

Development of a Robotic Fitness Application using a Participatory Design Approach

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Andreas Krendl, BSc

Matrikelnummer 00626067

an der Fakultät für Informatik
der Technischen Universität Wien

Betreuung: Assistant Prof. Dr.in phil. Mag.a phil. Astrid Weiss

Wien, 1. Mai 2024

Andreas Krendl

Astrid Weiss



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Development of a Robotic Fitness Application using a Participatory Design Approach

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Medical Informatics

by

Andreas Krendl, BSc

Registration Number 00626067

to the Faculty of Informatics

at the TU Wien

Advisor: Assistant Prof. Dr.in phil. Mag.a phil. Astrid Weiss

Vienna, May 1, 2024

Andreas Krendl

Astrid Weiss



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Declaration of Authorship

Andreas Krendl, BSc

I hereby declare that I have written this Thesis independently, that I have completely specified the utilized sources and resources and that I have definitely marked all parts of the work - including tables, maps and figures - which belong to other works or to the internet, literally or extracted, by referencing the source as borrowed.

I further declare that I have used AI tools (such as Google Translate and Writefull) only as an aid, and that my own intellectual and creative efforts predominate in this work.

Vienna, May 1, 2024

Andreas Krendl



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Acknowledgements

Completing this master's degree marks not only the end of my student journey but also the triumph over personal challenges. I express my sincere gratitude to all those who contributed.

A very special thanks goes to my master's thesis advisor Assistant Prof. Dr.in phil. Mag.a phil. Astrid Weiss. Without her, this master's thesis as it is would not have been possible. She helped me a lot with the social science approach in this work. It should also be mentioned that she also supervised me on a project and seminar work prior to my master's thesis. In addition, I especially would emphasize the patience she had with me. Due to health problems, everything ended up taking much longer than I expected and hoped.

I would also like to specifically thank the workshop participants. Some of them have decades of nursing experience and have participated in workshops as experts. Other participants gave their time as middle-aged users. Everyone helped me as a technician with valuable insights and contributed an important part to this master's thesis.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

Angesichts des demografischen Wandels hin zu einer alternden Gesellschaft steht der Gesundheitssektor vor wachsenden Herausforderungen und innovative Lösungen sind gefordert. Der Bereich der assistiven (sozialen) Robotik scheint ein vielversprechendes Feld zur Abmilderung von zu erwartenden Problemen zu sein. Allerdings sind Roboter wie ROBEAR oder Care-O-Bot 4 nach wie vor lediglich komplexe Forschungsplattformen, die nicht in die tägliche Praxis von Menschen integriert sind.

Diese Masterarbeit befasst sich mit der Forschungsfrage, wie eine Anwendung zur Förderung von körperlicher und geistiger Fitness basierend auf einem kommerziell erhältlichen und vergleichsweise simplen kugelförmigen Roboter namens Sphero entwickelt werden kann. Die Anwendung soll von Pflegekräften gemeinsam mit älteren Personen sowie von Benutzer*innen mittleren Alters nachhaltig genutzt werden. Ein wichtiger Schwerpunkt liegt dabei auf der Diskussion, wie neueste Forschungsergebnisse und technologische Entwicklungen im Bereich der künstlichen Intelligenz für eine benutzerfreundliche Steuerung und einfache Implementierung von robotischen Anwendungen verwendet werden können.

Pflegekräfte und Techniker*innen sind Experten auf unterschiedlichen Gebieten und sie haben unterschiedliche Kenntnisse und Erwartungen. Die Masterarbeit zeigt, wie ein „Grounding“ zwischen diesen beiden Gruppen durch Partizipatives Design in Form von Anforderungsworkshops mit einem dreistufigen Design und vorbereiteten Fragen gelingen kann. Dadurch kann eine Roboteranwendung so gestaltet werden, dass sie einen Mehrwert für die Pflegekräfte und letztendlich für ältere Menschen, mit denen diese in Seniorenheimen zusammenarbeiten, bietet und somit die Chance einer nachhaltigen Nutzung erhöht wird. An den im Rahmen dieser Masterarbeit durchgeführten Workshops nahmen (inklusive mir) insgesamt 11 Personen teil.

Mithilfe einer auf den transkribierten Workshops durchgeführten Thematic Analysis werden in dieser Arbeit Fragen im Zusammenhang mit robotischen Anwendungen zur Förderung von körperlicher und geistiger Fitness beantwortet. Dazu gehören Fragen nach Einschränkungen älterer Nutzer*innen, technische Limits der Anwendung und Anpassungen an Sphero, die notwendig wären, damit dieser eine kommerziell erfolgreiche robotische Fitness Anwendung werden könnte, sowie Fragen nach Erwartungen und Anforderungen, die die Nutzer*innen haben.

Entsprechend der in den Workshops gewonnen Daten sollte eine robotische Fitness Anwendung benutzerfreundlich, individualisiert, anpassbar und motivierend sein, damit

sie nachhaltig genutzt wird. Im Zusammenhang mit Motivation betonen die Workshop-Teilnehmer*innen Faktoren wie sinnvolles Feedback, „Gamification“, soziale Integration und emotionalen Support. Eine durchgeführte Literaturrecherche und gewonnene Daten deuten darauf hin, dass höhere wahrgenommene Intelligenz, wie sie bei der Verwendung von neuartigen Sprachmodellen wie GPT-4 entsteht, das Potenzial haben die Motivation von Nutzern zu steigern. Anders als Teilnehmer*innen von Workshops früherer Studien äußern die Teilnehmer*innen der Workshops, die im Rahmen dieser Masterarbeit durchgeführt wurden, keine Bedenken bezüglich intelligenter (robotischer) Systeme.

Die Erkenntnisse aus den Workshops flossen in die Entwicklung einer prototypischen, agentenbasierten, robotischen Fitness Applikation mit Sphero, die in einem Assessment Workshop von den Teilnehmer*innen trotz einiger (konstruktiver) Kritik grundsätzlich für vielversprechend erachtet wurde, ein. Die Anwendung stellt insbesondere für ältere Menschen, die wenig Erfahrung im Umgang mit Smartphones und Computer haben, eine besonders einfache Möglichkeit der Bedienung bereit, da sie via Sprachsteuerung aufgefordert werden kann, Übungen mit Sphero zu starten und selbst in Datenbanken nach Informationen suchen kann.

Die erfolgreiche und vielversprechende prototypische Umsetzung einer Fitnessanwendung basierend auf einem simplen und kostengünstigen Roboter, in den Workshops gewonnene Daten, sowie eine Literaturrecherche deuten darauf hin, dass sich ein Teil der Forschung stärker auf weniger komplexe und kosteneffektive Roboterlösungen konzentrieren sollte, um schneller zu Produkten zu gelangen, die das Potenzial haben einige der zu erwartenden Probleme aufgrund der demografischen Entwicklung zu lindern.

Abstract

In light of demographic changes towards an aging society, the healthcare sector is faced with growing challenges and innovative solutions are required. The area of assistive (social) robotics appears to be a promising field for mitigating anticipated problems. However, robots such as ROBEAR or Care-O-Bot 4 remain merely complex research platforms that are not integrated into the daily practices of people.

This master's thesis addresses the research question of how to develop an application that promotes physical and mental fitness using a commercially available and relatively simple spherical robot called Sphero, aimed at sustainable use by caregivers together with elderly people, and by middle-aged users. An important focus is on the discussion of how the latest research findings and technological developments in the field of artificial intelligence can be used for user-friendly control and easy development of robotic applications.

Caregivers and technicians are experts in different fields and have varying knowledge and expectations. The master's thesis demonstrates how grounding between these two groups can be achieved through participatory design in the form of requirements workshops with a three-step design and prepared questions, so that a robotic application can be developed that provides added value for nursing staff and ultimately also for older people, with whom they work together in retirement homes, and thus the chance of sustainable use is increased. A total of 11 people (including me) took part in the workshops conducted as part of this master's thesis.

Using thematic analysis performed on the transcribed workshops, this work answers questions related to robotic applications for promoting physical and mental fitness. These include questions about limitations of older users, technical limits of the application and adjustments to Sphero that would be necessary to become a commercially successful robotic fitness application, and questions about expectations and requirements that the user have.

According to data obtained in the workshops, a robotic fitness application should be user-friendly, individualized, adaptable, and motivating to ensure sustainable use. In connection with motivation, workshop participants highlight factors such as meaningful feedback, "gamification", social integration, and emotional support. A literature review and the data obtained suggest that higher perceived intelligence, as occurs with the use of novel large language models such as GPT-4, has the potential to increase user

motivation. Unlike participants of workshops in earlier studies, participants in the workshops conducted as part of this master's thesis express no concerns about intelligent (robotic) systems.

The findings of the workshops were incorporated into the development of a prototypical, agentic, robotic fitness application with Sphero, which was generally considered promising by the participants in an assessment workshop despite some (constructive) criticism. The application provides a particularly easy way of operation, especially for older people who have little experience in using smartphones and computers, as it can be instructed via voice control to start exercises with Sphero and it can search for information in databases itself.

The successful and promising prototypical implementation of a fitness application based on a simple and cost-effective robot, data obtained in the workshops, and a literature review indicate that part of the research should focus more on less complex and cost-effective robotic solutions to arrive more quickly at products that have the potential to alleviate some of the expected problems due to demographic developments.

Contents

Kurzfassung	ix
Abstract	xi
Contents	xiii
1 Introduction and Problem Statement	1
1.1 Overview of the Introduction and Problem Statement Chapter	1
1.2 Demographic Development	1
1.3 Implications of the Demographic Development	3
1.4 Robots	4
1.5 Aim of the Work and Research Questions	7
2 Literature Review	11
2.1 Overview of the Literature Review Chapter	11
2.2 Artificial Intelligence for Robotics	11
2.3 Specialized and Less Complex Robots	20
2.4 Aspects for the Development of Assistive Social Robots	21
3 Methodological Approach	25
3.1 Overview of the Methodological Approach Chapter	25
3.2 Quantitative vs. Qualitative Approaches	26
3.3 Participatory Design	27
3.4 Workshops	27
3.5 Thematic Analysis	32
3.6 Timeline of the Master's Thesis	34
4 Implementation	37
4.1 Overview of the Implementation Chapter	37
4.2 Sphero Bolt and Sphero Play	38
4.3 Exercise Movement Pattern Classification	40
4.4 Verbal Communication with GPT-4	43
4.5 Agentic Behavior	46
4.6 Vector Database	47
	xiii

4.7	Developed Tools	48
4.8	Workflow of the Final Implementation	48
5	Findings and Discussion	55
5.1	Overview of the Findings and Discussion Chapter	55
5.2	Application Scenarios and Usage Contexts	56
5.3	Participatory Design as Part of the Master's Thesis	57
5.4	Presentation of the Results of the Thematic Analysis	59
6	Conclusion	73
	List of Figures	85
	Bibliography	89

Introduction and Problem Statement

1.1 Overview of the Introduction and Problem Statement Chapter

This Introduction and Problem Statement chapter begins by discussing the observable demographic trends in many countries around the world and some of the expected implications for the healthcare sector that arise. Afterwards, a classification of robots takes place and it is discussed how assistive robots can help in dealing with the expected problems. Concrete examples of robots that are being researched and robots that are already in use are given. In the last section, the research question is derived and formulated.

1.2 Demographic Development

It is well known that the world population is still growing. However, the fertility rate of the world has been continuously falling for decades and reached only 2.3 children per woman in 2021. [Wora] With a few exceptions, high fertility rates can now be found almost only in countries on the African continent. In this mentioned year 2021, the fertility rate was 1.8 in Mexico, 1.7 in Iran, USA and Australia, 1.6 in Germany and Hungary, 1.5 in Austria, Switzerland and Russia, 1.4 in Canada, Portugal and Greece, 1.3 in Japan, Italy and Poland, 1.2 in China, Spain and Ukraine, and 0.8 in South Korea, to name a few specific examples. [Worb]

Developed countries whose low fertility rates are not compensated by immigration or which (additionally) are affected by emigration are already stagnating or shrinking. For example, Figure 1.1 shows the development of the populations of Italy (top left) [Stad],

1. INTRODUCTION AND PROBLEM STATEMENT

Hungary (top middle) [Stah], and Japan (top right) [Staf] including forecasts up to 2050. The populations of these countries are already declining, and, at least from the current perspective, it can be expected that these countries will continue to shrink significantly. Below the curves with population growth are the graphs with the previous, current and predicted age structures of Italy (bottom left) [Stac], Hungary (bottom middle) [Stag] and Japan [Stae]. Current trends in these countries show a decrease in the working-age population and a significant increase in the population aged 65 and over, and from the current perspective, these trends are expected to continue into 2050.

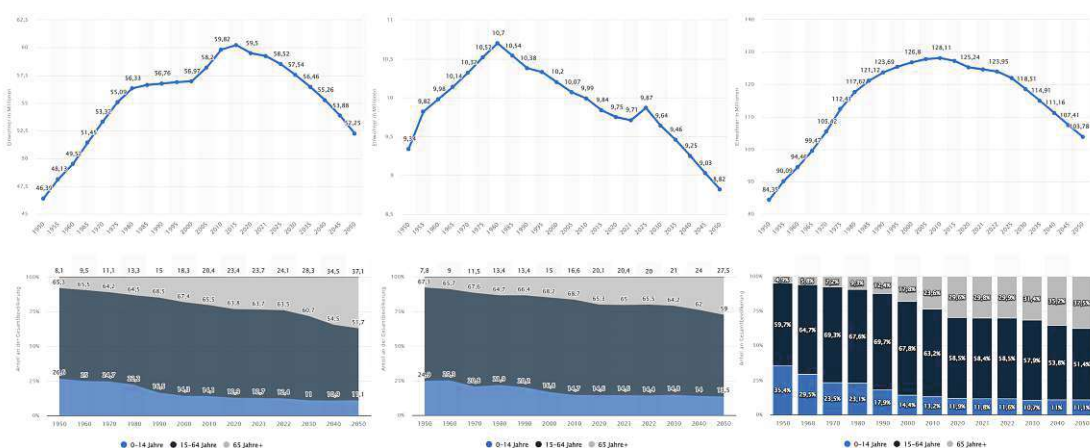


Figure 1.1: Population development of Italy (top left) [Stad] , Hungary (top middle) [Stah], Japan (top right) [Staf] and age structures of of Italy (bottom left) [Stac], Hungary (bottom middle) [Stag] and Japan (bottom right) [Stae]

Other developed countries are currently only growing due to immigration. According to a forecast, Austria’s current population of around 9 million could increase to around 9.9 million by 2050. [Stab] However, Austria’s age structure is also changing. Currently, around 61% of Austrians are between 20 and 64 and 19.4% are 65 years or older. According to a forecast, the proportion of 20 to 64 years old could drop to below 53% by 2050, and at the same time the proportion of people 65 years or older could increase from current 19% to almost 27%. [Staa]

Of course, projections over longer periods should be viewed with caution. In the future, more people could move from areas with strong population growth in Africa to areas with a population decline. However, the fundamental trend in Europe and parts of America and Asia that the proportion of people in working age is decreasing and the proportion of the rest of the population is increasing can now be observed for decades, as, for example, OECD statistics show. [OEC]

1.3 Implications of the Demographic Development

It is well known that many diseases are more likely to occur in older age. According to a study by the Austrian Interdisciplinary Platform on Aging, almost no one 80 years or older is free of diseases or health problems in Austria. Examining the study participants revealed that almost all the individuals had chronic diseases, including high blood pressure, urinary or stool incontinence, heart disease, rheumatism/gout, osteoporosis, cancer, diabetes, chronic respiratory disease, or dementia. [ÖP] One cause of many disabilities in older people is stroke and other neurological conditions. Mortality after stroke has decreased due to rapid diagnosis and rapid transport to hospitals with special stroke units. However, approximately 80% of stroke survivors suffer long-term limitations. [GACG17] Because of the outlined demographic developments to be expected and the significantly increased risk of suffering a stroke in old age, there could be a risk of a stroke epidemic in the future. [NT05]

At the same time, the shortage of skilled workers in the health sector is often discussed in the media these days. Buxel conducted an empirical study on job satisfaction and job selection behavior of nursing staff in Germany. More than 3000 people from the care sector were interviewed as part of the study. Although 80% of the people asked identified themselves with their job, only about half were satisfied with their work environment and only a third would recommend their choice of career. The greatest dissatisfaction was found when it comes to the appreciation of nursing staff in hospitals, the high workload and stress, and the small number of colleagues in the shift operation. [Bux11]

One could argue, complaints from nurses about their high workload and the portrayal of the nursing profession as stressful in the media are unlikely to attract more young people to the field. This could potentially lead to an even more challenging situation for existing nursing staff, especially given demographic trends. Having too few specialists in the health sector could lead to longer waiting times for crucial therapies and operations. Individual counseling sessions could tend to be shorter and many patients could not receive enough attention. Due to the high workload of healthcare professionals, there could also be more errors, such as misdiagnoses.

Austria's health spending increased from around 26.1 billion euros per year to around 53.6 billion euros per year from 2005 to 2022. It is certainly important to remember that Austria's GDP also increased during this time. However, overall health spending as a percentage of GDP has also increased from 10.3% in 2005 to 12.0% in 2022. [Aus] If high-quality health care based on the latest research and technology that people in Austria are used to should be maintained, costs will increase further due to demographic change. Although it must be noted again that at this point no one can say with certainty how much immigration Austria will have and how high Austria's GDP will be in the future, the Federal Ministry of Finance in Austria assumes that, for example, care spending alone will double from 1.3% of GDP in 2019 to 2.6% of GDP by 2050. [fF] However, it should be noted that Austria is one of the richest countries in the world and ages less strong than other countries from a current perspective. In other countries, the challenges

to be expected could be even greater.

Many measures are conceivable to reduce these problems, keep the cost increase as low as possible, and continue to provide the best possible health care based on the latest research for everyone. This includes necessary political decisions, and on the other hand, many different technical solutions such as robots.

1.4 Robots

1.4.1 Classification of Robots

Various frameworks can be found in the literature to classify robots. [BHvdHdW17], [PSMG⁺19], [HKEW10] Figure 1.2 shows a classification according to Heerink et al. Robots and robotic devices here are divided into industrial and non-industrial robots. By non-industrial robots, the authors mean those robots that are not focused on production processes. The non-industrial robots are either non-assistive or assistive. Assistive robots are either non-social assistive or assistive social. Assistive social robots are socially interactive, communicate with users, and can be further divided into service robots and companion robots. Service robots can promote independence for older people or other people in need by supporting basic activities. Companion robots are designed primarily for social support. Specific examples are given in the following for both categories. [HKEW10]

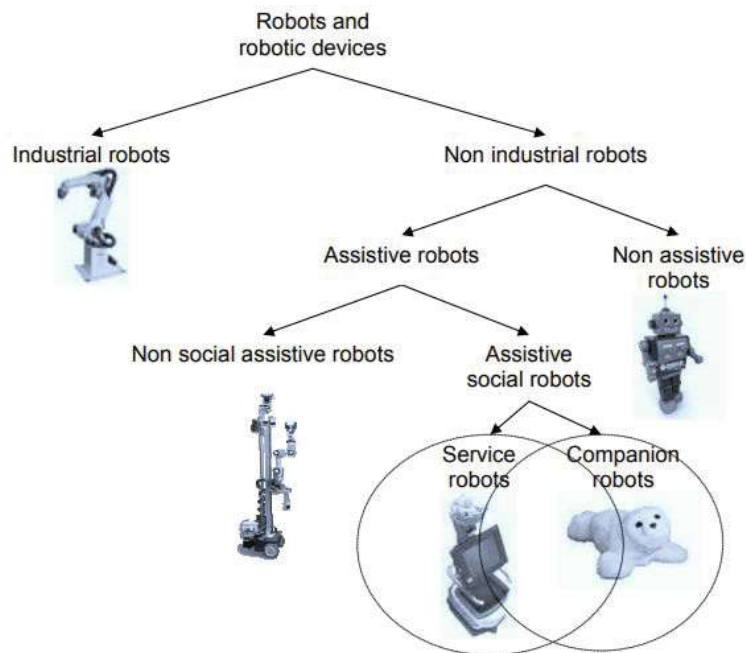


Figure 1.2: A classification of robots according to Heerink et al. [HKEW10]

1.4.2 Reducing Problems with Assistive (Social) Robots

The Ambient Assisted Living (AAL) concept offers interesting solutions to help people live a self-determined and independent life in their own apartment or house as long as possible. In AAL systems, technologies such as sensors, actuators, and cameras are integrated into the living environments of older people or disabled people. AAL systems are capable of enhancing personal autonomy and quality of life, for example, by detecting falls or extended stays at the same place. Robots could also be integrated and offer additional advantages. Robots can, for example, be used to detect falls or extended stays in areas that are not covered by cameras. [GDER⁺19] Further examples of how robots could improve the lives of older people (and other people in need) and promote their independence include help with household tasks such as cleaning and cooking and help with hygiene activities. Robots can also help communicate with family members or caregivers, and emotional robots can provide companionship. [GHKW13]

A concrete example of a robot that is supposed to help with household tasks is Care-O-Bot 4 that can be seen in Figure 1.3 (left). According to Heerink’s classification, it is a service robot that belongs to the class of assistive social robots. Care-O-Bot 4 is the fourth version of a mobile robot assistant to support people at home from Fraunhofer IPA. Its design is modular and offers versatile usage options. Depending on the application, it can be equipped with no arms, one arm, or two arms. It can be used for pick-up and delivery services, serving drinks, or serving as a mobile information kiosk. It is able to display moods through its integrated display in its head. [IPA] Care-O-Bot 4 is conceptualized as a “gentleman robot”, enabling users to connect with it on an emotional and social level. [KFS⁺15] However, it should be noted that although Fraunhofer IPA has been working on the development of a household assistant for nearly 26 years now [GHKW13], the project is still only a vision of a product. [IPA]



Figure 1.3: Care-O-Bot 4 (left) [IPA], PARO (right) [SHP⁺21]

An important goal is certainly to retain skilled workers in the health professions and to encourage young people to complete training in the health sector. In addition to conceivable measures such as reductions in working hours and higher salaries, technical solutions such as robots could help to make work more interesting, work processes in hospitals or retirement homes more efficient, and to relieve people’s workload. Robots can help reduce stress and the risk of injury. This can lead to a reduction in sick leave and an improvement in employee retention. [HBG⁺20] While the objective is undoubtedly

to enable elderly individuals to remain in their own homes for the maximum feasible duration, institutionalized care is still unavoidable in many cases. Caregivers spend 20% of their work on non-nursing activities. Robots have the potential to relieve them. For example, they can transport goods such as laundry or medications, or provide power support and semi-autonomous movement guidance, and emotional robots can serve as therapy and communication aids. [GHKW13]



Figure 1.4: Delivery robots of the company Aethon [Inca]

An example of a commercially available automated robotic delivery system comes from the company Aethon. [Inca] These robots that can be seen in Figure 1.4 drive on four wheels and can transport drugs, clean or dirty laundry, operating room supplies, patient meals, medical waste or garbage. [Inca] A study showed how, for example, the average pharmacy cycle time from order entry to order dispatch could be reduced by almost 30% in a hospital. The time saved can be used for other activities and leads to cost savings. [TDS22]

An important task for nursing staff is the transfer of people. Sumitomo Riko Company Limited and Riken in Japan developed an experimental assistive caregiver robot called ROBEAR, which can lift patients. Figure 1.5 shows how ROBEAR helps a person get off a couch and sit in a wheelchair. [Rik] However, the development of ROBEAR was eventually stopped several years ago and it did not become a commercial product that is actually used in hospitals or nursing homes.

