicators of its status throughout the system's operation and an interaction tutorial provided before the start of the training session. Some adjustments were also made to the exercise analysis algorithms, particularly for Exercise 2, which had produced highly inconsistent analyses. These changes were crucial for the final deployment in the Casa Viva+ environment.

Demonstration – The results wer showcased at events such as XPERiMENTA 2025 at the University of Aveiro, where it served as an example for students about to enter higher education of the work being done in the field of Computer Engineering at the University of Aveiro, and at the European Researchers' Night 2025, also in

Aveiro, where the project was on display for observation and interaction by the hundreds of attendees.

Development supported by continuous testing was essential for maturing the prototype into a stable, efficient, and user-centered solution, aimed at meeting the objectives set for this dissertation.

7.2 Main results

Considering the objectives of creating a solution that encourages and facilitates motivated physical exercise at home, capable of monitoring and providing feedback on performance, and introduces a points system to increase user engagement and foster a sense of healthy competition, all built on a 5G infrastructure, it can be concluded that all objectives were met except for the points system and gamification. As previously mentioned, it was unfortunately not possible to achieve this goal due to the need to prioritize other aspects of the implementation, particularly user interaction with the system, since without this, the system could not be used at all.

Thus, the main outcomes of this dissertation include:

Functional prototype of a home gym system: A complete solution that enables at-home exercise using HPE and its results to support a motivated and engaging practice.

Real-time exercise evaluation using MediaPipe Pose: The instantiation of the PEM with MPP models enabled the solution to provide real-time monitoring of exercises, delivering feedback on each performance and tracking correct repetitions without the need for specialized hardware.

Spoken interaction: The development of the Dialogue Service enabled a natural way of interacting with a machine, abstracting the user from the mundanity of a standard computer.

Integration with 5G: The system was successfully deployed on the 5G infrastructure of the CHHA project, demonstrating its compatibility with modern communication networks and smart home environments by offloading the most computationally intensive operations to the cloud.

Deployment at CasaViva+: The deployment in the CasaViva+ environment represents the culmination of all the work carried out in this dissertation.

7.3 Future work

Many continuations are possible for this system, some of which have been planned while others remain merely desirable. In order of relevance, they are:

- **Evolution of HPE and exercises**

  Use of machine learning models – One of the most obvious advancements, particularly given the increasing adoption of machine learning models in the industry, would be to replace the current exercise analysis algorithms with these models. It will be necessary to create datasets for each exercise and explore the best combinations of models to obtain all desired outputs for each performance.

  Test and evaluate alternative HPE models – Despite the good HPE results achieved by the MPP models, the lack of confidence in the Z-coordinate values makes their use impractical. As a result, the system is currently using a model that provides more information than required and is computationally demanding. This computational cost led to the inference processes being delegated to a remote Processing Service, introducing a delay between the user's actual movements and the on-screen skeleton display. MoveNet [83] emerged as a solution with an impressive low inference time and minimal computational resources for 2D coordinates, making it feasible to run directly on the machines where the Local Application operates. Given the current use of 2D cameras, MoveNet should be tested to determine whether it can deliver smoother motion and significantly lower latency in the Local Application, thereby improving the UX without compromising any other aspects of the current system.

  Use of 3D cameras and pose models – The clear limitations of 2D cameras suggest that the use of 3D cameras and sensors, such as the Kinect [30] and LiDAR, could significantly improve HPE quality. The Kinect, developed explicitly for real-time HPE tasks, demonstrates very high monitoring accuracy with surprisingly low response time and should be tested as an alternative to the cameras currently used in the system. Any decision to adopt solutions of this kind should be made before training and testing machine learning models, as the data to be evaluated would differ slightly.

  Creation of a professionally validated training plan – Although some research was conducted to select the exercises included in this prototype, none of them received approval from a professional in the field. It would be valuable to develop a training plan with a professional to physically stimulate older adults or individuals undergoing rehabilitation, which could be implemented in the prototype using algorithmic assessment or machine learning models.

- **Improvement of feedback**

Progressive color feedback – One of the main proposals of this dissertation is to provide the user with feedback on the quality of their exercise performance. Currently, this feedback is provided through a binary color system: if the movement is correctly executed, the skeleton turns green; if it is performed incorrectly, it turns red. It would be interesting to implement a system in which the base white color gradually shifts towards green as the movement is executed correctly, or towards red depending on the degree of error in the performance.