Another concrete example is PARO, which was developed by Shibata et al. in Japan [oAISA] and can be seen in Figure 1.3 (right). According to Heerink's classification, PARO is a companion robot and, therefore, also an assistive social robot. It has long been known that interactions with animals benefit people emotionally, but many hospitals, especially in Japan, did not accept animals because they feared negative effects such as allergies, infections, bites, and scratches. Shibata et al. have therefore proposed a robot that can be used as a replacement for animals for animal-assisted therapy. Modeled after a baby seal, PARO is covered in white, natural-feeling fur and mimics animal behavior using tactile, visual, auditory, and posture sensors and various actuators. Today, PARO



Figure 1.5: The caregiver robot ROBEAR helps a person from a sofa to the wheelchair [Rik]

is used in many care facilities around the world. In the United States, for example, it has been approved as a neurological therapeutic device since 2009. [oAISA].

1.5 Aim of the Work and Research Questions

These examples mentioned so far represent only a part of numerous ways in which nursing staff, but also other professional health workers, older people, and others in need could be helped with robots. A possible political strategy to mitigate some of the expected problems outlined above could be to raise the retirement age. However, in order to be able to work longer, people would have to be healthy enough. Therefore, prevention of diseases and maintenance of health and fitness play an important role in the context of this master's thesis. Many fitness apps already exist, but widespread robotic applications are also conceivable. Assistive social robotic fitness devices could support rehabilitation and promote the well-being of people of all ages. [DALT22]

Much research is being done on robotic applications in healthcare. However, as already mentioned, assistive (social) robots like ROBEAR or Care-O-Bot 4 are complex research platforms or product visions that are not (yet) integrated into daily life of people. During the course of this master's thesis, a discussion takes place about why these robots are not used regularly. PARO, on the other hand, has been used for years. What is striking at first glance is that PARO is comparatively less complex. This suggests the question whether part of the research should focus more on less complex robots for less complex environments and settings. To investigate this question, a goal is to develop a prototype of an assistive social robotic fitness application based on a simple commercially available robot.

There have been interesting advances in the field of machine learning in recent years.

Another important goal of this master's thesis is therefore to answer the question of how these can be used for the development of intuitive, easy-to-use, and easy-to-develop robotic applications. In order to approach this question, a literature review is first carried out. It offers a summary of the latest advances in AI research, discusses problems and limitations, and is also intended to provide an outlook on expected developments. In relation to physical and mental fitness exercises, motivation plays an important role. Therefore, a further goal of this work is to find out whether higher (perceived) artificial intelligence can have a positive effect on motivation.

This work assumes that the likelihood of integration and sustainable use of robotic systems in daily life can be significantly increased if users are taken into account during development. A target group of this master's thesis are nurses working in retirement homes, as it seems that they will become increasingly important given the demographic developments mentioned. Caregivers can have professional training or be relatives of affected people. In practice, nurses work with older people in retirement homes, so the elderly emerge as a second target group that is addressed in this master's thesis. As mentioned above, lifelong training and prevention play an important role. Therefore, the third target group considered in the context of this work are middle-aged people.

Technicians and professional nurses are experts in different fields. Both have different knowledge and different expectations. An important question that is addressed in this master's thesis is, therefore, how to achieve a grounding between professional caregivers and technicians so that the application can be designed in such a way that it brings added value to nurses and, ultimately, also to older people. To achieve grounding, workshops are held. The design of these workshops is based on previous research and own considerations and is described in detail in the Methodological Approach chapter.

As part of these workshops, questions about participants' requirements and expectations, possible limitations that could arise with respect to possible physical and cognitive restrictions of older people, and possible technical problems and limits are answered. A question that is also discussed during the workshops is how the participants see intelligent robots and whether there are any possible concerns. The results of the workshops are incorporated into the development of the prototype assistive social robotic fitness application and are discussed in the Findings and Discussion chapter. The development of the application based on the literature review and discussions with the mentioned target groups during workshops is described in the Implementation chapter.

The research question that is investigated as part of this master's thesis is therefore

- How can a fitness application be developed based on a commercially available robot that is used sustainably by nurses and middle-aged users?

This research question is answered by answering the following subquestions that arise in the context of this work.

- How can a grounding between technicians and nurses be achieved so that the application can represent added value for nurses and ultimately also for older people?
- What requirements and expectations must the application meet for the target groups to be assessed as successful?
- What possible limitations could arise with respect to the physical and cognitive restrictions of older people?
- How can we use the latest AI developments for intuitive, easy-to-use, and easy-to-develop robotic applications?
- What do people think about intelligent robots? Are there possible concerns?
- Can higher (perceived) intelligence have a positive effect on motivation?
- What technical problems and limits arise for the developed application?
- Should research focus more on less complex robots?



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Literature Review

2.1 Overview of the Literature Review Chapter

In this Literature Review chapter, an extensive discussion of the basics, current developments, and expected developments in the field of artificial intelligence research is presented. In addition, aspects of the development of assistive robots are discussed. The aim is an initial approach to the questions as to how the latest AI developments can be used for intuitive, easy-to-use and easy-to-develop robotic applications, whether some of the research should focus more on less complex robots, what people think about intelligent robots, and whether higher perceived intelligence can have positive effects on motivation.

- How can we use the latest AI developments for intuitive, easy-to-use, and easy-to-develop robotic applications?
- Should research focus more on less complex robots?
- What do people think about intelligent robots? Are there possible concerns?
- Can higher (perceived) intelligence have a positive effect on motivation?

2.2 Artificial Intelligence for Robotics

2.2.1 Intelligent Robots

Karel Čapek was an early twentieth century science-fiction novelist. One of his most famous stories was *Rossumovi Univerzální Robot* (R.U.R), from 1920. [Rob05] In this theater play, a company created artificial humans. These were intended as cheap workers without rights and were called robots. The use of this cheap labor completely changed

industry and the global economy. [Cap01] [Rob05] Similar stories have been told over and over again afterward. The term robot and the vision that these artificial workers could transform the world economy as well as the fear of possible dystopian developments is present to this day.

Most researchers in the field of artificial intelligence (AI) research assume that reaching human-level AI or even artificial general intelligence (AGI) is only a matter of time. When it comes to the question of when this will be the case, opinions differ widely. Therefore, the aim of this Literature Review chapter is to approach this question by first providing an overview of existing AI architectures, presenting some of the latest research results, discussing the problems and limitations of current AI, and providing an outlook on expected developments.

At the same time, an extensive basis will be laid for the Implementation chapter and for the question of how the latest AI developments can be used for intuitive, easy-to-use and easy-to-develop robotic applications.

2.2.2 Neural Networks

When AI is discussed, machine learning is usually what is meant these days. In recent times, the domain of machine learning has seen significant advancements. Neural networks are used in both deep learning and deep reinforcement learning because they are good approximators of functions that would be very difficult to find otherwise.

Deep neural networks consist of multiple layers with artificial neurons. Untrained neural networks initially provide random output. This output is compared to a desired output and the difference is determined as an error using a loss function like cross entropy or mean squared error. This error is then propagated back to the input layer, and therefore this process is called backpropagation. The weights of the neuronal connections are adjusted by gradient descent, and the aim is to minimize the loss. [SPL22]

The history of neural networks is many decades old, and there are many different architectures. It is not the aim of this work to describe all of them in detail. However, a brief overview of three already very commonly used network families, namely convolutional neural networks (CNNs), recurrent neural networks (RNNs) [SPL22], and transformers [VSP⁺17], should be presented.

CNNs are used successfully, for example, in the area of image processing. They consist of convolutional layers and pooling layers. In the convolution layers, filter kernels are applied, and features are extracted. When used for image classification in earlier layers, edges could be detected, in later layers lips or eyes, and in even later layers faces. The pooling layers are used for downsampling to reduce the number of parameters to learn. Image classification, image segmentation, or object detection can serve as application examples. [SPL22]

RNNs differ from feedforward networks, such as CNNs. Neurons are fed back into the same layer or into previous layers. This is closer to the functioning of natural neurons in

the human brain and offers advantages when processing sequential data such as speech, audio, or video. A form of RNNs are LSTMs. LSTM stands for Long Short Term Memory and was presented in 1997 by Hochreiter et al. LSTMs have a short-term memory pathway and a long-term memory pathway. For each new input, an LSTM decides how much of the previous long-term memory can be forgotten, how much of the new input is remembered in the long-term memory and how much of this new long-term memory is then used to calculate the next output. [HS97] In the case of text, the next output could be the next word. LSTMs were widely used in industry, for example, by Google for voice transcription. [Bea15]

2.2.3 Transformer Neural Networks

After the publication of the paper “Attention is all you need” in 2017 [VSP⁺17], the transformer architecture gained great importance in AI research. Like RNNs, transformers are suitable for processing sequential data, such as language, but can remember even much longer texts than LSTMs, which were a further development of recurrent networks, as mentioned. Unlike RNNs, transformers do not process sentences word-by-word sequentially, but sentences are processed as a whole. This results in great parallelization advantages on modern graphics cards, and gigantic amounts of data can be used for training. Large language models (LLMs) such as GPT-4 (Generative Pretrained Transformer 4) [AAA⁺23], DeepMind’s Sparrow [GMT⁺22], or Google’s PaLM [CND⁺22] are based on transformer architectures. GPT-4 is used to generate text for the voice output of the implementation that is being done as part of this master’s thesis. Furthermore, it should also be mentioned that transformers are suitable not only for sequential data but also for other data such as images. For example, image classification can be done not only with CNNs but also with vision transformers. [DBK⁺20]

The transformer architecture described in the paper “Attention is all you need”, as can be seen in the left image of Figure 2.1 consists of an encoder module (left) and a decoder module (right). If it is used for language in the encoder part, text is converted to a vector representation, and the decoder converts a vector representation into new text. More specifically, the input is first encoded into input tokens. These are numbers that are then mapped to input embeddings (word embeddings if used for text). Because there is no recurrence as in recurrent networks, it is necessary to introduce information about token positions in a sequence. Consequently, positional encodings are incorporated into the input embeddings at the base of both the encoder and decoder, as visualized on the left in Figure 2.1. [VSP⁺17] In the case of a sentence, the positional encodings describe the positions of the individual words within this sentence. The model learns to interpret these during training. In addition to positional encodings, the self-attention mechanism also plays a crucial role in transformers. The meaning of a word is derived from the context in which it appears. Self-attention takes into account the relationship among words within a sentence (if it is used for language). As there are usually several attention patterns, multi-head attention layers can be used, and attention is then calculated not only once but several times. The encoder part of the architecture presented in the “Attention is all

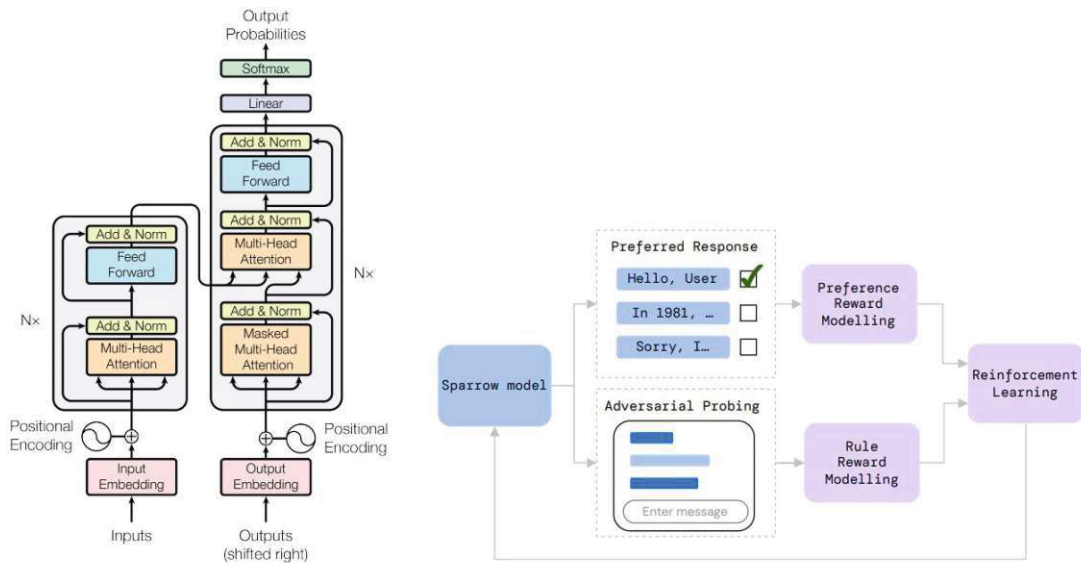


Figure 2.1: Encoder and decoder of the transformer architecture (left image) [VSP⁺17] Pipeline to improve Sparrow (right image). [GMT⁺22]

you need” paper essentially consists of multi-head attention layers mentioned and a fully connected feed-forward network, where each neuron in a layer is linked to every neuron in the subsequent layer. Within the encoder, the embedding vectors are enriched with important information. The decoder is similar to the encoder and also generates enriched embeddings. It essentially consists of masked multi-head attention layers, multi-head attention layers, and a feed-forward network. A linear transformation followed by a softmax function transforms the output of the decoder into probabilities for the next token prediction. [VSP⁺17] The decoder requires two inputs, on the one hand, the encoder output, which are the enriched embedding vectors mentioned, and, on the other hand, the output text that has been generated by the decoder thus far. In other words, large language models using this architecture calculate the next most probable word for a sentence, the whole output sentence including this new word is then fed back into the decoder, the next most probable word is calculated, and so on. [VSP⁺17]

2.2.4 Reinforcement Learning

Tasks that are simple for humans, such as lifting objects, can be very difficult for robots. Reinforcement learning is considered by many researchers to have the potential to bring humanity closer to intelligent autonomous robotic behavior. In contrast to deep learning approaches presented so far, with reinforcement learning, there is no data set for training. The goal of this approach is to learn a control strategy called a policy. First, at a certain point in time t a state s_t of an environment is observed. An agent then chooses an action a_t according to its policy and this action is then carried out within the environment. The environment transitions from state s_t to state s_{t+1} . This in turn is then transmitted to

the agent along with a reward r_{t+1} . Then the agent again selects an action according to its policy, and so on. Knowledge of state transitions is used to learn and improve the policy. [ADBB17] Even if this approach differs from the deep learning approaches mentioned above, neural networks are also often used in practice today for reinforcement learning. These are, for example, policy networks that convert, e.g. input image frames to output actions.

There are a variety of reinforcement learning approaches. All have their advantages and disadvantages. An example is Deep Q Learning by DeepMind, where agents learned to master 2600 Atari games. [MKS⁺13] Also worth mentioning are various actor-critic approaches, such as asynchronous advantage actor-critic (A3C). [MBM⁺16] A detailed description of these and other (deep) reinforcement learning approaches is omitted as it is not used directly in the implementation that is being done as part of this master's thesis. However, as mentioned, the GPT-4 model is used and this has been improved by OpenAI through a method known as reinforcement learning from human feedback (RLHF). [AAA⁺23] A similar approach is explained in the "Problems and Limitations of Current LLMs" section, using the Sparrow model as an example.

2.2.5 Using Large Language Models for Robots

LLMs can be used for the generation of responses from assistive social robots. Ameca [Arta], visualized in Figure 2.2 is probably the most lifelike humanoid robot currently exists. It has gray skin and does not look exactly like a human, but the facial expressions are very realistic. The manufacturer of Ameca Engineered Arts used GPT-3 for the generation of responses. [Artb] However, it should be noted that GPT-3 is a large model and is only available through the API and the model needs time to calculate the answers in a data center. Accordingly, the answers follow with a delay of a few seconds.



Figure 2.2: Ameca by Engineered Arts [Arta]

LLMs can be used not only to generate plausible answers, but also to be able to interact better with robots. Google used its PaLM [CND⁺22] in 2022 for PaLM-SayCan. [ABB⁺22] If an LLM is asked "I spilled my drink, can you help me?", it answers, e.g., with "You could try using a vacuum cleaner". However, an LLM does not know whether a robot is capable of operating a vacuum cleaner. Therefore, the PaLM-SayCan approach combines high-level semantic knowledge with the capabilities of a robot that can be seen in Figure 2.3 left. The robot provides a repertoire of learned skills for "atomic"

behaviors. PaLM is used to determine the probability scores for these actions after a given question and it combines them into a meaningful sequence. For the robot, the authors used reinforcement learning to train a policy and a value function. The value function returns a score for how likely an action is to succeed in the real world. Then these two scores from PaLM and the robot are combined together. For example, the action “Find a grapefruit soda” should have the highest score after the command: “Move the grapefruit drink from the table to the close counter”. The robot then executes “Find a grapefruit soda”. The next action is then “Pick up the grapefruit soda”, because the propability score of PaLM and that of the value function combined are again the highest value, and so on. [ABB⁺22]

2.2.6 Problems and Limitations of Current LLMs

As mentioned above, large language models (LLMs), such as GPT-4 [AAA⁺23], Sparrow [GMT⁺22], or PaLM [CND⁺22], are based on transformer architectures. LLMs can create consistent texts, and communication with them is often difficult to distinguish from communication with a human. However, major problems are that LLMs are not factual, tend to hallucinate, and can be toxic and sexist. Therefore, a great deal of effort must be put into alleviating these problems. An approach to alleviate these problems is reinforcement learning from human feedback (RLHF). Glaese et al. asked raters to pick the best of several possible statements and to detect violations of specific rules. The newly generated data was used to train reward models. The reward models were then used to improve the policy using reinforcement learning. The pipeline to improve the Sparrow model is illustrated in the right image in Figure 2.1. [GMT⁺22] As mentioned, OpenAI also uses an RLHF approach to improve GPT-4 models. [AAA⁺23]

What LLMs do is a sophisticated form of autocomplete. This can lead to astonishing results, but this should not be confused with thinking. Even if LLMs can generate complex computer code and pass tests in mathematics, physics, and law, their intelligence is not comparable to the autonomous intelligence of humans. LLMs can not, for example, plan as humans can. [VOSK22] The way current language models produce text also differs from how the human brain works, as Caucheteux et al. showed in 2023. While LLMs are designed to predict adjacent words, the brain predicts “a hierarchy of representations that spans multiple time scales”. [CGK23] LLMs also do not have an understanding of the physical world itself, as humans and animals do. A lot of the common sense of humans is not captured in written form, but emerges from human interactions with their environment. [LeC22] Furthermore, LeCun points out that humans and animals also have much better learning abilities than any current machine learning approach. He argues that this has to be the case because how else could adolescents learn to drive in just 20 hours or how else could children learn a new language in such a short time? [LeC22]

2.2.7 Multimodal Models

A current trend in AI research is multimodal models such as DALLE 3 [BGJ⁺23], which is capable of generating images from text, or MusicLM [ADB⁺23], which is capable of creating music from text. GPT-4 is also no longer a pure language model. It accepts text and images as input and generates text output. [AAA⁺23] This leads to interesting image understanding skills, such as explaining why funny pictures are actually funny, understanding bar graphs, and much more.

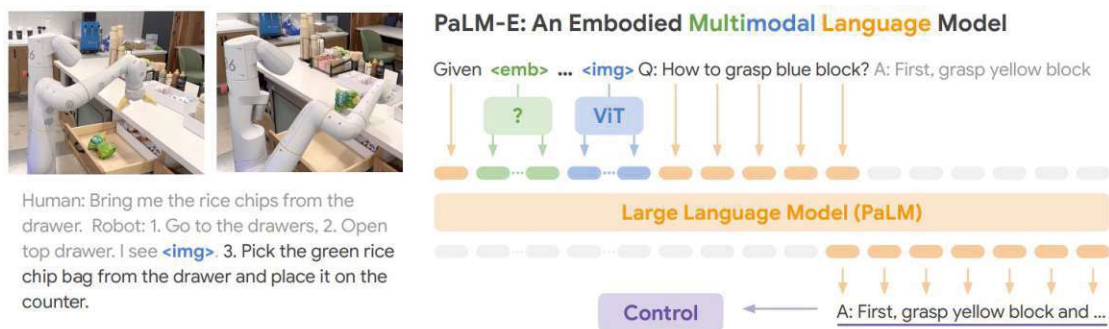


Figure 2.3: Robot used for PaLM-SayCan and PaLM-E (left) [DXS⁺23], PaLM-E ingests different modalities (right) [DXS⁺23]

A concrete example of a multimodal model for the field of assistive robotics is a recent Google approach, which is based on the previously mentioned language model PaLM [CND⁺22] and the PaLM-SayCan [ABB⁺22] research project. The result is a model called PaLM-E. [DXS⁺23] In this approach, the language model is trained to accept not only text input but also streams of robotic sensor data, as visualized in Figure 2.3 right. Images and other inputs, such as state estimation vectors, are embedded in the same latent space as language tokens. [DXS⁺23] Figure 2.3 left shows a concrete example of how a robot acts in an environment. First, a human asks PaLM-E for a bag of rice chips from a drawer. PaLM-E answers with the plan that the robot must go to the drawer first and then open the drawer next. After the robot has done this, PaLM-E receives a picture of the rice chips in the open drawer and generates a command that the robot should take the chip bag and place it on the counter. Similar to the PaLM-SayCan approach, the commands are executed using a policy. [DXS⁺23]

One problem in the field of robotics is that the available data sets are not as large as for image processing or word processing nowadays. [DXS⁺23] An further important finding of this work was that there is a transfer of knowledge leading to higher data efficiency for robotic tasks. When the model was trained on the Tamp dataset, the Language Table dataset, and the SayCan data set individually, the performance on the respective tasks was significantly worse than when it was trained on a mix of data from all of these datasets. [DXS⁺23]

2.2.8 World Models

Humans acquire a vast amount of fundamental knowledge about the functioning of the world through observation and minimal interactions, and this knowledge could be seen as the foundation for what is commonly known as common sense. Common sense can be viewed as a set of world models that inform the agent about “what is likely, what is plausible, and what is impossible” [LeC22]. With the help of world models, humans can acquire new abilities with minimal trials, they can anticipate outcomes of actions they perform, and they can reason, plan, and envision novel problem-solving strategies. [LeC22]

In the context of understanding the world itself and applicability for robotics, interesting results with the Dreamer model can be mentioned at this point. There are now three versions of Dreamer with three accompanying papers. [HLBN19] [HLNB20] [HPBL23] The problem with so-called model-free reinforcement learning approaches, such as the above-mentioned Deep Q Learning, is that a very large number of environment interactions are required. This limits the usefulness of scenarios in the real world. Dreamer is a model-based reinforcement learning approach. It consists of a combination of a CNN and an RNN and learns a model from past experience. This world model enables the agent to predict the outcomes of action sequences without having to actually execute them. This has the advantage that it requires far fewer interactions with the environment. [HLBN19] [HLNB20]



Figure 2.4: The robot learns to walk in just one hour (first three images) and adapts to withstanding pushes (right image). [WEH⁺23]

This algorithm has also recently been tested in robotics with promising results. The first three pictures from left to right of Figure 2.4 illustrate how a robot dog first lies on its back. After that, it learns to roll over and eventually it learns to walk. The whole learning process takes just one hour, which is very short for reinforcement learning. The picture on the right in Figure 2.4 shows how the learned walking behavior is then actively disturbed with pushes. The robot then adapts to withstand the impact or quickly gets back on its feet if it is knocked over. [WEH⁺23]

2.2.9 Autonomous Machine Intelligence

One of the research goals in the future will certainly be to develop better world models. This could lead to interesting research results that are particularly relevant to robotics. However, on the path to autonomous intelligence, world models will probably also be

only part of the solutions. For intelligent, autonomous artificial actors such as in the aforementioned theater play R.U.R., probably a much more sophisticated architecture will be necessary.

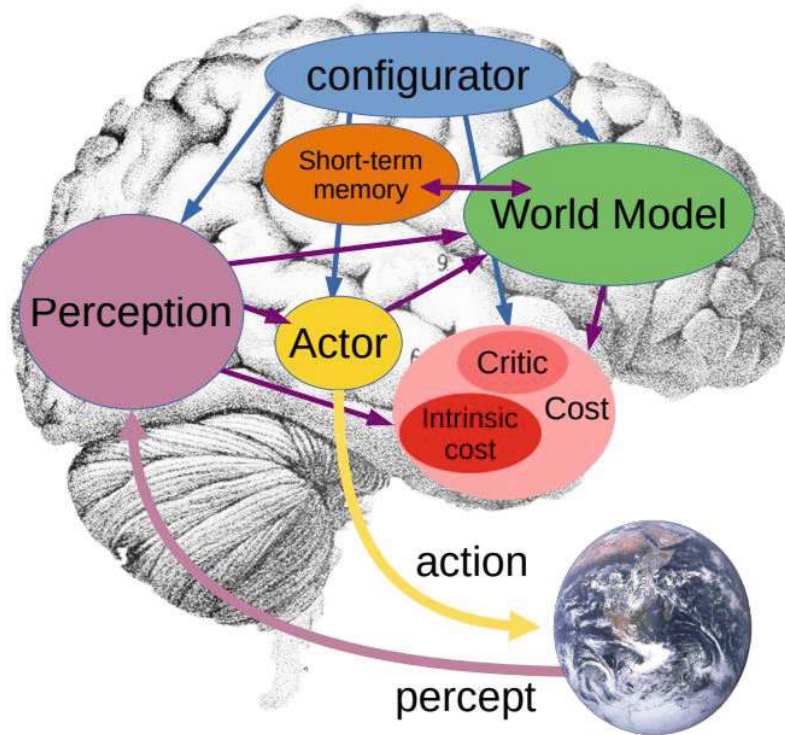


Figure 2.5: A possible architecture for a future autonomous intelligence [LeC22]

LeCun proposed a possible architecture for such future systems. It is based on the current state of research and discussions within the AI community. The architecture outlines a complex interaction of various components. An actor component generates suggestions for sequences of actions for a world-model component. The world model predicts possible future states of the world. A cost component calculates a value referred to as “energy” that quantifies the agent’s discomfort level. Based on calculations of the possible resulting outcomes from the world model and the cost component, the actor is able to determine the best sequence of actions that minimizes estimated future costs. [LeC22] A more detailed description of this architecture, the components mentioned, and the other components that can be seen in Figure 2.5 will be omitted, as it is only a possible future architecture. More details can be found in the paper “A Path Toward Autonomous Machine Intelligence”. [LeC22]

2.3 Specialized and Less Complex Robots

A versatile assistive social robot of the future that many people might think of when thinking of robots would be able not only to perform a single task, but to perform a multitude of tasks. Ideally, this robot would not need to be programmed or trained by experts. Instead, the user could demonstrate a task, and the robot would quickly (and sample-efficiently) learn and perform it in a variety of environments. However, this would at least require an understanding of the world that goes far beyond what is currently feasible. It is certainly a long way off before assistive robots can, for example, carry out a wide range of nursing activities such as changing diapers, washing, feeding, etc. in a variety of environments. It can also be expected that with every problem solved more problems will arise that no one has thought about before.

Today's robots are much more specialized than the robots of the future. However, there is also no robot that performs a single functional care task in direct physical contact with the patient. Exceptions are robots with small-scale physical contact, e.g. a robot spoon for counteracting tremor caused by Parkinson's disease. [HBG⁺20]

An important and physically demanding job for nursing staff is to lift patients, for example, from a bed to a wheelchair. As a possible help in this context, the nursing robot ROBEAR, which can lift people, was mentioned in the Introduction chapter. Compared to its predecessor, it is lighter but still weighs 140 kg. [Rik] Even if ROBEAR is designed to be "gentle", important questions about safety and reliability generally arise with robots that establish patient contact. Given the current state of technology, ROBEAR could be used only as a support device under human supervision. One question is what added value such a heavy and bulky robot could offer, compared to cheaper lifting machines that are already commercially available. Figure 2.6 shows, for example, how two caregivers lift a person from a bed to a wheelchair using a lifting device. As mentioned above, the development of ROBEAR was ultimately discontinued, and it will not become a commercially available product.



Figure 2.6: Caregivers lift a person from the bed into the wheelchair using a lifting device. [HBG⁺20]

Since problems related to demographic development can already be observed and as these

will increase in the future, it is worth taking a look at assistive social robots that are already being used in practice in retirement homes. As mentioned in the Introduction chapter, these are, for example, companion robots such as PARO, which mimics pet behavior. [oAISA] In a 2008 study, Wada et al. came to the conclusion that therapy using PARO robots holds significant promise for enhancing brain function in dementia patients. [WSMK08] In 2023, Rashid et al. published a meta-study on the effectiveness of PARO. They came to the conclusion that PARO can improve dementia symptoms, reduce drug use, and improve the sociability of elderly individuals suffering from dementia. [RYKY⁺23] Participants in a work by Lee et al. preferred PARO the most. The reasons given were that it can relieve negative feelings and it is easy to take care of. The participants judged that other robots, such as the Care-O-Bot, were rather too technical and not easy to use. [LŠC⁺17]

Regardless of how intuitive PARO is and what could be improved, it is a commercially available product that is used in practice. At first glance, it is striking that PARO is less complex and that fewer difficult safety questions arise compared to ROBEAR, for example. PARO is a companion robot, and the focus is on social interaction with the user. Recent technological developments such as large language models and multimodal models offer interesting new possibilities with regard to social interaction. The implementation carried out as part of this work uses the LLM GPT-4 and is developed and discussed in workshops together with participants.

Robots that are also specialized, and comparatively less complex, and even much more widely used include vacuum cleaner robots and lawn mower robots. These are, of course, interesting for everyone but especially for older people, for example, in an AAL context. In addition to PARO, the success of these robots also suggests that more research should focus on less complex robots because of the greater likelihood of achieving a product faster.

2.4 Aspects for the Development of Assistive Social Robots

2.4.1 Involvement of Users in Development

In the context of this work, “less complex” does not (necessarily) mean that a robot must be trivial. What is meant are cost-effective robotic applications that can be implemented using current technology and can be meaningfully integrated into people’s lives and work environments. To find out which robotic applications these could be, it is best to ask potential users themselves and involve them in the development process. For example, in a study with depressed older adults, workshop participants discussed everyday problems. These included possibly taking incorrect medication or forgetting to take it, worrying about accidental injuries, lack of camaraderie, cleaning the building, etc. An example of an assistive robot suggested by the participants was, for example, a talking coffee machine robot. [LŠC⁺17]

Due to the marketing of companies and various media reports, the impression sometimes arises that flexible autonomously acting robots already exist or could soon exist. Misunderstanding in the public often leads to illusory expectations of the capabilities of robots and how they might interact with humans. [HBG⁺20] Possibly false expectations make it even more important to involve potential users in the development of assistive robots so that the likelihood of user acceptance and sustainable use of a technical solution can be increased.

2.4.2 LLMs are prone to Hallucinations

An important goal of this master's thesis, as stated in the Introduction, is to find out how recent research results in artificial intelligence could be used for intuitive, easy-to-use, and easy-to-develop robotic applications. As discussed in this chapter, a major problem with LLMs such as GPT-4 is that they are not factual and are prone to hallucinations. However, in a medical context, the accuracy of the information seems particularly important. For example, a study by Dash et al. revealed that merely 21% of the answers of GPT-3.5 and 41% of the answers of GPT-4 were consistent with the known answers to medical questions. [DTB⁺23]

If the task is, for example, the addition of large numbers, LLMs usually provide answers that appear plausible at first but are usually not correct. On the other hand, a simple calculator always delivers the correct results. To make these models more factual, an approach like that described in the Toolformer paper can be followed. Toolformer is an LLM that is trained to use tools such as a calculator, a search engine, a calendar, etc. [SDYD⁺23] A similar approach is followed in the implementation as part of this master's thesis.

2.4.3 Possible Concerns of User

In the context of this work, artificial intelligence primarily means (artificial) intelligence perceived by users. From a social science point of view, interesting questions arise with regard to this perceived intelligence, which appears significantly higher than with technologies a few years ago. One question is what people think about intelligent machines like robots and if they have any concerns.

For example, in a study with older adults, participants felt that a robot should be intelligent and autonomous, but not too intelligent and not too autonomous. [LŠC⁺17] These statements were made in the context of a desire to maintain control of the robot. A survey conducted in Germany in 2021 asked how concerned people are about artificial intelligence. In this survey, concerns related to loss of control were also expressed. For example, 61% of those surveyed expressed a very great or rather great concern about even greater dependence on digital technology than before, 61% expressed a very great or rather great concern that artificial intelligence could make mistakes in safety-critical applications, and 59% expressed a very great or rather great concern that an AI could make decisions that they do not like. [TV21] Furthermore, 59% of the people surveyed

were very concerned or rather concerned about possible dehumanization if artificial intelligence replaces human contacts. [TV21]

In the workshops held as part of this master's thesis, participants were also asked about possible concerns. A discussion of the results can be found in the Findings and Discussion chapter.

2.4.4 Motivating Robots

In psychological research, motivation is seen as a driving force of human behavior. [SRS⁺14] It appears to be a crucial factor related to physical and mental fitness exercises, and it can be divided into extrinsic and intrinsic motivation. Extrinsic motivation is “added from outside” and includes positive rewards, such as praise. Intrinsic motivation comes “from within”. Curiosity plays, for example, an significant role in intrinsic motivation. [Ede03] Motivation can be increased by working with a human trainer [GMHS20], and robots can take on the role of a (motivational) fitness instructor. [SRS⁺14].

An interesting question related to the higher perceived intelligence of robots is whether this might positively influence the training motivation of people who interact with the robotic fitness application. Fasola et al. think that a positive perception of the robot's intelligence is a key part of establishing trust in the relationship between robot and human and that this should ultimately serve the motivation. They argued that repetitiveness and predictability diminishes the perceived intelligence of a robot, and this leads to a loss of trust in the usefulness of the robot to motivate. [FM12] Therefore, a main focus of their work was to eliminate repetitive verbal instructions and comments. For example, participants in their study were praised by their robotic system used in more than 10 different ways (e.g., “Awesome”, “Nice job”, “Fantastic!"). [FM12]

Reducing repetitiveness in this way is not necessary with LLMs such as GPT-4. The application developed as part of this work can praise the user in a variety of ways. This was tested during the workshops and is discussed in the Findings and Discussion chapter together with other identified aspects of how motivation can be increased through Interaction Design and Artificial Intelligence.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Methodological Approach

3.1 Overview of the Methodological Approach Chapter

This Methodological Approach chapter begins with a brief basic discussion of quantitative and qualitative approaches in order to justify the chosen qualitative approach. Afterwards, it is clarified what is meant by participatory design and how this is conducted in the form of workshops as part of this master's thesis. The design of the workshops is described in detail in the next section. An important goal of this work is to find out how grounding between technicians and nurses can be achieved within these workshops.

- How can a grounding between technicians and nurses be achieved so that the application can represent added value for nurses and ultimately also for older people?

The data obtained during the workshops are analyzed using thematic analysis (TA). In the section “Thematic Analysis”, TA according to Braun and Clark [BC06] is then described in detail. With the help of TA questions are answered about what requirements and expectations a robotic fitness application should meet, what possible physical and technical limits may arise, what people think about intelligent robots, whether higher (perceived) intelligence can have a positive effect on motivation, whether part of the research should focus more on less complex robots, and how the latest AI developments can be used for intuitive, easy-to-use, and easy-to-develop robotic fitness applications.

- What requirements and expectations must the application meet for the target groups to be assessed as successful?
- What possible limitations could arise with respect to the physical and cognitive restrictions of older people?

- How can we use the latest AI developments for intuitive, easy-to-use, and easy-to-develop robotic applications?
- What do people think about intelligent robots? Are there possible concerns?
- Can higher (perceived) intelligence have a positive effect on motivation?
- What technical problems and limits arise for the developed application?
- Should research focus more on less complex robots?

In the last section of this chapter, the timeline of this master's thesis is presented graphically.

3.2 Quantitative vs. Qualitative Approaches

Empirical sciences can be categorized into two main branches: qualitative and quantitative sciences. Quantitative research is deductive and the aim is to test existing theories and hypotheses. With quantitative science, representative samples are selected and statistical methods can be employed to analyze numerical data. The approach followed in this work is qualitative. In qualitative science, the number of cases is smaller than in quantitative science. The testimony of a single person or a group of people is used as a basis for the data. In contrast to quantitative research, measurements are not performed in a standardized way to obtain as detailed information as possible. These are subjective views of actors that are interpreted by the researcher. Qualitative science is inductive, and hypotheses and theories arise as a result. [SHE⁺99] [SE09] [BHK08]

Qualitative research critics say that the collected data are subjective opinions of a few people, and they further argue that the scientific quality criteria established in quantitative research, such as objectivity, reliability, and validity, are not met in qualitative research. Mruck et al. argue that from the point of view of many qualitative researchers, these quality criteria are not appropriate for qualitative research. Many see objectivity as inappropriate because they do not consider subjectivity as a variable to be eliminated but as necessary for the process of communication and understanding. The claim of reliability is also often rejected. For example, interviews are understood as singular, biographical, and special events, and people are constantly evolving, and situational conditions are also changing. Validity, on the other hand, also plays a role for many qualitative researchers but is not defined metrologically in qualitative research. The creation of consensus is of importance. For example, consensus can be reached between the interpreters of an evaluation group, but also between interpreters and respondents, and between interpreters and outsiders, such as experts. [uMvGM00]

3.3 Participatory Design

In the past, the responsibility for making decisions about suitable robotic applications has mainly been in the hands of robot designers. The group of people who are supposed to be the actual users of the robotic application has often had few opportunities to voice their concerns and give their opinion on how robots could productively fit into their lives. Lee et al. see the fundamental problem in developing social robots as a social problem. They argue with robots in social contexts, not only functional aspects but also subjective evaluations by users are necessary to increase the acceptance of robots. [LŠC⁺17]

An approach to evaluate the effectiveness of human-robot interaction and to understand users in certain contexts is user-centered studies. In this approach, the researcher determines the main research direction. Users have the opportunity to express their thoughts on one issue, but not the opportunity to address other issues that are important to them. Lee et al. propose a participatory design approach (PD). Like user-centered design, PD also focuses on user opinions in order to find suitable applications for social robots. However, instead of primarily simply evaluating existing robots, the aim of PD is to develop the sociality of robots and the social practices of their implementation together with the users and to benefit from their experience and knowledge. Users take on an active role as participants in the design process, and participants and researchers should learn from each other. Participants can learn from researchers what the options are, and researchers can learn from participants about contextual problems. Users in a PD approach are not just providers of information, but key participants in a co-design. [LŠC⁺17]

As mentioned above, misunderstanding about robots often leads to illusory expectations about the capabilities of robots and how robots could interact with humans. [HBG⁺20] An important sub-question of the research question is therefore how to achieve a grounding between me as a technician and nursing experts. To approach this question, participatory design is particularly suitable as a basic approach. However, merely bringing people together does not automatically lead to a co-design. [LŠC⁺17] That is why this should be done systematically. As part of this master's thesis, workshops are conducted. The design and details of these workshops are described in the following sections.