Audio feedback – In the context of execution feedback, it is also desirable for the assistant to inform the user which part of their body is compromising the movement and what actions should be taken to correct it.

- **More evaluations**

  New user evaluations – With the implementation of the system at Casa Viva+, it is crucial to test the prototype again, both in terms of the interaction, which has undergone significant improvements, and the refinements made to the WEM algorithms, with new users, to assess whether these changes have contributed to an improved user experience.

- **Improvement of modules**

  Study the replacement of the Web Speech API – Despite the relative satisfaction with the Web Speech API used in the STT engine, it introduces certain limitations, namely its compatibility with only a small number of browsers and the constant need for an internet connection to function. OpenAI's Whisper [84] system was tested with both Portuguese and English audio files and demonstrated considerably good transcription performance, perhaps even superior to that of the Web Speech API. However, the main advantage of this system is its ability to run locally, eliminating the need for an external connection and resolving browser compatibility issues. The feasibility of implementation will depend on whether the system can accurately transcribe the user's speech in real-time with an adequate response time.

- **Implementation of new functionalities**

  Implementation and validation of the gamification system – As mentioned earlier, one of the proposed objectives was to use gamification and scoring strategies to motivate users to engage in more physical exercise. This implementation was not possible due to time constraints and the prioritization of other goals; therefore, in a future iteration, creating a Gamification Module would be the most natural next step. At this stage, it would still take advantage of the results produced by the WEM algorithms.

Implementation of the group training feature – Another proposed functionality to enhance user engagement with the system and promote regular exercise was the ability to perform joint workouts with others via video calls. The backend required for this functionality was implemented using the Jitsi Videobridge [85]. However, for the same reasons that prevented the implementation of the gamification module, this feature was not completed. Developing a frontend page to enable this functionality is undoubtedly one of the next steps for the project.

Initial calibration system using the user's body – The limitations arising from the use of 2D cameras, algorithmic approaches instead of machine learning models, and the low quality of the Z-coordinate values returned by MPP made the reliability of the implemented algorithms highly dependent on the camera's angle and height, as well as on the user's relative position to it. Performing an initial calibration that records the proportions of the user's different body parts would help infer, using only 2D coordinates, the user's posture, the extension or contraction of body segments, and other aspects, thereby providing additional confidence to the algorithms as they operate with only two coordinates per joint.

- **Evolution to a complete usable system**

  Integration into a fully featured gym/fitness system – The current proof of concept only allows starting a workout and following it through to completion, beginning with an interaction tutorial. From a UX perspective, it would be far more interesting to separate the workout from the tutorial and integrate them into a comprehensive gym system with multiple training plans and additional features, such as viewing workout history, tracking results, managing user profiles, and customizing user preferences and needs.

  Extension to more complex exercises – Finally, the natural evolution of this solution's application domain would be towards gym and calisthenics exercises, potentially even integrating cameras into the equipment itself to monitor performances more closely without the need for a complex setup. The creation of Persona 2 in section 3.1 already reflects this evolution, as it targets a younger audience than that of the current prototype.

References

[1] Eurostat. «Ageing Europe - statistics on population developments». Data extraída em julho de 2020, European Commission. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Ageing_Europe_-_statistics_on_population_developments.

[2] R. Schweighart, J. L. O'Sullivan, M. Klemmt, A. Teti, and S. Neuderth, «Wishes and needs of nursing home residents: A scoping review», *Healthcare*, vol. 10, no. 5, 2022, ISSN: 2227-9032. DOI: 10.3390/healthcare10050854. [Online]. Available: https://www.mdpi.com/2227-9032/10/5/854.

[3] «Oli». [Online]. Available: https://www.oli-world.com/pt/.

[4] M. Justine, A. Azizan, V. Hassan, Z. Salleh, and H. Manaf, «Barriers to participation in physical activity and exercise among middle-aged and elderly individuals», *Singapore medical journal*, vol. 54, pp. 581–6, Oct. 2013. DOI: 10.11622/smedj.2013203.

[5] E. Hoare, B. Stavreski, G. Jennings, and B. Kingwell, «Exploring motivation and barriers to physical activity among active and inactive australian adults», *Sports*, vol. 5, p. 47, Jun. 2017. DOI: 10.3390/sports5030047.