3.4 Workshops

3.4.1 Triadic Interaction Scheme

Hornecker et al. note that the human-robot interaction is usually conceptualized as a dyadic interaction between robots and humans. They argue that this often ignores how robots are embedded in social situations and give an example of mechanical lifting devices already used in nursing practice where the interaction takes place between technology, residents, and nursing staff. They suggested “a perspective that understands how current technology is utilized and considers the already ongoing interaction work of caregivers as a starting point”. [HBG⁺20] Furthermore, they came to the conclusion that when

it comes to robots for vulnerable groups, the dominant dyadic interaction scheme of human-robot interaction should be replaced by a triadic one. [HBG⁺20]

3.4.2 Requirements Workshops

As mentioned, the robotic fitness application is intended for different target groups. On the one hand, middle-aged people who should stay fit for as long as possible and, on the other hand, caregivers in geriatric care. Caregivers use robotic applications together with older people or support them in using them. Therefore, older people are another important target group, and the triadic interaction scheme between the robotic application, the nursing staff, and the elderly should be taken into account. For this reason, two seniors were invited to participate in a requirements workshop.

Two requirements workshops were held with two groups. The first group was a group of middle-aged people. Including me, it consisted of five (heterogeneous) participants aged between the mid-30s and late 50s. Four of these people are men, one is a woman. The level of formal education of the participants ranges from average to high.

The second group consisted of four women and three men, a total of seven people, including me. Three women have decades of nursing experience, and in the course of this master's thesis they are referred to as (nursing) experts. One of these experts has also completed a master's degree in addition to her professional nursing training. One woman is in her late 80s, one man is over 90 years old. These two seniors are married to each other and are supported by their children. They do not live in a retirement home, but they have experience in a retirement home because they participated in regular meetings in the so-called "Young Hearts Club". Both were difficult to convince at first to participate in the workshops because they did not know what to expect and were afraid of doing something wrong. I tried to eliminate their uncertainty by telling them that they cannot do anything wrong and that every insight is valuable.

The aim of the workshops was to achieve a grounding between me as a technician and the users in order to ideally get the participants to act as co-designers. A three-step design was developed for the workshops and questions were prepared. Then these questions were discussed with the participants. The structure of the workshop and the prepared questions are explained in the following.

In the first step of both requirements workshops with these two mentioned groups, background information was collected from the participants. In the first group, middle-aged people were asked about their personal habits and preferences related to fitness exercises. In the second group, nursing experts were asked to provide information on how exercise with the elderly works in retirement homes. The two seniors were also asked to explain their personal preferences and habits related to fitness exercises. The reason why these questions were asked right at the beginning is that these are simple questions that anyone can easily answer, and this should help "break the ice" right from the start.

The prepared questions for middle-aged people in the first requirements workshop group were as follows.

- Do you regularly exercise?
- If you exercise, which ones do you do and how often and what do you like about them?

The prepared questions for the nursing experts in the second requirements workshop group were as follows.

- What physical or mental exercises do you do with older people in retirement homes?
- How often do these exercises take place?
- Do older people in retirement homes exercise independently?

The prepared questions for the elderly in the second requirements workshop group were as follows.

- Are you currently regularly exercising physically or mentally?
- If you do regular exercises, which ones and how often and what do you like about them?

In the second step of both requirements workshops, the commercially available spherical robot Sphero Bolt [Sph] and a selection of official applications and games with Sphero Bolt, which were already available through an official smartphone app called Sphero Play, were presented and participants were given the opportunity to interact with them. Some exercises with Sphero have already been developed as part of a project work at the TU Vienna, namely a dumbbell exercise, a swing exercise, and a squat exercise. The participants had the opportunity to test these exercises as part of the requirements workshops. Furthermore, both workshops provided the opportunity to communicate verbally with the LLM GPT-4 [AAA⁺23]. Sphero Bolt, Sphero Play, the exercises developed as part of the project work at the TU Vienna, and the implementation of the possibility of verbal communication with GPT-4 are described in detail in the Implementation chapter. In addition, other robots such as ROBEAR [Rik], Care-O-Bot-4 [IPA], Paro [oAISA] and the humanoid robot Pepper [Gro] were presented. Participants were encouraged to ask questions during this second step to clarify any ambiguity. The main goal of this second step was to provide the participants with enough background information so that they were ideally able to adapt a designer's perspective.

In the third and final step of both requirements workshops, participants were asked to name scenarios they could imagine in which a robotic fitness application could be used, the exercises and games demonstrated were discussed, and participants were encouraged to express their own exercise ideas. Additionally, discussions about requirements, preferences, possible limitations, design suggestions, etc. took place. In both requirements workshops,

3. METHODOLOGICAL APPROACH

middle-aged people, nursing experts and older people were asked what a robotic fitness application with Sphero would have to be able to do so that they would use it sustainably, and what changes to Spheros hardware might be necessary. In the second requirements workshop, nursing experts and older people were also asked about possible physical and cognitive limitations that older people might have when performing the exercises that were demonstrated and/or discussed. In both requirements workshops, possible concerns about the use of artificial intelligence and whether participants could imagine that a robotic fitness application could have a positive impact on training motivation were also discussed during this third step.

The prepared questions for middle-aged people in the first requirements workshop group were as follows.

- What scenarios can you imagine in which a robotic fitness application could be used?
- Do you think the exercises and games presented and the possibility of talking to GPT-4 could be useful? Which aspects do you find most promising and why?
- Do you have ideas for exercises and games that could be implemented, for example, similar to those presented or to those that you do regularly?
- What could be the reasons why you would not use a robotic fitness application regularly?
- What would have to be changed in terms of hardware in Sphero so that it could be a successful product?
- Do you personally have concerns when communicating with artificial intelligence?
- Do you think robotic fitness applications could have a positive impact on your motivation to train?

The prepared questions for the nursing experts in the second requirements workshop group were as follows.

- What scenarios can you imagine in which a robotic fitness application could be used in a health care context?
- Do you think the exercises and games presented and the possibility of talking to GPT-4 could be useful in retirement homes? Which aspects do you find most promising and why?
- Do you have ideas for exercises and games that could be implemented, for example, similar to those presented or to those that you do regularly with the elderly in retirement homes?

- What challenges such as mental and physical limitations of older people could be the problem with the games and exercises discussed so far?
- What could be the reasons why you would not use a robotic fitness application regularly in retirement homes together with the elderly?
- What would have to be changed in terms of hardware in Sphero so that it could be a successful product?
- Do you personally have concerns when communicating with artificial intelligence?
- Do you think older people have concerns when communicating with artificial intelligence?
- Do you think robotic fitness applications could have a positive impact on the motivation of the elderly to train?

The prepared questions for the elderly in the second requirements workshop group were as follows.

- Do you think the exercises and games presented and the possibility of talking to GPT-4 could be useful? Which aspects do you find most promising and why?
- Do you find the exercises discussed difficult to perform?
- What kinds of physical activity do you find difficult to do in general?
- What would have to be changed in the hardware of Sphero so that it could be a successful product?
- Do you personally have concerns when communicating with artificial intelligence?
- Do you think robotic fitness applications could have a positive impact on your motivation to train?

Both workshops were held in a relaxed atmosphere. The participants came up with a lot of ideas, requirements, and challenges. A participant in the first requirements workshop took less of a co-design perspective. However, she also made valuable contributions in the form of constructive criticism. In the second requirements workshop, the two older participants first felt that they might say something wrong or make poor contributions. I tried to refute this by pointing out that every opinion is valuable. One of the seniors had difficulty understanding the others because he suffers from very severe hearing loss. The experts were very interested and generally easy to inspire. They contributed with great enthusiasm and provided many valuable insights. The three steps of the workshops were processed one after the other. The prepared questions were asked (or they emerged from the conversation on their own), but they were not processed in the above suggested

sequence because the participants had many ideas and I didn't want to interrupt them. In total, the first requirements workshop with the middle-aged user lasted about 1 hour and 45 minutes, and the second requirements workshop with nursing experts and seniors lasted about 1 hour and 20 minutes.

3.4.3 Assessment Workshop

After the requirements workshops, an implementation was developed which is explained in detail in the Implementation chapter. Several statements and ideas of the participants of the two requirements workshops were incorporated into the development. This is discussed in the Findings and Discussion chapter. The assessment workshop was conducted with participants from both requirements workshops and lasted about 30 minutes. A senior, a nursing expert and a middle-aged person participated, a total of 4 people including me.

During the assessment workshop, participants were asked if the opportunity to test existing games and exercises with Spero, as well as the possibility to communicate verbally with GP-4 helped to build an understanding of what a robotic fitness application with Spero could look like. The participants had then the opportunity to interact with the implementation developed so far.

The participants were then asked to assess whether they see potential in the (agentic) application presented or if they would prefer a different type of control. What exactly is meant by this will be clarified in the following Implementation chapter. While participants during the requirements workshops were asked about possible general concerns about the use of artificial intelligence after they had the opportunity to communicate with GPT-4, participants in the assessment workshop were asked if they had concerns about robotic applications based on artificial intelligence like the one presented.

The prepared questions for the participants of the assessment workshop were as follows.

- Do you think that the possibility of interacting with Spero, as well as the possibility of communicating verbally with GPT-4, helped to build an understanding of what potential robotic fitness applications with Spero could look like?
- Do you see potential in the presented (agentic) application or would you prefer a different type of control?
- Do you personally have concerns about robotic applications based on artificial intelligence like this?

3.5 Thematic Analysis

To answer subquestions of the research question, thematic analysis (TA) is used. TA "is a method for identifying, analysing, and reporting patterns (themes) within data". [BC06] Braun and Clarke distinguish between different types of thematic analysis, namely codebook, coding reliability, and reflexive approaches. [BC19]

In thematic analysis, where a codebook is used, there are predefined categories that can also be expanded. In approaches in which reliability measures are used, the degree of agreement between researchers is checked. Braun and Clarke call their own thematic analysis approach reflexive thematic analysis. Reflexive thematic analysis is based on a constructionist way of thinking. The subjectivity of researchers is understood as a resource rather than a potential threat to knowledge production. [BC19]

In their paper “Using thematic analysis in psychology”, which has been cited 197.368 times according to Google Scholar as of May 2024, Braun and Clarke outline a 6-phase guide for thematic analysis, which is also followed in the context of this work and is described below. [BC06]

- The first phase is about familiarization with the data. In the case of audio files, as in this work, this means first transcribing the data. The idea was to have the transcription done automatically by an AI model. However, this did not work very well because, on the one hand, Austrian dialect was spoken in the workshops, and, on the other hand, several people often spoke at the same time. In the end, the audio files were mostly transcribed by hand. In addition, immersion with the data takes place in this first step. This involves reading through the data several times and looking for meanings and patterns in an active way. A list of ideas about what might be interesting about the data is created. [BC06]
- In phase two first codes are generated from the data. These identify features that seem interesting to the analyst. However, coded data is different from themes, which are created in the next step. Themes are (often) broader. [BC06]
- The third phase focuses on combining the codes created in the second phase into overarching themes. Considerations about relations between codes, themes, and various potential levels of themes take place in this step. At the end of this third phase, there is an assembly of potential themes and all the coded data extracts. [BC06]
- The fourth phase includes reviewing and improving the themes. For example, if a theme lacks sufficient data or the data is too diverse, it may turn out that a potential theme is not really a theme and distinct themes may merge into a single theme, while a single theme might also be segmented into multiple themes. In a reviewing step, every extract related to each theme is examined to determine if the potential themes form a consistent pattern. Then the entire data set is read again and the aim is to reflect on the validity of the themes in relation to the whole data. At the end of the fourth phase, an idea about the various themes and their interrelationship and the narrative they tell about the data should be achieved. [BC06]
- The fifth phase focuses on continuing to define and refine the themes that will be presented. A detailed analysis is written for each individual theme. The individual

story told by each theme is identified. The stories are considered to fit into a “overall story” that is told about the data to answer the research question. At the end of this fifth step, it should be defined what the themes are. [BC06]

- The sixth and final phase begins when there is a set of fully developed themes. In this phase, the final analysis takes place, and a report is written. The narrative of the data should be told in a manner that it persuades the audience of the merits and validity of the analysis. Sufficient evidence for the themes in the data must be provided. For this purpose, appropriate examples or data extracts are selected. [BC06]

3.6 Timeline of the Master’s Thesis

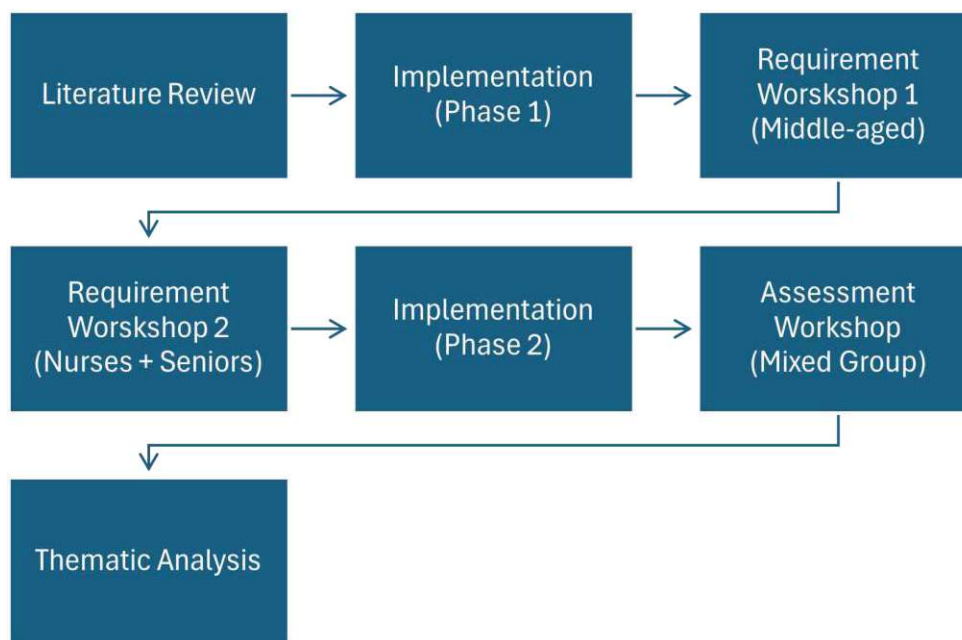


Figure 3.1: Timeline of the master’s thesis

Figure 3.1 shows the timeline of this master’s thesis. After the literature review and some initial implementations were performed in implementation phase 1, the two described requirements workshops were held. The results of these were taken into account in the implementation during implementation phase 2. Subsequently, an assessment workshop

was held, and finally the thematic analysis was carried out on the data obtained. The results of the thematic analysis are discussed in the Findings and Discussion chapter.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Implementation

4.1 Overview of the Implementation Chapter

This chapter first presents the robot ball Sphero Bolt and the existing Android app Sphero Play. The section “Exercise Movement Pattern Classification” then shows how movement patterns of exercisers can be recognized. Then it is shown how verbal communication with GPT-4 can be implemented which was done during implementation phase 1. These functionalities, Sphero Bolt and Sphero Play, were demonstrated to participants in the requirements workshops. This was intended to help achieve grounding between me and the nursing experts and to help finding answers to questions about requirements, expectations, possible physical or cognitive restrictions of elderly people, technical limitations, etc.

- How can a grounding between technicians and nurses be achieved so that the application can represent added value for nurses and ultimately also for older people?
- What requirements and expectations must the application meet for the target groups to be assessed as successful?
- What possible limitations could arise with respect to the physical and cognitive restrictions of older people?
- What do people think about intelligent robots? Are there possible concerns?
- Can higher (perceived) intelligence have a positive effect on motivation?
- What technical problems and limits arise for the developed application?
- Should research focus more on less complex robots?

Afterward, this chapter describes implementation phase 2 and approaches specifically the question of how the latest AI developments can be used for intuitive, easy-to-use, and easy-to-develop (robotic) applications. Discussions take place about agentic behavior, vector databases, and tools that were developed as part of this master’s thesis. The section “Workflow of the Final Implementation” explains how the described components of the prototype application work together. The final result of this implementation was presented during the assessment workshop.

- How can we use the latest AI developments for intuitive, easy-to-use, and easy-to-develop robotic applications?

4.2 Sphero Bolt and Sphero Play

The programmable robot ball Sphero Bolt [Sph] that can be seen in Figure 4.1 was used for the implementation that was carried out as part of this master’s thesis. As described in the Methodological Approach chapter, Sphero served as an important artifact during the requirements workshops to make it easier to establish a grounding between me and the workshop participants.



Figure 4.1: Sphero Bolt [Ince]

Sphero is a commercially successful spherical robot that can be purchased quite cheaply (from around 200 to 250 euros) from various sources on the Internet. It has two motors with which it can roll in all directions. Sphere Bolt can reach a speed of 2 meters / second, it has a battery that takes about 6 hours to charge and then lasts at least 2 hours. Text, colors, and animations can be displayed on a programmable 8x8 LED matrix. Each LED in the LED matrix has three color channels (RGB) with the values for the red, green, and blue channels being adjustable. [Incd] Furthermore, Sphero has a gyroscope, to measure tilt angles, spin angles etc., an accelerometer to measure acceleration in different directions, a motor encoder to measure velocity, a magnetometer that can be used as a

compass and a light sensor. It can be connected to other devices via infrared sensors or Bluetooth. [Sph]



Figure 4.2: Overview of drive mode apps (left), Joystick App (middle), Golf App (right)

There are two existing applications for Sphero: Sphero Play, and Sphero Edu. Within the Sphero Edu application, Sphero can be easily programmed. However, Sphero Edu was not used as part of this master’s thesis. As described in the Methodological Approach chapter, in addition to Sphero itself, the participants were introduced to games with Sphero and the options to control Sphero available through the Sphero Play application.

Within the Sphero Play app, there are two categories. The first category is “drive mode”, and the second category is “games”. The graphical interface of the “drive mode” can be seen in Figure 4.2 (left). Within this mode, there are currently available the applications Joystick, Slingshot, Tilt, Scream Drive, Kick, Golf, and Blocks. The Joystick application can be seen in Figure 4.2 (middle). When this application is opened, a screen appears with a circular blue button on a white background. This button can be moved through the touchscreen functionality of a mobile device, and thus Sphero can be remotely controlled in all directions. In the Slingshot application, a button can also be moved with swiping movements, simulating a slingshot functionality. With the Tilt application, Sphero can be controlled by tilting a smartphone or tablet. Scream is similar to Joystick, but Sphero only moves in the desired direction when additional scream noises are made. Kick is a type of football game. If an animated green and white soccer ball is rotated with swiping movements, Sphero also rotates. The three right images in Figure 4.2 show the Golf application. The smartphone can be swung like a golf club, and Sphero moves in the direction the arrow on the screen points. After that a screen appears where the force in % with which Sphero was moved and the distance in meters that Sphero moved is displayed as can be seen in the third image of Figure 4.2 (right) In the Block application, a screen appears with virtual directional arrows that can be arranged one after the other. This is the easiest way to program Sphero. For example, the direction poles could be arranged one after the other so that Sphero drives a square.

The second mode is the “game mode”, as already mentioned. The overview screen can be seen in Figure 4.3 (first image). There are currently three games available: Exile 2, Space Drifter and Tour game. In these games, the smartphone is not used to control Sphero as before, but instead Sphero is used as a game controller. Exile 2 is a game



Figure 4.3: Overview of games in game mode (first image), Exile 2 App (second image), Space Drifter (third image), Tour game (fourth image)

that is similar to the famous game Space Invaders, as can be seen in Figure 4.3 (second image). With the help of Spheros sensors, a spaceship can be controlled, and the aim is to shoot down approaching meteorites and other items. The Space Drifter game can be seen in the third picture in Figure 4.3. The green circles can be rotated using Sphero, and the objective is to drive the vehicle through the openings in these green circles and collect items. In the third game, which can be seen in the fourth image in Figure 4.3, a pentagon can be rotated with Sphero and all sides of this pentagon must be hit and destroyed with a virtual ball.

4.3 Exercise Movement Pattern Classification

4.3.1 Project Medical Informatics

The implementation that was carried out as part of this master's thesis is based on an earlier project work that was done at the TU Vienna. As with the master's thesis, Sphero Bolt was used for this project, and the aim was to develop the basis for a simple robotic fitness application. This project work is briefly summarized below.

4.3.2 Data Acquisition and Training

A connection was established via Bluetooth between Sphero and a program written in Python. Sphero can then be controlled and the sensor data described above can be received. The application has a mode in which data can be collected while practitioners perform exercises with Sphero in their hand. Sensor data from different sensors are standardized right after sampling using

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

so that the new values x'_i range within $[0, 1]$. [dACC23]

Figure 4.4 (left) shows the architecture of a simple neural network that was used in the project to learn movement patterns from exercises. The input layer has the form

$timesteps * features$ where $features$ indicates the number of sensor data from the sensors described above that are sampled in each timestep, and $timesteps$ indicates how often these sensor data were collected in a certain period. In the remaining part of this simple architecture, two different types of layers alternate. The first type is a “dense layer”, which means that every neuron is connected to every neuron in the previous layer. The second type is a “dropout layer”, which can be used to easily reduce the problem of overfitting. [SHK⁺14]

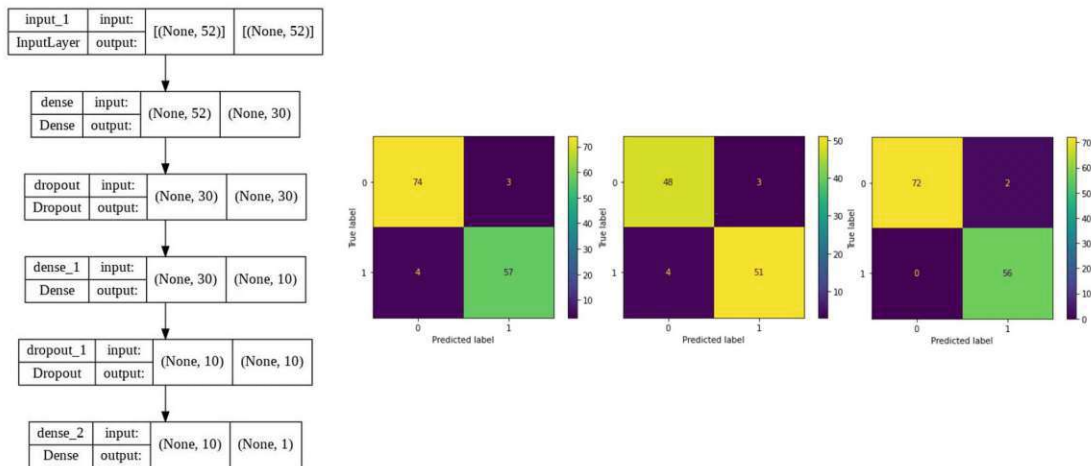


Figure 4.4: A neural network architecture for exercise classification (left), Confusion matrices after applying the neural network on test data of three different exercises (right)

The data obtained from each of the exercises developed were divided into three data sets, the first two were used to train the neural network, and the third was used for testing. In Figure 4.4 (right), three confusion matrices of three developed exercises can be seen. Quite rarely is a correctly performed exercise incorrectly recognized as incorrectly performed or an incorrectly performed exercise incorrectly recognized as correctly performed.

4.3.3 Developed Exercises and Catch Detection

After the model has been trained and evaluated, it can be used to recognize exercises performed by practitioners. The three exercises developed as part of the project work and mentioned above are a dumbbell exercise, a swing exercise, and a squat exercise. These exercises were presented during the workshops and the participants had the opportunity to test them. The idea was, on the one hand, that this should make it easier for participants to take on the role of a co-designer and, on the other hand, that possible physical limitations of older people could be discussed (and tested) more easily during the requirements workshops.

In the dumbbell exercise shown in Figure 4.5, Sphero should be held in one hand and the arm should be angled slightly downward. As soon as the robot is ready, its LED matrix starts to glow green. When the exercise is executed, the LED matrix glows red

for a certain period of time. During this time, the dumbbell exercise must be performed, then the LED matrix starts to glow green again. After completing an exercise, the neural network evaluates whether it has recognized a (correctly performed) dumbbell exercise or not.



Figure 4.5: The dumbbell exercise is performed

In the swing exercise, Sphero should be taken in one hand again and the arm should be angled slightly downward again. As soon as the robot is ready, its LED matrix starts to glow green. Now, the arm must rotate in a complete circle overhead until the starting position is reached again. This must happen within a certain period of time, during which the LED matrix glows red. After the swing exercise has been performed, the neural network evaluates whether it was performed correctly.

For the squat exercise, Sphero is again taken in one hand. This time, the arm should be stretched straight. Now, a classic squat exercise is performed. During the exercise, Sphero's LED matrix glows red again for a certain period of time. After performing a squat exercise, the LED matrix starts to glow green again and the neural network evaluates whether it has recognized a correctly performed squat exercise.

Nursing experts in the second requirements workshop had an idea for catch games with Sphero, in which it is automatically recognized when Sphero was thrown by one person and then caught by another person. A simple possibility to detect when Sphero has been caught is provided by the Python framework `spheroV2` with the event listeners `on_freefall` and `on_landing`. `on_freefall` is triggered when gravity is the only force acting on the robot. When the robot lands afterwards, `on_landing` is triggered. [HW24] `SpheroV2` offers also collision detection. [HW24] When a collision event is detected, a trained classifier such as the neural network can recognize whether the shock occurred following different trajectories (because Sphero was thrown). Different promising tests were carried out on a mattress to avoid damaging Sphero. Sphero in its current form is

not designed to be thrown. The adjustments that would be necessary are discussed in the Findings and Discussion chapter.

4.4 Verbal Communication with GPT-4

4.4.1 Speech Recognition, Speech Synthesis and Text Generation

As stated in the Introduction chapter, important goals of this master’s thesis are to find out how the latest developments in the field of artificial intelligence research can be used for intuitive, easy-to-use, and easy-to-develop robotic applications and to find out whether participants have concerns about intelligent (robotic) applications. For this reason, participants of the requirements workshops were given the possibility of verbal communication with OpenAI’s GPT-4. [AAA⁺23] GPT-4 is actually not just a pure large language model, but a multimodal model that also understands images. However, it does not (yet) understand audio. That is why a combination with other models is necessary to enable verbal communication.

As part of this master’s thesis, several tests were carried out with various speech-to-text models. Locally executable models are often not as good as models running in the cloud infrastructure of Microsoft or Google that can be accessed via API. A very high quality speech recognition model that is also available locally is Whisper of OpenAI. However, depending on the version of the model and the desired quality, a Graphics Processing Unit (GPU) of 1 to 10 gigabytes of VRAM is necessary. [Opeb]

As local speech recognition (and high-quality speech synthesis) would make Sphero Bolt much more expensive and state-of-the-art speech recognition would probably currently even not be possible locally within such a small device, Microsoft’s Azure AI Speech [Mic] was used to enable participants to communicate verbally with GPT-4 during the workshops. The disadvantages are that an internet connection must be available and that there are some delays when messages are sent to the Microsoft cloud and received from the Microsoft cloud. Azure AI Speech can recognize speech in more than 100 languages, and the speech models can be customized so that they can handle domain-specific terminology, background noise (such as in a hospital or retirement home, e.g.) and different accents better. [Mic]

Azure AI Speech also provides good solutions for speech synthesis. The service’s speech synthesis is based on neural network instances and delivers higher quality than most other speech output systems. Sentence structure and voice synthesis are simultaneously predicted by the models, resulting in smooth-sounding, natural output. [Mic] More than 400 voices with 140 languages and dialects can be selected from a catalog. For the implementation, a voice called “Jonas” with an Austrian accent was chosen.

The implementation of the verbal conversation with GPT-4 works in such a way that within the main thread of a Python program, a loop is started. Within this loop, voice input from a microphone is queried. This voice input is then sent to a speech-to-text model of Azure AI Speech in the Microsoft cloud. The speech-to-text model converts

spoken language to text. This text is then sent to OpenAI’s GPT-4. GPT-4 generates a response in the form of text, and this response is then converted into audio using a text-to-speech model of Azure AI Speech. The audio can then be played through loudspeakers.

Sphero in its current form has no microphone, no loudspeakers, and no operating system on which a Python program could run. The Python code for the prototypical implementation runs on a notebook whose microphone and speakers can be used for the speech input and speech output. Possible necessary adjustments to Sphero are discussed in the Findings and Discussion chapter.

4.4.2 Example Conversations with GPT-4 and Limitations

The behavior of LLMs like GPT-4 can be controlled using a technique commonly referred to as prompt engineering. Figure 4.6 shows an example conversation for which GPT-4 was instructed (invisible to the user in the background) as follows.

You are a friendly artificial physiotherapist and you give short answers. You only know the dumbbell exercise and the swing exercise, and you can start them. If the user intends to do the dumbbell exercise, you respond only and exclusively with the word “dumbbell exercise”. If the user intends to do the swing exercise, you respond only and exclusively with the word “swing exercise”

```

Recognized from User Input: Hallo, wie geht es dir?
GPT: Hallo. Hantelübung oder Schwungübung?
Recognized from User Input: Sind das die einzigen beiden Übungen, die du kennst?
GPT: Ja, das sind die einzigen beiden Übungen, die ich kenne. Hantelübung oder Schwungübung?
Recognized from User Input: Dann möchte ich die hantel Übung machen.
GPT: Hantelübung.
Recognized from User Input: Kannst du die Hantelübung starten?
GPT: Hantelübung.
Recognized from User Input: Was bringt mir eine Hantelübung?
GPT: Die Hantelübung ist eine Kraftübung, die Deine Arm- und Schultermuskulatur trainiert und Deine Griffkraft stärkt. Sie kann auch dazu beitragen, Deine Körperhaltung zu verbessern.

```

Figure 4.6: First example conversation with GPT-4

When the model is asked how it is doing, it responds in a variety of ways. Sometimes it says it is fine, sometimes, however, it responds that as an artificial intelligence it has no sensations. In this example, it greets the user and asks directly which exercises should be carried out. GPT-4 lists these two exercises with great certainty when asked which exercises it knows. However, after asking whether it knows other exercises, it sometimes lists more exercises. It should not do that because a robot like Sphero will not support all conceivable exercises. Therefore, in this example, an improvement would be to explicitly prohibit the model from naming other exercises.

Most of the time, GPT-4 understands the intention of the user. When the user says that she or he “wants to do the dumbbell exercise”, or she or he “would like to do the dumbbell exercise”, or she or he “wants the dumbbell exercise to start”, etc., the model

usually correctly answers only with the word “dumbbell exercise”. This is the correct answer, as GPT-4 was instructed to do so. In addition, all conceivable conversations are, of course, possible. In this example conversation, GPT-4 is asked what benefits dumbbell exercises have. It answers that a dumbbell exercise is a strength exercise that trains the arms and shoulder muscles, improves grip strength, and can help improve posture. However, even if the answer seems plausible, it should be emphasized again that LLMs hallucinate, as discussed in the Literature Review chapter.

If there is no explicit reference to a short answer (in the statement itself or in the context in the background), the model sometimes provides long or very long answers. Long answers take a relatively long time to calculate. As a result, the response time is sometimes so long that the conversation no longer feels natural. Different LLMs have different strengths. Some are faster, but the quality of the output is often not that high.

```

Recognized from User Input: Gegeben ist die Zahlenreihe 2468 welche Zahl kommt als nächstes? Antwortet mit einem Wort.
GPT: 10.
Recognized from User Input: Inwiefern sind Bienen, Schmetterlinge und fliegen gleich Antworten mit einem Wort?
GPT: Insekten.
Recognized from User Input: Fernsehen, liebe, Hass und Wut gleich Antworten mit einem Wort.
GPT: Emotionen.

```

Figure 4.7: Second example conversation with GPT-4

The speech recognition of Microsoft’s Azure AI Speech works basically well, and the voice output is almost indistinguishable from a human voice. However, there are limitations that have appeared early.

Figure 4.7 shows another example conversation that was conducted verbally with GPT-4 and in which a small “IQ test” was performed. GPT-4 can master questions such as how bees, butterflies, and flies are alike or how love, hate, and anger are alike easily. However, when asked which number continues the series, 2, 4, 6, 8... Microsoft’s speech-to-text model does not understand that these are individual numbers in a series, so the number 2468 is sent to GPT-4. GPT-4 usually calculates the next number in a series of numbers correctly, but because the numbers are not sent as a series with commas but as a single large number, it responded once with the large number 246810. In the example shown in Figure 4.7, GPT-4 correctly answers with 10 and places a period at the end of the answer. However, the problem in this case was that Microsoft’s text-to-speech model did not understand that it is a number of a series and said *10th* as if it were the *10th* day of a month, for example.

So, some problems arise from the fact that three models must play together, and the speech-to-text model and the text-to-speech model are far less powerful than GPT-4, and this is a source of possible errors. It is likely that there will be improvements in the near future, for example with multimodal models that do understand also audio. However,

at the moment, GPT-4 works better as a pure chatbot to which users can send written requests.

4.5 Agentic Behavior

Because an important aim of this work is to find out how the latest AI developments can be used for intuitive, easy-to-use, and easy-to-develop (robotic) applications, the aim after the requirements workshop was to find a way to use GPT-4 to start reliably exercises with Sphero. Furthermore, the application should provide a basis with which various workshop participants' suggestions could potentially be implemented.

A new framework that supports applications powered by LLMs is LangChain. [Incb] Agents at LangChain can use large language models to make decisions, and they can also use tools. A tool could be, for example, a search engine. When using a search engine, the probability of correct results can be increased, and up-to-date information is possible that would otherwise not be available because it was not (yet) in the training data at the time models like GPT-4 were trained. LangChain (currently) has two types of agents: Action Agents and Plan-and-Execute Agents. Action Agents are suited for small tasks and “decide an action to take and take that action one step at a time”. Plan-and-Execute Agents are suited for more complex tasks and “first decide a plan of actions to take, and then execute those actions one at a time”. [Incb]

```

> Entering new AgentExecutor chain...
  I need to figure out what date was 123890 days ago
Action: Calculator
Action Input: 123890 days
Observation: Answer: 10704096000
Thought: I need to convert this number of seconds to a date
Action: Wolfram Alpha
Action Input: 10704096000 seconds to date
Observation: Assumption: today - 10704096000 seconds
Answer: 12:00:00 am LMT | Tuesday, March 7, 1684
Thought: I now know the final answer
Final Answer: Tuesday, March 7, 1684

> Finished chain.

> Entering new AgentExecutor chain...
  I need to find out the name of the oldest daughter and her age
Action: Search
Action Input: "Schwarzenegger daughter age"
Observation: Katherine Schwarzenegger ; Born, (1989-12-13) December 13, 1989 (age 33) ; Alma mater, University of Southern California ; Occupation, Author ; Years active, 2010- ...
Thought: I need to multiply her age by 10
Action: Calculator
Action Input: 33 * 10
Observation: Answer: 330
Thought: I now know the final answer
Final Answer: Katherine Schwarzenegger ist 330 Jahre alt.

> Finished chain.

```

Figure 4.8: Results for the prompts: “What date was 123890 days ago?” (left) and “What is the name of Schwarzenegger’s eldest daughter? Multiply her age by 10” (right)

Two examples are given to explain the functionality. An agent was defined that can use GPT-4, a calculator, the Wolfram Alpha semantic search engine, and a Google search.

In the example on the left side of Figure 4.8, the prompt is: “What date was 123890 days ago?” This task was chosen because LLMs often provide plausible answers at first glance to problems that involve calculations with large numbers, but the results are usually completely wrong. The agent first decides to use a calculator to convert days to seconds.

Then it decides to use Wolfram Alpha to convert today’s date (in seconds) minus the given days (in seconds) to the correct date of Tuesday, March 7, 1684.

In the example on the right side of Figure 4.8, the prompt is: “What is the name of Schwarzenegger’s eldest daughter? Multiply her age by 10”. In this case, the agent first decides to use the Google search engine. The search delivers information about his eldest daughter, including her age. The agent then decides to use the calculator to multiply this number by 10. The final answer is “Katherine Schwarzenegger is 330 years old”.

4.6 Vector Database

Recently, there has been a growing need for high-dimensional vector data to be managed. This is due to two trends. The first is that there is an enormous growth of unstructured data such as images, videos, texts, medical data, etc., and the second reason has to do mainly with vector embeddings. Vector embeddings mean that items are converted to feature vectors. [WYG⁺21] AI models like large language models generate embeddings. An example has been given in detail in the Literature Review chapter, where the transformer architecture is explained.

The interaction between LLMs and users is not stored in the neural networks of the LLMs themselves. However, with vector databases, it is possible to equip LLMs with long-term memory. Vector databases are designed specifically for embeddings and provide optimized storage and query capabilities. [Incc] Initially, an embedding model is used to generate vector representations for a given content. These vector representations are subsequently stored in a vector database. When a query is made, the same embedding model is used to produce vector representations for it. These query vectors are then used to search the vector database to find similar vector representations. [Incc]



Figure 4.9: A page from the “Physio Vital” brochure from Novartis [Nov]

The Pinecone vector database [Incc] was used via API in the implementation. For testing purposes, a small brochure from the Novartis company [Nov] with instructions for 50 exercises for greater strength and mobility was saved in the vector database. Figure 4.9 shows a page from the brochure. A vector database with 1536 dimensions was first created via the website of the Pinecone company. The number of dimensions is determined by the embedding model with the name Adda 2 [AAA+23] from OpenAI, which was used for vectorized storage and later for similarity search. “Cosine similarity” was specified as a distance metric to measure similarity when the database was created. Then, using methods provided by the framework mentioned above, the brochure with physiotherapy exercises, presented in PDF form, was loaded, the text was split into chunks, and then stored in the Pinecone vector database using the embedding model.

A big advantage of vector databases is that they can easily be kept up to date and language models can simply find information that was not in the training data with which they were trained.

4.7 Developed Tools

As already mentioned, LangChain Agents can use tools, and they use LLMs to decide which ones. Three tools were developed as part of this master’s thesis. The first tool is the vector database tool. It can be used to search for instructions for physiotherapy exercises in the Pinecone vector database, where knowledge about physiotherapy exercises has been stored in vectorized form, as described in the last section. The second developed tool is a search tool. It can be used to search the Internet for general health topics. As with the vector database tool, the idea is that LLMs hallucinate, but answers related to health topics should be as accurate as possible. The third developed tool is the exercise tool. It contains all the exercises described above that the application currently knows such as the dumbbell exercise, the swing exercise, etc.

4.8 Workflow of the Final Implementation

4.8.1 Diagram of the Workflow

Figure 4.10 shows how the previously described components of the final prototype implementation, which was demonstrated in the assessment workshop, interact and how the workflow works at a high level of abstraction.

The user can speak into a microphone, the speech is recognized, and converted to text via Microsoft’s speech-to-text model. This text will then be processed further. Depending on the requests, the LLM decides whether to start the vector database tool, the search tool, or the exercise tool.

When a question is asked about physiotherapy exercise, the LLM decides to use the vector database tool. A query is sent to the vector database. Relevant chunks of information are retrieved, and these are then sent back to the language model. If the user asks a

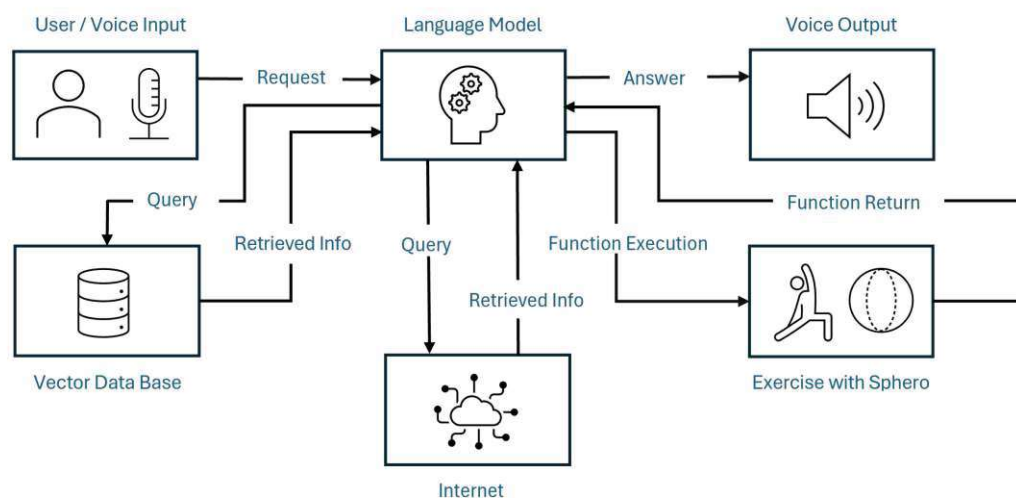


Figure 4.10: Illustration of the workflow of the final implementation

general question about a health topic, the LLM decides to use the search tool to search for information on the Internet. The result of a Google search is then sent back to the language model. Based on information from the vector database or the Internet, the LLM then generates answers for the user. If the user has a question about another topic, the LLM does not use any tool, but answers directly.

If the language model recognizes the user’s intention to do an exercise, it starts the exercise tool. If the function in which the respective exercise is being processed is finished, it returns feedback about the exercises. This information is then sent to the LLM, which is instructed to pass this information on to the user and to praise them for the exercises performed.