[6] J. Park, J. Moon, H. Kim, M. Kong, and Y. Oh, «Sedentary lifestyle: Overview of updated evidence of potential health risks», *Korean Journal of Family Medicine*, vol. 41, pp. 365–373, Nov. 2020. DOI: 10.4082/kjfm.20.0165.

[7] K. Nallaperumal, «Engineering research methodology: A computer science and engineering and information and communication technologies perspective», *ResearchGate*, 2013. [Online]. Available: https://www.researchgate.net/publication/259183120_Engineering_Research_Methodology_A_Computer_Science_and_Engineering_and_Information_and_Communication_Technologies_Perspective.

[8] W. H. Organization, *Active ageing : A policy framework*, 2002. [Online]. Available: https://iris.who.int/handle/10665/67215.

[9] J. S. Novotný, J. P. Gonzalez-Rivas, M. Vassilaki, J. Krell-Roesch, Y. E. Geda, and G. B. Stokin, «Natural pattern of cognitive aging», *Journal of Alzheimer's Disease*, vol. 88, no. 3, pp. 1147–1155, 2022, PMID: 35754277. DOI: 10.3233/JAD-220312. eprint: https://doi.org/10.3233/JAD-220312. [Online]. Available: https://doi.org/10.3233/JAD-220312.

[10] S. Dogra, D. W. Dunstan, T. Sugiyama, A. Stathi, P. A. Gardiner, and N. Owen, «Active aging and public health: Evidence, implications, and opportunities», en, *Annu Rev Public Health*, vol. 43, pp. 439–459, Dec. 2021.

[11] Eurostat. «Mortality and life expectancy statistics». Data extracted in March 2024, European Commission. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Mortality_and_life_expectancy_statistics.

[12] V. Stara *et al.*, «Intrinsic capacity and active and healthy aging domains supported by personalized digital coaching: Survey study among geriatricians in europe and japan on ehealth opportunities for older adults», en, *J Med Internet Res*, vol. 25, e41035, Oct. 2023.

[13] A. Ramalho and J. Petrica, «The quiet epidemic: An overview of emerging qualitative research trends on sedentary behavior in aging populations», en, *Healthcare (Basel)*, vol. 11, no. 15, Aug. 2023.

[14] J. L. Copeland, J. Good, and S. Dogra, «Strength training is associated with better functional fitness and perceived healthy aging among physically active older adults: A cross-sectional analysis of the canadian longitudinal study on aging», *Aging Clinical and Experimental Research*, vol. 31, no. 9, pp. 1257–1263, Sep. 2019, ISSN: 1720-8319. DOI: 10.1007/s40520-018-1079-6. [Online]. Available: https://doi.org/10.1007/s40520-018-1079-6.

[15] E. B. Larson and R. A. Bruce, «Health benefits of exercise in an aging society», en, *Arch Intern Med*, vol. 147, no. 2, pp. 353–356, Feb. 1987.

[16] L. M. Vecchio, Y. Meng, K. Xhima, N. Lipsman, C. Hamani, and I. Aubert, «The neuroprotective effects of exercise: Maintaining a healthy brain throughout aging», *Brain Plasticity*, vol. 4, pp. 17–52, 2018, 1, ISSN: 2213-6312. DOI: 10.3233/BPL-180069. [Online]. Available: https://doi.org/10.3233/BPL-180069.

[17] L. Piccardi *et al.*, «The contribution of being physically active to successful aging», en, *Front Hum Neurosci*, vol. 17, p. 1 274 151, Nov. 2023.

[18] D. V. de Oliveira, C. C. Ribeiro, and S. F. Pinto, «Finding purpose in life: The way of the masters' athletes», *Geriatrics Gerontology and Aging*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:273959475.

[19] S. Dubey and M. Dixit, «A comprehensive survey on human pose estimation approaches», *Multimedia Systems*, vol. 29, Aug. 2022. DOI: 10.1007/s00530-022-00980-0.

[20] J. Wang *et al.*, «Deep 3d human pose estimation: A review», *Computer Vision and Image Understanding*, vol. 210, p. 103 225, 2021, ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2021.103225. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314221000692.

[21] R. Josyula and S. Ostadabbas, *A review on human pose estimation*, 2021. arXiv: 2110.06877 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2110.06877.