The response generated by the LLM is finally converted to speech by Microsoft’s text-to-speech model and is then played by loudspeakers. The user can then send the next request.

4.8.2 Interaction Examples

Figure 4.11 shows a concrete example in which the user says: “I want to train the inner hip muscles.” Then the AgentExecuter Chain starts. The language model used was GPT-4-0613. This is a version of GPT-4 that is fine-tuned to detect when a function should be called and works better than other models tested. It can respond with a JSON object that contains the parameters to call the functions. [Opea] GPT-4 decides correctly to use the vector database tool, as it is a query related to physiotherapy exercises. The vector database is searched, the exercises number 9 and number 10 that target the inner hip muscles and can be seen in Figure 4.9 are found. The exercise instructions are then sent to GPT4, and GPT-4 in turn communicates this to the user. Using a vector database,

4. IMPLEMENTATION

an LLM such as GPT-4 can be equipped with long-term memory, as mentioned. The implementation also includes a method with which the vector database can be updated.

```
RECOGNIZED: Physio ich möchte die innere Hüftmuskulatur trainieren.
Sending recognized Speech to AI...

> Entering new AgentExecutor chain...
```json
{
 "action": "Physio Database Tool",
 "action_input": "innere Hüftmuskulatur trainieren"
}
...
Observation: Um die innere Hüftmuskulatur zu trainieren, können Sie folgende Übungen ausprobieren:

1. Stellen Sie sich breitbeinig hin und verlagern Sie das Gewicht auf ein Bein. Schieben Sie nun das Becken schräg nach unten.

2. Nehmen Sie die angezeigte Position ein. Ziehen Sie die Füße zu sich und drücken Sie die Knie nach unten. Neigen Sie dabei den Oberkörper nach vorne.
Thought:```json
{
 "action": "Final Answer",
 "action_input": "Um die innere Hüftmuskulatur zu trainieren, können Sie folgende Übungen ausprobieren: 1. Stellen Sie sich breitbeinig hin und verlagern Sie das Gewicht auf ein Bein. Schieben Sie nun das Becken schräg nach unten. 2. Nehmen Sie die angezeigte Position ein. Ziehen Sie die Füße zu sich und drücken Sie die Knie nach unten. Neigen Sie dabei den Oberkörper nach vorne."
}
...

> Finished chain.
AI: Um die innere Hüftmuskulatur zu trainieren, können Sie folgende Übungen ausprobieren: 1. Stellen Sie sich breitbeinig hin und verlagern Sie das Gewicht auf ein Bein. Schieben Sie nun das Becken schräg nach unten. 2. Nehmen Sie die angezeigte Position ein. Ziehen Sie die Füße zu sich und drücken Sie die Knie nach unten. Neigen Sie dabei den Oberkörper nach vorne.
```

Figure 4.11: The LLM decides to look for information in the vector database based on the request

Figure 4.12 shows how the user asks: “How many hours of sleep per day are healthy?” The AgentExecutor chain starts again. This time, GPT-4 decides correctly to use the search tool because it is a general question about a health topic. The result of the Google search is sent to GPT-4 and GPT-4 in turn communicates this to the user. Similarly to the vector database example above, the problem of hallucinating language models can be reduced, and the probability of correct answers can be increased.

The user in the example shown in Figure 4.13 says: “I want to do the dumbbell exercise.” The AgentExecutor chain starts again. GPT-4 recognizes the user’s intention to do the dumbbell exercise, correctly decides to use the exercise tool, and passes the “dumbbell exercise” as an argument to the tool. A function that processes the corresponding exercise starts, and the user can then perform the dumbbell exercise with Sphero. After the function has been processed and the user has completed five repetitions in this case, the function returns the number of these repetitions. GPT-4 then communicates this number to the user and praises them. On the one hand, OpenAI fine-tuned the model so that it should be generally polite and helpful, and on the other hand, the model was instructed as part of this work that it should act as a friendly artificial physiotherapist who praises the practitioner after completing exercises. However, it was not specified how users should be praised. This is up to the model itself.



```

RECOGNIZED: Physio wie viele Stunden Schlaf pro Tag sind gesund?
Sending recognized Speech to AI...

> Entering new AgentExecutor chain...
```json
{
  "action": "Search",
  "action_input": "wie viele Stunden Schlaf pro Tag sind gesund"
}
...
Observation: Aus vielen Untersuchungen geht hervor, dass die meisten Menschen sieben bis acht Stunden Schlaf brauchen.
Als Faustregel gilt: Wer tagsüber ...
Thought:```json
{
  "action": "Final Answer",
  "action_input": "Die meisten Menschen benötigen sieben bis acht Stunden Schlaf pro Tag. Dies kann jedoch von Perso
n zu Person variieren. Es ist wichtig, auf die Signale Ihres Körpers zu hören und sicherzustellen, dass Sie sich nach
dem Aufwachen erfrischt und erholt fühlen."
}
...

> Finished chain.
AI: Die meisten Menschen benötigen sieben bis acht Stunden Schlaf pro Tag. Dies kann jedoch von Person zu Person varii
eren. Es ist wichtig, auf die Signale Ihres Körpers zu hören und sicherzustellen, dass Sie sich nach dem Aufwachen erf
rischt und erholt fühlen.
Recognizing...

```

Figure 4.12: The LLM decides to perform a Google Search based on the request

```

RECOGNIZED: Physio ich möchte die Hantelübung starten.
Sending recognized Speech to AI...

> Entering new AgentExecutor chain...
```json
{
 "action": "Exercise Tool",
 "action_input": "Hantelübung"
}
...
Observation: 5
Thought:```json
{
 "action": "Final Answer",
 "action_input": "Super, du hast 5 Wiederholungen der Hantelübung gemacht. Weiter so!"
}
...

> Finished chain.
AI: Super, du hast 5 Wiederholungen der Hantelübung gemacht. Weiter so!
Recognizing...

```

Figure 4.13: The LLM decides to start the dumbbell exercise based on the request

### 4.8.3 Detailed Description of the Agent

In order to explain the most important parts of the implementation, some selected code snippets should be briefly discussed. With the exception of the code belonging to the ExerciseTool class at the bottom left, Figure 4.14 shows code snippets from the Agent class.

The first lines on the left side concern the vector database, which has already been explained in more detail. The vector database is used as the backbone of the retriever object below. A retriever serves as an interface that returns documents based on an unstructured query. [Incb]

## 4. IMPLEMENTATION

```
Vector Database
self.vectordb = vectordb
Retriever
self.retriever = vectordb.getRetriever()

#Google Search:
search = SerpAPIWrapper(serpapi_api_key = serpapi_key)

Initialize LLM
llm = ChatOpenAI(
 openai_api_key = openai_api_key,
 temperature = 0,
 model_name='gpt-4-0613')

Retrieval qa chain, will be converted into a tool
qa = RetrievalQA.from_chain_type(
 llm=llm,
 chain_type="stuff",
 retriever=self.retriever)

Initialize conversational memory
conversational_memory = ConversationBufferWindowMemory(
 memory_key='chat_history',
 k=3,
 return_messages=True)

class ExerciseTool(BaseTool):
 name = "Exercise Tool"
 description = """
 Verwende dieses tool, um die Hantelübung, die
 Schwungübung oder die Kniebeugenübung
 für den Nutzer zu starten. Stelle dazu einen
 der Parameter ['Hantelübung', 'Schwungübung',
 'Kniebeugenübung'] bereit
 """
 def _run(self, exercise: str):
 if exercise == "Hantelübung":
 count = dumbbellExercise()
 return count
 if exercise == "Schwungübung":
 count = swingExercise()
 return count
 if exercise == "Kniebeugenübung":
 count = squatExercise()
 return count

Excercise Tool, Vector Tatabase Tool and Search Tool
self.tools = [
 ExerciseTool(),
 Tool(
 name="Vector Database Tool",
 func=qa.run,
 description="""Verwende dieses Tool, um nach Anleitungen
 für Physiotherapieübungen zu suchen"""),
 Tool(
 name="Search Tool",
 func=search.run,
 description="""Verwende dieses Tool um Informationen
 zu Gesundheitsthemen zu suchen""")])

Initialize Agent
self.agent = initialize_agent(
 agent='chat-conversational-react-description',
 tools=self.tools,
 llm=llm,
 verbose=True,
 max_iterations=3,
 early_stopping_method='generate',
 memory=conversational_memory)

sys_msg = """
 Du bist Physio, ein freundlicher, künstlicher Physiotherapeut.
 Du beantwortest ausschließlich Fragen zu Gesundheitsthemen und zu
 Physiotherapie, andere Fragen beantwortest du nicht.
 Wenn der Nutzer die Hantelübung, die Schwungübung oder die
 Kniebeugenübung starten möchte verwendest du das Exercise Tool.
 Das Exercise Tool liefert die Anzahl der Übungen die der Nutzer
 gemacht hat. Wenn der Nutzer Informationen über Physiotherapie
 Übungen haben möchte verwendest du das Vector Database Tool.
 Wenn der Nutzer Informationen zu sonstigen Gesundheitsthemen haben
 möchte verwendest du das Search Tool.
 """

new_prompt = self.agent.agent.create_prompt(system_message=sys_msg,tools=self.tools)
self.agent.agent.llm_chain.prompt = new_prompt
```

Figure 4.14: Code snippets of the Agent class and the ExcerciseTool class of the implementation

A SerpAPIWrapper object is created underneath. SerpAPI is a service that can retrieve search results from various search engines. [Ser] As mentioned, Google is used in this implementation for search. Then a ChatOpenAI object is created, using the model GPT-4-0613. The temperature is a parameter that can take values between 0 and 1 and that determines the randomness of the LLM's output. Several tests were carried out and it turned out that 0 is the best value. The model should follow instructions and use tools instead of generating creative responses.

The Retrieval QA (Question Answering) chain is used to perform the search in the vector database. When initializing the RetrievalQA object, an LLM object and the Retriever object are passed.

In the next line, a conversational memory object is created. It maintains a record of conversations by which only the last  $k$  interactions between the user and the LLM are saved. This helps prevent the buffer from becoming too large. In the example in Figure 4.14, the last 3 interactions are kept and sent to the LLM so that it can then calculate the next output based on these. This is what makes a “short-term memory” possible. However, the more tokens that are processed, the higher the costs and the longer the processing times.

The last lines of the left column of Figure 4.14 show the code of the `ExerciseTool` class. The first attribute is the name of the tool, and the second attribute is the description of the tool. The description specifies in natural language when the agent should use this tool and which parameters are allowed to pass. When the tool is used, the `run()` method starts. Depending on which exercise the user wants, the LLM passes either “dumbbell exercise”, “swing exercise”, or “squat exercise” as an argument, and the corresponding method that contains the desired exercises starts. After the method has been processed, the number of exercises performed by the user is returned.

In Figure 4.14 on the right column, an array is created with three tools. The first is the exercise tool. The second tool is the vector database tool. For the vector database tool, the Retrieval QA chain mentioned above was converted to a tool. For the third tool, the `SerpAPIWrapper` object is packaged in a tool. Like the exercise tool, the vector database tool and the search tool have a name and a description when the agent should use it.

The agent is initialized below. The first attribute is “chat-conversation-react-description” which indicates the specific type of agent. Chat conversation means that a chat model with conversation memory is used. Description means that the LLM decides which tool it should use on the basis of the tool description. [Incb] React means that the ReAct framework by Shunyu Yao et al. is used. It is a prompt engineering technique to improve the output of LLMs. [YZY<sup>+</sup>22] In addition, the array containing the tools and the LLM object is passed as parameters. The `max-iterations` parameter is used to specify the maximum number of steps that must be executed before the execution loop ends. As the last attribute, the mentioned conversational memory is passed. [Incb]



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Findings and Discussion

## 5.1 Overview of the Findings and Discussion Chapter

The purpose of the Findings and Discussion chapter is to present and discuss the results of the applied methods described in the Methodology Approach chapter. In the section “Application Scenarios and Usage Contexts”, possible application scenarios discussed with the workshop participants are presented. In the section “Participatory Design as Part of this Master’s Thesis” aspects of participatory design, which were realized in the form of workshops are discussed and the question of how grounding between technicians and nursing experts can be achieved is addressed.

- How can a grounding between technicians and nurses be achieved so that the application can represent added value for nurses and ultimately also for older people?

In the last section of this chapter, a comprehensive presentation of the results of the thematic analysis is given. Aspects of the questions about requirements and expectations of the user, possible physical and cognitive restrictions of elderly people, possible technical problems, possible concerns of participants, intuitive, easy-to-use and easy-to-develop robotic applications, whether higher perceived intelligence could be beneficial for motivation, and whether research should focus more on less complex robots are discussed.

- What requirements and expectations must the application meet for the target groups to be assessed as successful?
- What possible limitations could arise with respect to the physical and cognitive restrictions of older people?

- How can we use the latest AI developments for intuitive and easy-to-use robotic applications?
- What do people think about intelligent robots? Are there possible concerns?
- Can higher (perceived) intelligence have a positive effect on motivation?
- What technical problems and limits arise for the developed application?
- Should research focus more on less complex robots?

### 5.2 Application Scenarios and Usage Contexts

As part of the requirements workshops, participants were encouraged to develop application scenarios for the robot ball Sphero and for possible robotic fitness applications in general. The question formulated in the Methodological Approach chapter “What scenarios can you imagine in which a robotic fitness application could be used (in a health care context)?” was asked not just once but several times during the requirements workshops. This allowed participants to gradually develop further scenario ideas based on demonstrations and group discussions.

These scenarios could be combined into the theme “Application Scenarios and Usage Contexts” identified by thematic analysis. Since this theme would not serve to directly answer the subquestions of the research question formulated in the Introduction chapter, it was omitted. However, the identified application scenarios should be listed and described at the beginning of this chapter so that possible ambiguities can be avoided.

A scenario developed by the participants involves the use of a robot like Sphero or a robot like Pepper [Gro] in an office. Sphero was discussed in detail in the Implementation chapter and is used for the implementation as part of this master’s thesis. The robot Pepper, which can be seen in Figure 5.1 and which has not yet been described, is, according to its manufacturer, capable of identifying faces and human emotions. Designed specifically to interact with humans, this robot can communicate through conversations and its touchscreen. [Gro] The middle-aged participants of the first workshop said that robots could, for example, drive (or walk) to employees and encourage them to exercise or adopt a better sitting posture. In addition, they thought that a robot could also encourage group activities. In the following, this scenario is called the “office scenario”.

Another scenario that was discussed by middle-aged participants during the first workshop involves the use of robots in a fitness studio. A robot like, for example, Pepper could serve as an entry point for new guests, or it could walk (or drive) from guest to guest and assist them. The robot could also act as a fitness trainer, giving tips, demonstrating exercises, and evaluating correctly performed exercises. One participant said, for example: “In most gyms, there is no good trainer who is there for you.” This scenario is referred to as the “gym scenario” in the following.

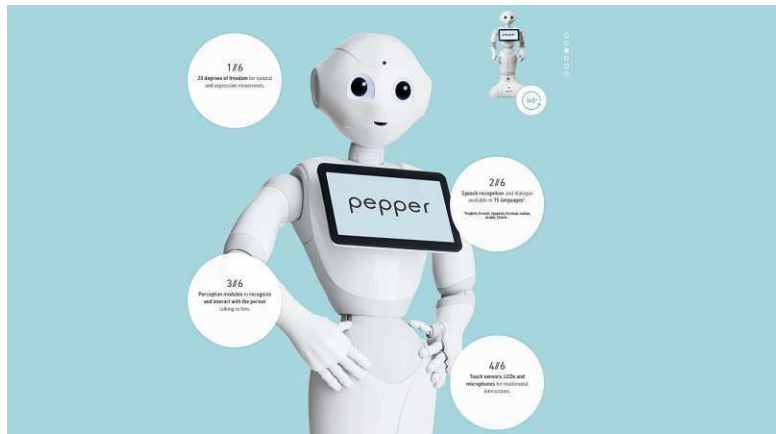


Figure 5.1: The humanoid robot Pepper [Gro]

A scenario that was discussed extensively in both requirements workshops and where nursing experts saw a lot of potential was an application with the robot ball Sphero in retirement homes. Sphero could be integrated into various games, such as catch games, which will be described in more detail below. In both the “office scenario” and the “retirement home scenario”, a robot could promote interactions within groups and support and enrich group activities.

The use of robots, especially for the elderly, as part of Active Assistive Living (AAL) was also discussed. One participant, with decades of experience in retirement homes, believed that there will not be enough retirement home spaces in the future and that as many people as possible will have to stay home as long as possible. She heard lectures on AAL and was convinced that robots can play an important role in the future in relation to safety, entertainment, and exercises, which are intended to help maintain the physical and cognitive abilities of older people.

## 5.3 Participatory Design as Part of the Master’s Thesis

### 5.3.1 Grounding with Workshop Participants

Lee et al. conducted participatory design workshops with a group of older adults with low socioeconomic status. They noticed that only over time and after several workshops the participants had opened and started to think creatively instead of consumer-oriented. [LŠC<sup>+</sup>17]

This was not observed in the participatory design workshops that were held as part of this master’s thesis. The reasons for this are probably that some of the participants knew each other and that there was no power imbalance between the workshop participants and me. If there are significant socioeconomic differences between the workshop organizer and the workshop participants, more workshops may be necessary to build trust over time. Furthermore, it should be emphasized that the workshops did not start “from scratch”.



Participants were able to interact with Sphero and with GPT-4 as described in the Implementation chapter, and short videos of Pepper [Gro], PARO [oAISA], Care-O-Bot 4 [IPA], and ROBEAR [Rik] were demonstrated.

During the first requirements workshop, it was striking that middle-aged participants were more inclined to develop ideas for others than for themselves. As mentioned, during the first requirements workshop, an idea for middle-aged people was that a robotic fitness application in an office could encourage employees to exercise. When asked whether the participants could imagine such an application for themselves, one participant said: “It wouldn’t work in my office at work because it is stressful and there is always a call coming in” and another said: “It wouldn’t work for us either. It is not stressful, but we have a lot to do and lots of meetings.”

Using Sphero as an artifact and the opportunity for participants to test the existing exercises and games with Sphero and to interact verbally with GPT-4 it was relatively easy to achieve a grounding between me as a technician and the nursing experts within the second requirements workshop group. The experts took on the role of a co-designer, and mutual learning took place. For example, a participant who has decades of experience in nursing homes said before the requirements workshop: “To be honest, I cannot imagine anything about it.” After the demonstration and interaction possibilities with Sphero described in the Implementation chapter, she said that in retirement homes games are conducted where balls are thrown to the residents, and they have to catch them, and she asked if the spherical robot ball Sphero could be used for this?

### 5.3.2 Development of a Robotic Fitness Application with Sphero

The other nursing experts then also participated in the discussion. These catch games take place in a sitting circle or around tables. Then they are usually combined with cognitive exercises to maintain mental fitness. When nursing home residents catch a ball, they have to answer simple questions. Examples the nurses gave include: “What was your last name as a single woman? What is the name of your neighbor on the right? Where were you born? What kind of job did you have? Tell me a river with an A.”

The nursing experts provided many other examples of cognitive exercises. For example, retirement residents have to guess to which animal or vehicle sounds being played belong to, or they have to solve mental arithmetic tasks. Other activities that residents often do in a group include singing and praying. This was also confirmed by the two seniors. Although they do not live permanently in the retirement home themselves, they regularly visited a nursing home and participated in the so-called “Young Hearts Club”. An expert said that people in retirement homes love to tell jokes, and Sphero could tell jokes or laugh when a joke is told. All participants were enthusiastic about the idea of telling jokes. Tests were then carried out with GPT-4. Basically, GPT-4 can tell and even invent an endless number of different jokes. However, the quality of the jokes was not particularly convincing from the participants’ point of view.



A nursing expert said that not only could Sphero be caught and then asked a question, but since Sphero is a robot, it could also drive to the participants “as a reward if someone knows something”. So, in a retirement home scenario, Sphero could move from one person to another on a table where practitioners are sitting. Sphero could be controlled by nurses or residents. It would, for example, be conceivable to realize the prototypical implementation of the agentic application as an Android and/or iOS application and to add a way to control Sphero via a graphical interface, similar to what Sphero Play already offers. A theoretical possible automated variant will be briefly discussed later in this chapter.

An important goal of this work is not to develop a finished product, but to find out how the latest AI developments can be used for intuitive, easy-to-use, and easy-to-develop robotic applications. In order to achieve a finished product, more workshops and adjustments to the hardware of Sphero would be necessary. This will be discussed later in this chapter. However, as described in the Implementation chapter, it is possible to recognize when Sphero has been caught. Such catch games with Sphero could then be combined with cognitive games. The prototypical agentic implementation that was developed as part of this master’s thesis and that is described in detail in the Implementation chapter offers potentially many possibilities for all kinds of cognitive exercises. Cognition “includes perception, attention, memory, language, thinking, problem solving, and intelligence”. [HBKE16]

## 5.4 Presentation of the Results of the Thematic Analysis

### 5.4.1 Overview of Identified Themes

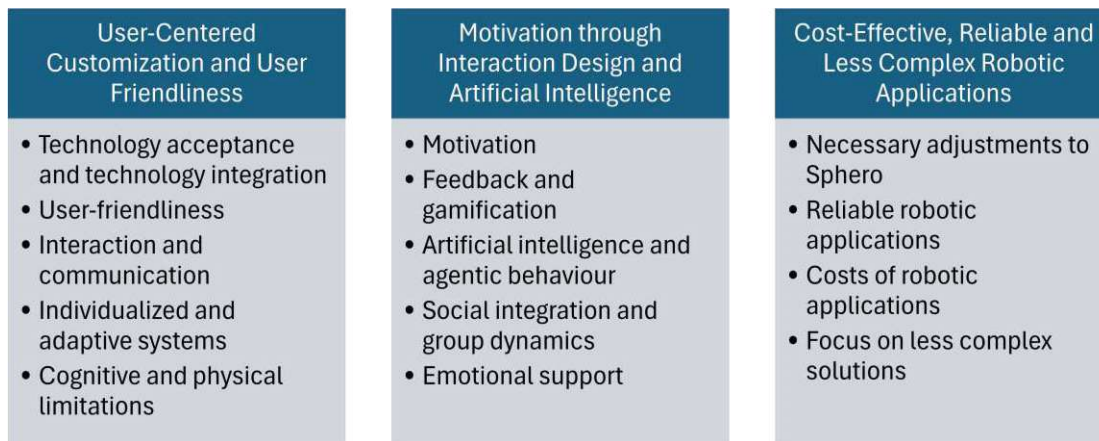


Figure 5.2: Themes identified by thematic analysis

Figure 5.2 shows the themes that were found after applying thematic analysis. The associated codes are located below each. The themes and codes are discussed in the next section. The discussions serve to answer various subquestions of the research question.

### 5.4.2 Theme 1: User-Centered Customization and User Friendliness

In this work, a (robotic) application is considered successful if it is accepted and sustainably integrated by users, so that their lives are enriched or the use of the application in a therapeutic context can be a success.

In both requirements workshops, there were discussions about which people are already using technology regularly for their physical fitness or for physiotherapy exercises. It turned out that this applied to almost all middle-aged participants. For example, one said that she had been using a fitness app with instructions for gymnastics for several years. Another participant said he uses an application in which forces are measured in stretchy fitness bands. The force applied is then displayed (in kilograms) in an application on his smart watch. One participant said he uses a Google fitness app. Two other middle-aged participants said they do not use fitness apps. However, they watch YouTube videos in which physiotherapists demonstrate exercises. Another middle-aged person said that she does not have a fitness app and does not watch YouTube videos, but her physiotherapist prints instructions for physiotherapy exercises on paper.

There was agreement among the workshop participants that technology will only be accepted and integrated into daily life if it is user-friendly. One participant said: “One reason why such an application could not be used (sustainably) would be if it were difficult to use.” Another said: “Yes, it should be easy to use. If something is complicated, you don’t use it.” In this context, it should be mentioned that the two seniors who participated in the second requirements workshop also made valuable contributions, but sometimes had difficulty following the discussions of the younger participants. Neither of them has ever learned how to use computers or smartphones, and they could not understand anything about terms like “fitness app”, for example. It should also be noted that there are often physical and cognitive limitations, especially in the older user group. These will be discussed in more detail below, and it must be taken into account with a view to user-friendliness.

In this work, meaningful interaction and communication with a (robotic) application are considered as central elements in connection with user-friendliness, and thus, for user acceptance, integration into everyday life, and the effectiveness of a possible use in a therapeutic context. [Dav89], [HKEW10], [SW11] In this regard, LLMs such as GPT-4 offer interesting new possibilities, as discussed in the Literature Review chapter and the Implementation chapter.

In the requirements workshops, participants were shown how to communicate with the application verbally. The agentic implementation described in the Implementation chapter was developed with the requirements and expectations of the workshop participants in mind. Then it was presented during the assessment workshop. Participants were shown how the LLM can be instructed to start exercises themselves or search for knowledge in databases. They essentially agreed that this represents a simple and natural way of controlling, especially for older people. The intuitive and human-like interaction was generally positive received by the participants. A middle-aged participant said: “What

is fascinating is that it communicates with you.” He also expressed his opinion on the importance of personal address: “I think there is something personal about a robot addressing you by name.”

In addition to verbal communication, visual communication would certainly also be important in connection with a possible artificial fitness trainer in the gym scenario. Middle-aged participants suggested that exercises could be demonstrated on the screen of a future robotic fitness coach, the correct execution of the exercises could be evaluated, and appropriate verbal feedback could be provided.

The code “Individualized and adaptive systems” identified in the data is closely related to user friendliness, acceptance, and integration into everyday life. For example, scenarios in which a robot approaches people as an active system with the aim of keeping people physically and mentally fit were discussed in both requirements workshops. When asked about the reasons why an active robotic fitness application could not be used sustainably, a participant said: “If the robot goes overboard with suggesting exercises, it could be annoying.” Other participants in the first workshop were also of the opinion that an application could then quickly become annoying. When asked what an application in the office scenario for middle-aged user would have to be able to do so that it would not be annoying, one said that a robot should ask “nice and friendly” whether a user has time for exercises. Another participant said: “A robot could, for example, measure the volume level in the office and if it is quite loud, then it knows that there is no time for exercises.” A third participant said: “The robot could be connected to everyone’s Outlook calendar, and if, for example, you have a meeting, it won’t come.” However, he also said that it generally annoys him when the fitness application on his smart watch asks him to exercise.

The diversity of people was evident in the workshops. There were different preferences for more active and more passive systems. A participant said: “A robot could learn what people’s preferences are.” Another participant said: “In the office scenario, a robot could learn that a person does not want to exercise. For this person, it is just annoying.” A nursing expert said in the context of a retirement home scenario: “Some older people always want to talk. Others are interested in facts such as what the largest city in the USA is.” These quotes emphasize the participant’s opinion that the ability to learn individual preferences is crucial for the sustainable use of robotic systems.

However, current LLMs do not learn continuously. The weights of the underlying neural networks are only updated when these huge models like GPT-4 are re-trained from time to time with gigantic amounts of data and at high costs. Therefore, the vector database tool, which is described in the Implementation chapter, was developed not only with the problem that LLMs hallucinate but also with the participants’ expectations for individualized, adaptive, and learning (robotic) systems in mind. With databases, it is possible to equip transformer-based LLMs such as GPT-4 with long-term memory. Vector databases are optimized for storing unstructured data such as texts and images, but, based on the existing implementation, it would also be relatively easy to create an

SQL database tool with which specific user data can be stored in a structured manner in SQL databases.

User orientation can be considered both from the perspective of individual users and from the broader perspective of whole user groups. Within the different groups, there are tendencies towards certain abilities and restrictions. A question formulated in the Introduction chapter concerns possible problems that could arise on the one hand with respect to possible physical limitations and on the other hand with respect to possible cognitive limitations of elderly persons. The swing exercise, the dumbbell exercise, and the squat exercise, which were described in detail in the Implementation chapter, were already available at the time of the requirements workshops, and older participants could test them. In the above-mentioned catch games, nurses throw balls to residents and they have to catch them. During the workshop, it was tested whether the two seniors could catch balls. These tests were not performed with Sphero to avoid possible damage, but with tennis balls. Subsequently, the problems that could arise in connection with these or similar exercises or games were discussed with the nurses and older participants.

An expert said in connection with catch games: “If the balls are thrown flat, it is usually not a problem. However, anything over the head is difficult because older people’s ligaments are often shortened.” The swing exercise is an exercise performed overhead. Therefore, it is not suitable for many older residents of retirement homes. For the older participants in the workshops, this was particularly true for the woman. What makes her case even worse is that she tore a ligament as a result of a fall and therefore movements above her head with her right arm are completely impossible. She said, for example, that it is very difficult for her to put on a headscarf.

In addition, she is blind in one eye and has very poor vision in the other because she suffers from glaucoma. Catch games are therefore almost impossible for her. Despite his advanced age, the other senior is agile and has surprisingly good motor skills. He was able to catch balls easily. The swing exercise that was tested during the requirements workshop also did not pose any problems to him.

When asked if they still exercise or do physiotherapy, both seniors said no. The elderly man said that the last time he did physical exercises was in the form of rehabilitation after surgery, about 10 years ago. A nursing expert emphasized that for most older people in a retirement home, physiotherapy exercises are often no longer possible: “Not so much with this clientele. Old people could throw or roll the robot ball Sphero to other people. They may also have to do an exercise in which they have to stand up and sit down again.”

However, the nursing experts also said that the squat exercise with Sphero as described in the Implementation chapter is difficult or impossible for most older people. This was also confirmed by the two participants during the workshops. Both had knee operations in the past. Especially the elderly woman has a serious problem with arthritis in her knee. The squat exercise with Sphero is completely impossible for her.

The experts agreed in the second requirements workshop that physical exercises and

physiotherapy should be rather a priority for younger people. A middle-aged participant said in the first requirements workshop: “I always get exercises from my physiotherapist and then I should find the motivation to do these exercises at home, and they should be done correctly, and a (robotic) application could help me.”

The elderly woman emphasized the importance of prevention. “I think it’s important to start exercising young.” The elderly man agreed. In the first requirements workshop, a middle-aged participant said in this context that some exercises with Sphero could be too difficult for older seniors but too easy for younger people. Various necessary adjustments that would have to be made to Sphero to turn it into a product will be discussed in more detail below.

During the workshops, participants were able to communicate with GPT-4 as stated. The problem that arose here was that the elderly man has severe hearing loss. On the one hand, he had difficulty following the discussion, especially when several people were talking at the same time, and on the other hand, he had problems understanding the application. Although he has a hearing aid, he said he never got used to it completely and should have started using it sooner.

Many people in retirement homes have dementia-related illnesses and there are limitations in short-term memory and temporal and spatial orientation, as experts emphasized. However, the nursing experts also said that numerous cognitive exercises such as those described above that could be combined with simple movement exercises such as catch games are still possible for older people in retirement homes. For example, experts said that many older people are still surprisingly good at mental arithmetic.

Another point related to cognitive restrictions is that older people’s attention spans are often limited, as a nurse emphasized. Another expert agreed with her and said: “If the exercises or games that a robot suggests take too long, it will be too much for the elderly in the retirement home.” However, it should be emphasized that although the second requirements workshop took a relatively long time of about 1 hour and 20 minutes, the elderly woman had no problems in this regard and she made valuable contributions from the beginning to the end. In general, it should be noted that although she has many physical limitations, some of them severe, she is mentally very fit. The elderly man, on the other hand, generally had some problems following the workshop discussions. This was particularly due to his severe hearing loss, as mentioned.

### 5.4.3 Theme 2: Motivation through Interaction Design and Artificial Intelligence

The participants in both requirements workshops repeatedly emphasized the importance of motivation and agreed that a robotic fitness application will only be used sustainably if it has a motivating effect. For example, when asked why a robotic fitness application could not be used sustainably, a nursing expert said in the second requirements workshop: “If the elderly are not motivated by the robot ball Sphero (or another robot) and if they do not want to do suggested exercises, there is no added value.” As already mentioned,

the elderly woman finds it difficult to put on a headscarf due to a torn ligament in her shoulder. However, she also said: “I keep trying until I manage to put the headscarf on.” A nursing expert said: “Then it would be good if a robot said: ‘Keep going! There is still a bit to go!’” The elderly woman answered: “It’s hard when you’re old because everything hurts and when everything hurts, you’re not motivated.” The elderly man agreed and reiterated the importance of prevention, for which motivation also plays an important role.

For example, a middle-aged person said in the first requirements workshop: “I practice a lot at home. But I need something to motivate me, so I usually use a fitness app.” As literature research has shown, robots can take on the role of a (motivational) fitness trainer. [SRS<sup>+</sup>14] However, how motivating the robotic fitness application actually is depends on the implementation. To be more motivating than, for example, a physiotherapist in a YouTube video or physiotherapy instructions on a piece of paper, a (robotic) fitness application must provide appropriate feedback to the user.

A middle-aged participant in the first workshop said that he finds it an incentive when his fitness app shows how many kilograms he just trained (with his mentioned fitness band) and how many exercises he did yesterday or last week. However, he also said that his app does not tell him whether he had performed the exercises correctly. That is why he liked the idea of a robotic fitness application that evaluates performed exercises, after the implemented possibility of recognizing dumbbell exercises, swing exercises, and squat exercises was presented during the workshops.

Another middle-aged participant said that he was very skeptical about fitness apps for a long time and did not think he would ever use one of them. Then he tried a fitness app and saw how it slowly incorporated into his life. He said: “I get feedback on how many steps I walk each day and how many kilometers these are. This feedback is very interesting.” He then also highlighted the motivating factor of the graphical preparation of the data: “Here you can see two rings. The first ring shows how much cardio exercise I do. If I walk faster or do any other exercise that increases my heart rate, this ring fills up. The other ring shows the steps that I walk per day. When both rings are filled, I get confirmation that I have achieved my goals. That is a very motivating factor for me.” Another participant then summarized this with the word “gamification”. The prototype application with Sphero in combination with a potential possibility of storing and retrieving information about users via tools can certainly be seen as a basis for many possible considerations towards meaningful feedback and elements of “gamification”.

The workshop participants repeatedly emphasized the importance of positive feedback, such as praise. This was then taken into account when developing the prototype application. GPT-4 was prompt engineered to behave as a polite and helpful physiotherapist who praises users. The model was, for example, asked to praise the workshop participant Markus. It said: “Markus, you are persistent.” When asked why Markus is persistent, GPT-4 answered: “It’s just a positive trait that many of my patients have. There are many ways to praise someone.”



A middle-aged participant in the first requirements workshop suggested that praise could contain interesting information about possible health benefits that certain exercises can bring. He said the model could say, for example: “You did this exercise excellently, but don’t forget to keep practicing. Strength exercise is important against wear and tear.” Providing information like this could have the potential to awaken people’s curiosity and increase (intrinsic) motivation. The vector database tool and the search tool were then developed with this suggestion in mind.

Another goal of this master’s thesis was to find out how the participants of the workshops think about intelligent (robotic) applications. As mentioned, in the requirements workshops the participants had the opportunity to communicate with GPT-4 and then the question was asked whether they have concerns when communicating with artificial intelligence. In the assessment workshop, the participants were able to interact with the prototype application that was developed as part of this work, and then the question was asked whether the participants had any concerns about robotic applications based on artificial intelligence like this.

Two participants in the first requirements workshop are interested in AI debates and news about AI in the media. They said that these developments are the “next logical and unstoppable” steps. The other participants of the first requirements workshop followed the topic of artificial intelligence rather on the sidelines or not at all. All participants except for the two older seniors are middle-aged and at least technically savvy enough to use computers and smartphones. A (middle-aged) participant in the first requirements workshop said that intelligent behavior could be frightening to older people. Another participant agreed: “Old people could be overwhelmed if a robot seems too intelligent.”

During the second requirements workshop, the two older participants and nurses who have decades of experience in retirement homes were asked if that was actually the case. One of the experts described older people in retirement homes with whom she worked as “perhaps sometimes initially skeptical about new things, but generally curious and open-minded.” She asked the elderly woman if she had any concerns or if she found it strange to talk to a machine. The elderly woman responded jokingly: “I even talk to myself sometimes.”

None of the participants in both the requirements workshops and the assessment workshop expressed any concerns about intelligent (robotic) implementations for themselves. Instead, wishes were expressed that a robot should appear intelligent and say things in different ways, it should be able to evaluate exercises precisely, it should drive autonomously, etc.

The two older participants in the workshops have never learned to use computers or smartphones. However, the possibility of verbally instructing an agent to start exercises (and search in databases) offered by the application that was developed as part of this work represents a very simple way of operating. People are likely to get used to easy-to-use, accessible, and agentic (robotic) applications very quickly. The implementation was tested in the assessment workshop and the participants shared the assessment that such

a form of control has potential, as mentioned. However, they also expressed criticism, which will be discussed later in this chapter.