[22] G. A. for Developers, *Pose landmark detection guide*. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=en.

[23] G. A. for Developers, *Face landmark detection guide*. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker?hl=en.

[24] G. A. for Developers, *Hand landmarks detection guide*. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker?hl=en.

[25] C. Zheng *et al.*, «Deep learning-based human pose estimation: A survey», *ACM Comput. Surv.*, vol. 56, no. 1, Aug. 2023, ISSN: 0360-0300. DOI: 10.1145/3603618. [Online]. Available: https://doi.org/10.1145/3603618.

[26] A. Singh, S. Agarwal, P. Nagrath, A. Saxena, and N. Thakur, «Human pose estimation using convolutional neural networks», in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 2019, pp. 946–952. DOI: 10.1109/AICAI.2019.8701267.

[27] A. K, P. P, and J. Paulose, «Human body pose estimation and applications», in *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, 2021, pp. 1–6. DOI: 10.1109/i-PACT52855.2021.9696513.

[28] F. Hatta Antah, M. A. Khoiry, K. N. Abdul Maulud, and A. Abdullah, «Perceived usefulness of airborne lidar technology in road design and management: A review», *Sustainability*, vol. 13, no. 21, 2021, ISSN: 2071-1050. DOI: 10.3390/su132111773. [Online]. Available: https://www.mdpi.com/2071-1050/13/21/11773.

[29] Š. Obdržálek *et al.*, «Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population», in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 1188–1193. DOI: 10.1109/EMBC.2012.6346149.

[30] «Azure kinect dk - desenvolver modelos de ia | microsoft azure». [Online]. Available: https://azure.microsoft.com/pt-pt/products/kinect-dk.

[31] OpenPose, *Openpose advanced doc - 3-d reconstruction module and demo*. [Online]. Available: https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_advanced_3d_reconstruction_module.html.

[32] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, «Openpose: Realtime multi-person 2d pose estimation using part affinity fields», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[33] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, «Hand keypoint detection in single images using multiview bootstrapping», in *CVPR*, 2017.

[34] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, «Realtime multi-person 2d pose estimation using part affinity fields», in *CVPR*, 2017.

[35] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, «Convolutional pose machines», in *CVPR*, 2016.

[36] OpenPose, *Openpose advanced doc - 3-d reconstruction module and demo*. [Online]. Available: https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_advanced_3d_reconstruction_module.html.

[37] D. Martinho, J. Carneiro, J. M. Corchado, and G. Marreiros, «A systematic review of gamification techniques applied to elderly care», *Artificial Intelligence Review*, vol. 53, no. 7, pp. 4863–4901, Oct. 2020, ISSN: 1573-7462. DOI: 10.1007/s10462-020-09809-6. [Online]. Available: https://doi.org/10.1007/s10462-020-09809-6.

[38] M. S. Staller and S. Koerner, «Beyond classical definition: The non-definition of gamification», *SN Computer Science*, vol. 2, no. 2, p. 88, Feb. 2021, ISSN: 2661-8907. DOI: 10.1007/s42979-021-00472-4. [Online]. Available: https://doi.org/10.1007/s42979-021-00472-4.

[39] A. N. Saleem, N. M. Noori, and F. Ozdamli, «Gamification applications in e-learning: A literature review», *Technology, Knowledge and Learning*, vol. 27, no. 1, pp. 139–159, Mar. 2022, ISSN: 2211-1670. DOI: 10.1007/s10758-020-09487-x. [Online]. Available: https://doi.org/10.1007/s10758-020-09487-x.

[40] R. Damaševičius, R. Maskeliūnas, and T. Blažauskas, «Serious games and gamification in healthcare: A meta-review», *Information*, vol. 14, no. 2, 2023, ISSN: 2078-2489. DOI: 10.3390/info14020105. [Online]. Available: https://www.mdpi.com/2078-2489/14/2/105.

[41] Y. Li, H. Phan, A. V. Law, A. Baskys, and D. Roosan, «Gamification to improve medication adherence: A mixed-method usability study for medscrab», *Journal of Medical Systems*, vol. 47, no. 1, p. 108, Oct. 2023, ISSN: 1573-689X. DOI: 10.1007/s10916-023-02006-2. [Online]. Available: https://doi.org/10.1007/s10916-023-02006-2.

[42] A. Stefoska-Needham and A. L. Goldman, «Perspectives of australian healthcare professionals towards gamification in practice», en, *Nutr Diet*, Nov. 2024.

[43] M. V. Villasana *et al.*, «Promotion of healthy lifestyles to teenagers with mobile devices: A case study in portugal», *Healthcare*, vol. 8, no. 3, 2020, ISSN: 2227-9032. DOI: 10.3390/healthcare8030315. [Online]. Available: https://www.mdpi.com/2227-9032/8/3/315.

[44] J. E. S. David Hayes and T. A. Harwell, «Preventing pollution: A scoping review of immersive learning environments and gamified systems for children and young people», *Journal of Research on Technology in Education*, vol. 55, no. 6, pp. 1061–1079, 2023. DOI: 10.1080/15391523.2022.2107589. eprint: https://doi.org/10.1080/15391523.2022.2107589. [Online]. Available: https://doi.org/10.1080/15391523.2022.2107589.

[45] M. A. Johnson, «Gamification and its impact on hospitalized children», *Journal of Cancer Research and Cellular Therapeutics*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:272244506.

[46] Y. An, «Designing effective gamified learning experiences», *International Journal of Technology in Education*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:216267566.

[47] «Peloton app: Your on-demand fitness companion». [Online]. Available: https://www.onepeloton.com/app.

[48] «Nike training club app. home workouts. nike.com». [Online]. Available: http://nike.com/pt/ntc-app.

[49] N. Habib, F. Kamal, and M. Munir, «Comparison of the effectiveness of home-based workouts and gym training according to caloric intake», *International Health Review*, vol. 1, pp. 13–29, Dec. 2021. DOI: 10.32350/ihr.0102.02.

[50] M. C. Maccarone *et al.*, «Enhancing quality of life in sedentary elderly individuals: The impact of the home-based full-body in-bed gym program - a prospective, observational, single-arm study», *Bulletin of Rehabilitation Medicine*, Jan. 2023. DOI: 10.38025/2078-1962-2023-22-5-8-14.

[51] «Bed-gym - 2022 - youtube». [Online]. Available: https://www.youtube.com/watch?v=pcHKmxCLYFs.

[52] Y. Wu *et al.*, «Ar-enhanced workouts: Exploring visual cues for at-home workout videos in ar environment», in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '23, San Francisco, CA, USA: Association for Computing Machinery, 2023, ISBN: 9798400701320. DOI: 10.1145/3586183.3606796. [Online]. Available: https://doi.org/10.1145/3586183.3606796.

[53] S. Health. «Thrive». [Online]. Available: https://swordhealth.com/solutions/thrive.

[54] F. D. Correia *et al.*, «Home-based rehabilitation with a novel digital biofeedback system versus conventional in-person rehabilitation after total knee replacement: A feasibility study», *Scientific Reports*, vol. 8, no. 1, p. 11 299, Jul. 2018, ISSN: 2045-2322. DOI: 10.1038/s41598-018-29668-0. [Online]. Available: https://doi.org/10.1038/s41598-018-29668-0.

[55] D. Kim, M. Cho, Y. Park, and Y. Yang, «Effect of an exercise program for posture correction on musculoskeletal pain», *Journal of Physical Therapy Science*, vol. 27, no. 6, pp. 1791–1794, 2015. DOI: 10.1589/jpts.27.1791.

[56] M. Salsali, R. Sheikhhoseini, P. Sayyadi, J. A. Hides, M. Dadfar, and H. Piri, «Association between physical activity and body posture: A systematic review and meta-analysis», *BMC Public Health*, vol. 23, no. 1, p. 1670, Aug. 2023, ISSN: 1471-2458. DOI: 10.1186/s12889-023-16617-4. [Online]. Available: https://doi.org/10.1186/s12889-023-16617-4.

[57] A. Rahmadani, B. S. Bayu Dewantara, and D. M. Sari, «Human pose estimation for fitness exercise movement correction», in *2022 International Electronics Symposium (IES)*, 2022, pp. 484–490. DOI: 10.1109/IES55876.2022.9888451.

[58] N. Kumar Reddy Boyalla, «Real-time exercise posture correction using human pose detection technique», Master's thesis, Department of Computer and Information Science, SUNY Polytechnic Institute, University of New York, 2021. [Online]. Available: http://hdl.handle.net/20.500.12648/8622.

[59] L. Yang, Y. Li, D. Zeng, and D. Wang, «Human exercise posture analysis based on pose estimation», in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, 2021, pp. 1715–1719. DOI: 10.1109/IAEAC50856.2021.9390870.

[60] A. Tharatipyakul, T. Srikaewsiew, and S. Pongnumkul, «Deep learning-based human body pose estimation in providing feedback for physical movement: A review», *Heliyon*, vol. 10, no. 17, e36589, 2024, ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2024.e36589. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844024126205.

[61] M. Kolla, P. V. Gadiraju, and D. Tondanoorthy, «Form check: Exercise posture correction application», English, *Journal of Electrical Systems*, vol. 20, no. 4s, pp. 25–33, 2024. DOI: https://doi.org/10.52783/jes.1819.

[62] T.-Y. Lin *et al.*, *Microsoft coco: Common objects in context*, 2015. arXiv: 1405.0312 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1405.0312.

[63] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, «2d human pose estimation: New benchmark and state of the art analysis», in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014.

[64] Q. Yu, H. Wang, F. Laamarti, and A. El Saddik, «Deep learning-enabled multitask system for exercise recognition and counting», *Multimodal Technologies and Interaction*, vol. 5, p. 55, Sep. 2021. DOI: 10.3390/mti5090055.

[65] A. L. Rosenhaim, «Human action evaluation applied to weightlifting», Accepted: 2023-12-18T00:35:28Z, Dissertation, Jul. 12, 2023. Accessed: Jan. 19, 2025. [Online]. Available: https://hdl.handle.net/10216/152213.

[66] L. Xu *et al.*, «The effects of mhealth-based gamification interventions on participation in physical activity: Systematic review», *JMIR mHealth and uHealth*, vol. 10, e27794, Feb. 2022. DOI: 10.2196/27794.

[67] C. Estgren, *From gamer boy to gym boy : A design study on gamification*, 2023.

[68] D. Kappen, P. Mirza-Babaei, and L. Nacke, «Technology facilitates physical activity through gamification: A thematic analysis of an 8-week study», *Frontiers in Computer Science*, vol. 2, Oct. 2020. DOI: 10.3389/fcomp.2020.530309.

[69] «Adobe: Creative, marketing and document management solutions». [Online]. Available: https://www.adobe.com/.

[70] «Rtmp streaming: Everything you need to know». [Online]. Available: https://restream.io/blog/rtmp-streaming/.

[71] «Srt protocol». [Online]. Available: https://getstream.io/glossary/srt-protocol/.

[72] «Apple». [Online]. Available: https://www.apple.com/.

[73] «What is http live streaming? | hls streaming». [Online]. Available: https://www.cloudflare.com/learning/video/what-is-http-live-streaming/.

[74] «Websocket and its difference from http». [Online]. Available: https://www.geeksforgeeks.org/web-tech/what-is-web-socket-and-how-it-is-different-from-the-http/.

[75] «Webrtc». [Online]. Available: https://webrtc.org.

[76] «Websockets - fastapi». [Online]. Available: https://fastapi.tiangolo.com/advanced/websockets/#install-websockets.

[77] «Google gemini». [Online]. Available: https://gemini.google.com/app.

[78] «Porcupine wake word detection & keyword spotting - picovoice». [Online]. Available: https://picovoice.ai/platform/porcupine/.

[79] «Web speech api - web apis | mdn». [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API.

[80] «Using nlu only». [Online]. Available: https://legacy-docs-oss.rasa.com/docs/rasa/nlu-only/.

[81] «Text to speech - openai api». [Online]. Available: https://platform.openai.com/docs/guides/text-to-speech.

[82] «Gemini-tts | text-to-speech | google cloud documentation». [Online]. Available: https://docs.cloud.google.com/text-to-speech/docs/gemini-tts?hl=pt.

[83] «Next-generation pose detection with movenet and tensorflow.js — the tensorflow blog». [Online]. Available: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html.

[84] «Introducing whisper | openai». [Online]. Available: https://openai.com/index/whisper/.

[85] «Jitsi videobridge | video conferencing for developers». [Online]. Available: https://jitsi.org/jitsi-videobridge/.