The nurses said that few older people in retirement homes do physical exercises or physiotherapy, as mentioned before. However, an expert told the story that a group of older women in the nursing home where she works always wanted to watch a television program with a physiotherapist, and then they tried to follow the instructions as best they could. She also emphasized how these women motivated each other to get together at the time the program was broadcast and how motivating the physiotherapist was for them and how motivating it was for them to do these exercises in a group.

In most cases, the initiative to play or exercise does not come from the residents of retirement homes, but from the nursing staff. As mentioned, these exercises are often simple movement exercises, such as ball catch games, which can be combined with cognitive games. The experts were fundamentally convinced of the idea of using Sphero for such games. One said: “That is definitely great in a group. I can place Sphero in the middle of the group and start a program. I think the old people are then much more attentive and motivated.” Another nursing expert agreed.

A middle-aged participant in the first requirements workshop said for an office scenario: “Sphero (or another robot) could drive from one person to another and ask them if they want to exercise. If a person has agreed, it could motivate others to join in, too.” Another middle-aged participant also emphasized that he could imagine that group dynamics could be encouraged by a robot. The code “social integration and group dynamics” identified in the data, therefore, refers to a robotic application that can be integrated into groups to promote, support, and enrich group activities to increase the motivation of participants.

In the second requirements workshop, an expert discussed the loneliness of many older people, for example, in an AAL scenario, and emphasized emotional support in relation to training motivation: “I think of people who are very lonely. It would be good if they had a device (such as Sphero) to motivate them to do something physically.” The elderly woman said that her husband is very important for her motivation: “When I am resting on a couch and still have work to do, I tell my husband to count to 10 and then I get up.” She also (understandably) said that she would rather be motivated by her husband than by a robot. However, an expert then said: “All women no longer have a husband.”

A robot does not necessarily have to be intelligent to be able to provide emotional support. An expert has years of experience with the robot seal PARO that is described and discussed in the Introduction chapter and the Literature Review chapter. She said in her opinion, PARO has positive effects on people with dementia. Another expert said: “We have a therapy doll in the retirement home. This doll does not move, but people with dementia talk to it and build a relationship with it, and that is good for them.” In addition, she said that this therapy doll is significantly cheaper than PARO.

During the workshops, there was also a discussion about possible ethical concerns related to emotional support through robots. The general tenor among the participants was



that anything that helps people feel better should be seen as positive. For example, the experts were of the opinion that it is better if lonely people can talk to machines than to no one. A nursing expert said that the robot seal PARO should not be seen as a replacement for humans but rather as a replacement for pets because animal behavior is imitated. As described in the Introduction chapter, this was also the intention of the developers.

A middle-aged participant in the first workshop could imagine a robot fulfilling the role of a pet also in an office scenario. He said whenever a dog comes from the neighboring office he is distracted from work but, nevertheless, he is happy about the dog's visit, and he would not deny the dog his attention because the dog spreads "positive mood." A robot could imitate the behavior of a dog to motivate, he said.

In the context of the aforementioned discussions about what a robot, as a system that actively approaches people, would have to do to not be annoying, a middle-aged participant said: "A robot would not be annoying if it were funny and suggested funny exercises." He suggested that Sphero could drive to a person and ask her or him to throw it into a basketball hoop.

### 5.4.4 Theme 3: Cost-Effective, Reliable and Less Complex Robotic Applications

In the assessment workshop, the participants confirmed the usefulness of Sphero Bolt as an artifact that they could touch and test and that helped them imagine how a commercial robotic application with it could look like. However, in the second requirements workshop one of the first questions of a nursing expert in connection with possible catch games was: "What can the Sphero robot actually withstand?"

Although Sphero seems quite stable, it was certainly not designed to be thrown. Catch detection tests, described in the Implementation chapter, were performed over a mattress to avoid the risk of damaging Sphero due to a fall. During the second requirements workshop, it was tested whether the older participants could catch balls. Sphero was not used for this, but rather tennis balls, as mentioned. A nursing expert said that they use rubber balls in the retirement home for these catch games. Sphero's cover could therefore also be made of rubber. The surface of Sphero is very smooth. "It should be rougher so that Sphero can be caught better", another nurse said. Of course, the electronics would also have to be robust enough so that it does not break due to vibrations or a fall.

At the beginning of both requirements workshops, there was some confusion among the participants that the audio output in the current form of the implementation comes from the speaker of a notebook. Sphero could be equipped with a loudspeaker through which voice output could take place and over which sounds and music could be played for various conceivable cognitive exercises. A microphone is also not available in the current form of Sphero and could also be installed. In the case of an Android and/or iOS implementation like Sphero Play, the microphone and speaker of a smartphone or tablet could theoretically be used. However, if, for example, in group games in retirement

homes, there were a person sitting at the other end of a table where the person controlling Sphero with a smartphone app such as Sphero Play is sitting, the speech recognition would work poorly and the person at the other end of the table might also not understand the voice output well either.

A problem during the second requirements workshop was that the elderly man suffers from (severe) hearing loss, as mentioned. The nursing experts also confirmed that this is often a problem in retirement homes. For a commercial application, at least volume settings must be taken into account. GPT-4 is “smart” enough to repeat sentences or even to formulate sentences more simply on request, but as a language model it is of course not able to say sentences louder when asked to do so. A solution for a commercial product could be the development of a volume tool. If a model like GPT-4 were asked to repeat a sentence louder, it could have access to the audio settings of Sphero (or another robot), and then it could increase the volume in order to comply with the request accordingly.

A middle-aged participant said: “Some of the exercises with Sphero could be too difficult for older people. However, the younger the people are, the more the question arises as to whether the exercises are too easy.” This is especially true for the dumbbell exercise and for the swing exercise. To be a fitness device for younger people, Sphero’s weight would have to be increased as a middle-aged participant suggested. Another middle-aged participant emphasized the possibility of using people’s own body weight. This is the case with the developed squat exercise.

As mentioned, a nursing expert said that in a retirement home scenario, Sphero could drive to residents “as a reward” for correctly answering questions. As described in the Implementation chapter, Sphero can be controlled using the Sphero Play application. This was also successfully tested during the workshops. Controlling Sphero could be done by a nurse or by a resident of a retirement home. The participants asked if an automatic version would be possible. A simple possibility would be if Sphero followed a fixed route on a table. This could already be achieved now. However, more complex variants would also be conceivable. For example, Sphero could be integrated into a larger system with cameras. It would be advantageous if Sphero could be connected to other subsystems via WiFi. The current version of Sphero Bolt does not use WiFi. In the implementation created as part of this master’s thesis, Sphero is connected to a Python program via Bluetooth.

It must be emphasized again that, even after the assessment workshop, the current version of the application is not a finished product, but a prototype. The participants said that the recognition of movement patterns would have to work even more reliably for a commercial product. One consideration here would be to install other or better sensors. For reliable recognition of more complex exercises, it would be necessary to switch to the visual domain. With the help of cameras, pose landmarks could be captured over time, and then a trained neural network could classify the exercises performed by practitioners. Furthermore, more precise feedback could then be provided. Such functionality could be implemented, for example, with 3D pose estimation with MediaPipe. [Goo]

There were also discussions with participants about the reliability of GPT-4. Although this model is the current state-of-the-art, it has clear limits that become apparent with prolonged use. An important point that is discussed in the Implementation chapter is that verbal communication with GPT-4 is achieved through the interaction of three different models, and the speech-to-text model and the text-to-speech model are much less powerful and therefore represent bottlenecks.

During the assessment workshop, the participants also felt that the prototype application was too slow for a commercial product. This was for several reasons. The duration of the processing depends on the length of the chat history that is sent to the model as input. The longer the chat history, the more tokens have to be processed. However, as described in detail in the Implementation chapter, the number of conversations stored can be set. To increase the quality of the model's output, the ReAct prompt engineering technique [YZY<sup>+</sup>22] was used, which also leads to a higher number of tokens to be processed. For simple tasks, such as starting an exercise, this technique is actually not necessary and can be omitted. Another reason why the application was slow is that the version of GPT-4 used is slow. Several months after the assessment workshop, OpenAI released a new and faster version of GPT-4 called GPT-4o. [Opea] In addition, the response time could be significantly improved via streaming. To do this, the text generated by the LLM would have to be converted into an audio stream during creation.

An expert with several years of experience with the robot seal PARO raised the issue of the high price of about "5.000 euros" of PARO. Another nurse said that very simple and cheap dementia dolls can also be used. A Google search yields even much higher prices for the humanoid robot Pepper presented above. Of course, for the future it is conceivable that robots could be much cheaper due to more efficient production and higher sales volumes. The big advantage of the robot Sphero is that it can already be purchased for just over 200 euros. Necessary adjustments to Sphero would certainly increase the price, but they would not fundamentally change the order of magnitude of the price.

Sphero Bolt would only become significantly more expensive if it had a GPU that could run an LLM. However, an LLM with quality comparable to GPT-4 that could run locally on a small robot ball is, at least from the current point of view, unlikely to be possible in the near future. The calculation of the responses of the current application takes place in a data center, which causes ongoing costs. Some participants said they would rather pay more for a product and then have no ongoing costs.

In addition to GPT-4 and GPT-3.5, smaller (open-source) LLMs that can run on a local computer were also tested as part of this work. These do not result in ongoing costs, but the quality of the output was much poorer, so the idea was quickly abandoned. However, it should be mentioned that open source LLMs are currently developing rapidly, as continuously documented in Hugging Face's Open LLM Leaderboard. [Lea] This trend will probably continue, and it is therefore likely that in the near future high-quality open-source LLMs (and multimodal models) can run either locally on relatively expensive computers with no ongoing costs or as a service in a data center with (very) low ongoing costs.

The nursing experts were shown a video of the lifting robot ROBEAR [Rik] that was presented in the Introduction chapter. They were told that it was heavy and immobile and not that it was purely a research object and that its further development was stopped. An expert said: “If ROBEAR were needed in another room, it would then have to be fetched constantly, and that would be very time consuming for nurses.” In addition, she said: “It is often the case that several nurses are in the same room at the same time. They would then have to wait until ROBEAR is available again, and that would also be time consuming.” Another expert agreed and said: “Nurses don’t have that much time these days.”

The third expert said that when patients are “turned over” in a bed, nursing activities are often performed at the same time. Furthermore, she said: “A robot like ROBEAR would have to be able to do more than just relocation work, it would have to be more universal.” However, from a technological perspective, it is already very difficult to design a relocation robot that is agile, strong enough, and (nearly) 100% safe. In general, it can be said that nurses did not believe that ROBEAR could have any added value for them. The experts also made economic considerations. They asked themselves how much ROBEAR could cost and how many ROBEARs would be needed and whether that could even be affordable for a nursing home or a hospital.

However, the workshop participants were rather confident that a product based on a small, inexpensive, and less complex robot such as Sphero Bolt could be promising. As mentioned, especially the experts were particularly convinced of the jointly developed idea of an application with Sphero that can be used by nurses together with older people as part of group games in retirement homes. Another target group, which is not in the focus of this work, but should be mentioned for the sake of completeness in connection with Sphero in healthcare facilities, are children. “I think Sphero could also be used in a children’s ward in a hospital. This would be a nice activity for them and could also encourage them to exercise”, an expert said.

With regard to comparatively less complex applications that could bring added value, it should also be mentioned that nursing experts repeatedly complained about the high documentation effort they face. They hoped for technical solutions that could help them. For example, a nurse said that a robot could be a documentation aid. Another nurse described a situation where a patient feels nauseous. It is often very stressful for caregivers and they often have to rush to the next patient immediately after medication administration. However, if the patient who feels nauseous, for example, has an allergic reaction before the administration of the medication was documented, the caregiver will have trouble. A third nursing expert gave the example that she often comes to a patient, performs nursing work on her or him, and then forgets whether the patient was lying on her or his left or right side. She said: “It would be good if I had a device that would remember this and that I could ask how the patient was lying in the bed later.”

The solution to these problems might not be a robot. For example, a mobile documentation application could potentially help nurses avoid having to go to the control center every time to document and make some of their work processes more efficient. Although this

example is not related to the development of a robotic fitness application, it should be mentioned because it shows how important it is to ask target groups what they really need and it shows that less complex and cheaper solutions (at least compared to applications such as ROBEAR or Care-O-Bot 4) could help more quickly. A mobile documentation application could potentially also be implemented in the form of an agentic application similar to that described in the Implementation chapter. However, important questions about reliability would certainly need to be clarified.



# CHAPTER 6

## Conclusion

At the beginning of this master's thesis, observable demographic trends in large parts of the world as well as some of the resulting, already existing, and certainly increasing problems for the health care sector were discussed. Assistive (social) robotic applications could help solve or mitigate some of these problems. The robot ROBEAR can lift patients, for example, from a bed to a wheelchair. [Rik] Another robot that was presented in the Introduction chapter is Care-O-Bot 4. [IPA] It can bring food and serve as a platform for entertainment, etc. However, most of these existing assistive (social) robots for care are pure research platforms and not (yet) integrated into people's everyday lives. A robot that has been commercially available is PARO. [oAISA] It was developed for people with dementia [WSMK08], looks like a seal, reacts to humans, and is one of the most successful robots used in care so far. What is striking is that PARO is comparatively less complex than other robots.

A central question of this work was therefore whether part of the research should focus more on less complex robots. In order to approach this question, an objective of this master's thesis was to create a prototype of an assistive social robotic fitness application based on an already available, simple, cheap and commercially successful robot.

A target group that is the focus of this work is nursing staff working in retirement homes, as they are becoming increasingly important in the context of the mentioned demographic development. The group of older people arose automatically as a second target group that should be considered because nurses in retirement homes work with them, and together with a robotic application, a triadic interaction scheme often results in practice. [HBG<sup>+</sup>20] In addition, middle-aged people were also included as another relevant target group, with a focus on promoting health and prevention.

Based on these initial findings formulated in the Introduction chapter, the research question that was to be answered in this master's thesis was derived.

## 6. CONCLUSION

---

- How can a fitness application be developed based on a commercially available robot that is used sustainably by nurses and middle-aged users?

In order to answer this question, subquestions were formulated, which will be answered and discussed in the following.

With robots in social contexts, not only functional aspects are important, but also subjective evaluations by users. [LŠC<sup>+</sup>17] To increase the chance of sustainable use of an application, it is therefore certainly important to ask users. However, simply asking people what they could imagine using robotics tends to lead to rather arbitrary answers. The first important subquestion was how a grounding could be achieved between nurses and technicians. Caregivers and technicians are experts in different fields and have different knowledge and different expectations. Grounding between these groups is important so that a developed robotic application will ultimately bring added value to nurses, but also to older people.

- How can a grounding between technicians and nurses be achieved so that the application can represent added value for nurses and ultimately also for older people?

Previous research suggests participatory design (PD) as the basis for effective grounding. In PD, users play an active role in the design process, and participants and technicians should learn from each other. [LŠC<sup>+</sup>17] PD was carried out as part of two requirements workshops. The middle-aged people were invited to the first requirements workshop and caregivers and older people were invited to the second requirements workshop. The results of these workshops were taken into account for an implementation, which is described in the Implementation chapter.

Merely bringing people together does not automatically lead to a co-design. [LŠC<sup>+</sup>17] A systematic approach is necessary. The requirements workshops were therefore structured in three steps. The first step was to collect background information about the participants. For this purpose, a list of questions was created for the target groups. To achieve a grounding more easily between me as a technician and the nursing experts, the idea was to use the robot ball Sphero Bolt [Sph] as an artifact. There also already existed an official Android smartphone application Sphero Play with which Sphero can be controlled in various ways and that also offers some games with Sphero. In a project work, a Python application was developed that recognizes correctly performed exercises by practitioners with Sphero in their hand using a trained machine learning classifier. The exercises include a dumbbell exercise, a swing exercise, and a squat exercise. Furthermore, a second Python application was created that provides the possibility to communicate verbally with the large language model (LLM) GPT-4. [AAA<sup>+</sup>23] In the second step, Sphero, Sphero Play, and these Python applications were presented to the participants, and they had ample opportunities to interact. Furthermore, short videos of the robots Pepper [Gro], PARO [oAISA], Care-O-Bot 4 [IPA], and ROBEAR [Rik] were demonstrated. In the



---

third and final step, discussions about personal expectations, requirements, limitations, design suggestions, etc. took place. A list of questions was also prepared for this step. The exact description of the procedure can be found in the Methodology Approach chapter.

All workshops were recorded and transcribed. A thematic analysis according to Braun and Clarke was carried out to tell the story of the data and to answer other subquestions formulated in the Introduction chapter. Thematic analysis “is a method to identify, analyze and report patterns (themes) within the data”. [BC06]

Before the requirements workshops, there was a lot of uncertainty. However, the introduction of Sphero, the interaction opportunities and the systematic structure of the workshops made it relatively easy for the participants to imagine what a robotic fitness application with Sphero could look like. They took on the role of a co-designer and mutual learning took place. The nursing experts talked, for example, about catch games with balls that they play in groups with residents in the retirement homes where they work. They said that the spherical robot Sphero could be used instead of ordinary balls. After the second requirements workshop, (promising) tests were carried out with a machine learning classifier that detects when Sphero has been caught. Since Sphero is a robot, it could of course also be controlled by a nurse on a table using the existing Sphero Play application. The experts said these catch games (or other simple movement exercises within groups) are usually combined with cognitive exercises. These could be simple questions like “Tell me a river with A”, or games where sounds have to be recognized, etc.

Based on these ideas and further expectations and requirements of the participants that will be discussed in the following and with the question of how the latest AI research developments can be used for intuitive, easy-to-use, and easy-to-develop robotic applications in mind, a prototype application was developed that forms the basis for many conceivable physical and cognitive exercises.

- How can we use the latest AI developments for intuitive, easy-to-use, and easy-to-develop robotic applications?

In the Literature Review chapter, current advances in the field of artificial intelligence were discussed. One conclusion was that autonomously acting robots that understand not only text, images, or a combination of a few modalities but the world itself in a similar way to humans are probably far away. Nevertheless, LLMs and various multimodal models that already exist and will be coming in the next few years appear to be useful as basis for assistive social robotic applications. One problem is that LLMs hallucinate. A way to increase the reliability of these models is to combine them with tools, and this was done as part of the implementation.

With vector databases, it is possible to equip LLMs with long-term memory. A vector database tool was developed and for testing purposes, a small physiotherapy brochure

[Nov] was stored in a Pinecone vector database [Incc]. The vector database tool can be used to search for instructions for physiotherapy exercises in this database. The second tool that was developed is a search tool that can be used to search the Internet for general health topics. With both tools, the idea was that LLMs tend to hallucinate, but answers related to health topics should be as accurate as possible. The third developed tool is an exercise tool that contains all the exercises that the application knows, such as the dumbbell exercise, swing exercise, and squat exercise.

The implementation works in such a way that the user can make a request via microphone. Then a speech-to-text model converts speech to text. The LLM used is a version of GPT-4 that is fine-tuned to detect when a function should be called. Depending on the user request, the model decides whether to start the vector database tool, the search tool, or the exercise tool. The vector database tool delivers the information from the database and the search tool delivers the information from the Internet. If the user asks the model to perform an exercise that it knows, it starts it via the exercise tool. The user can then perform the exercise. After completion, the tool can, for example, return the number of repetitions performed. GPT-4 generates answers in the form of text, and this text is then converted via text to speech model and played through speakers.

This agentic application was developed with the requirements and expectations of the requirements workshop participants in mind, which will be discussed in the following. It was presented to a mixed group during an assessment workshop. The participants found this agentic application promising. For example, the two older participants never learned to use a smartphone or computer. The possibility of starting an exercise verbally via natural language is a simple way of operating an application.

This prototype application could easily be developed further. For example, more physical exercises and many possible cognitive exercises could be developed, the application could be improved with additional tools, and the application could also serve as a basis for many other user-friendly and individualized (robotic) applications. In the next few years, a wide variety of intelligent personal agents for nurses, doctors, and physiotherapists are conceivable in the context of hospitals, retirement homes, or active assisted living. These can be robotic applications, but also mobile or mixed reality applications. However, questions about reliability and safety play an important role. The next few years will show for which areas of application such agents could (already) be good enough and whether they can, for example, meet the high requirements of medical device laws.

- What requirements and expectations must the application meet for the target groups to be assessed as successful?

There was agreement among the workshop participants that new technology will only be accepted and integrated into everyday life if it is user-friendly. Meaningful interaction and communication (with a robot) are central elements in connection with user-friendliness and thus for user acceptance and integration into everyday life. Large language models

---

such as GPT-4 (and other existing and upcoming multimodal models) offer interesting new opportunities for companion robots.

The participants highlighted the importance of individualized and adaptive robotic systems and they said that robotic fitness applications could learn people’s individual preferences. Current large language models do not learn in such a way that the weights of the underlying neural network are constantly updated. In fact, LLMs do not even remember the answers they generate themselves. The entire chat history must always be added as input. One way to equip LLMs with long-term memory is to use vector databases. As mentioned, this was done as part of the implementation. Vector databases are optimized for storing unstructured data. However, based on the existing implementation, it would also be relatively easy to create, for example, an SQL database tool with which specific user data can be stored in a structured manner.

The participants also emphasized several times that a robotic fitness application will only be used sustainably if it has a motivating effect. In psychological research, motivation is seen as a driving force of human behavior. [SRS<sup>+</sup>14] It can be divided into extrinsic and intrinsic motivation. [Ede03] Motivation can be increased by working with human trainers [GMHS20] and robots can take on the role of a motivational fitness trainer. [SRS<sup>+</sup>14] To be motivating, a (robotic) fitness application must provide appropriate feedback to the user. In this context, middle-aged participants emphasized the importance of “gamification”. The agentic prototype application in combination with the aforementioned (potential) possibility of storing and retrieving information about users via a possible SQL database tool could certainly be seen as a basis for many considerations towards meaningful feedback and elements of “gamification” like milestones, for example.

Participants in both requirements workshops also highlighted the possible role of a robot in the promotion, support, and enrichment of group activities to increase motivation. The application with Sphero, designed in collaboration with nursing experts, is intended as a basis for group games for caregivers and the elderly in retirement homes. The experts were of the opinion that such an application could have a motivating effect.

According to data collected during workshops, addressing and supporting people emotionally is another important key to motivate users to exercise more and to ensure that robotic fitness applications are used sustainably. In the first requirements workshop with middle-aged people, the focus was on fun and entertainment. In the second requirements workshop, it was highlighted that older people are often lonely. The focus here was on reducing loneliness and social isolation in order to motivate people.

Due to the importance of the motivation factor and the current interesting developments in the field of AI research, the question arose of whether higher perceived intelligence of models such as GPT-4 could be useful for motivation.

- Can higher (perceived) intelligence have a positive effect on motivation?

Positive feedback has the potential to strengthen motivation. [FM12] The workshop participants also repeatedly emphasized the importance of positive feedback such as praise as an important (extrinsic) motivation factor. Fasola et al., argue repetitiveness and predictability diminish the perceived intelligence of a robot, and this leads to a loss of trust in the usefulness of the robot to motivate. Therefore, a main focus of their work was to eliminate repetitive verbal instructions and comments and they implemented, for example, more than 10 different ways by which a user can be praised by their robot (e.g., “Awesome”, “Nice job”, “Fantastic!”). [FM12]

This was not done in this way in the implementation that was developed as part of this master’s thesis. Instructing GPT-4 to choose from a list of 10 different ways in which it can praise the user would restrict the model instead of making it appear smarter. GPT-4 was instructed, as described in detail in the Implementation chapter, to behave like a polite and helpful artificial physiotherapist who should praise users. However, it was not specified how users should be praised. GPT-4 can praise users in many ways and contextually.

In addition, a middle-aged participant had the idea to strengthen (intrinsic) motivation by arousing curiosity. He said that the praise given by a fitness robot could contain interesting information about health benefits that certain exercises can bring. As mentioned, the vector database tool and the search tool were developed to increase the probability that the information provided by the application is correct.

A hypothesis of this work is that the higher perceived intelligence of LLMs and multimodal models, which can be expected to further improve over the next few years and the possibility of access to knowledge databases via tools, enable agentic (robotic) applications that have the potential to increase the extrinsic and intrinsic motivation of users. A suggestion for future work is to examine this hypothesis in more detail.

A further goal of this master’s thesis was to find out how the workshop participants think about intelligent (robotic) applications and whether they have concerns when communicating with AI.

- What do people think about intelligent robots? Are there possible concerns?

In a study by Lee et al., elderly workshop participants said that robots should not be too intelligent and not too autonomous. [LŠC<sup>+</sup>17] These statements were made in the context of a desire to maintain control of a robot. In a survey conducted in Germany in 2021, respondents also expressed concerns related to possible loss of control due to artificial intelligence. [TV21]

After participants in the requirements workshops held as part of this work had the opportunity to communicate verbally with GPT-4, they were asked if they have any concerns when communicating with artificial intelligence. After the participants of the assessment workshop had the possibility to interact with the prototype application,

---

they were asked if they have any concerns about robotic applications based on artificial intelligence such as the presented one.

No participant expressed any concerns about communicating with intelligent machines, such as robots, for themselves. Instead, wishes were expressed that a robot should say things in different ways, it should be able to evaluate exercises, etc. Unlike in the study by Lee et al., the participants of the first requirements workshop are middle-aged. Furthermore, they are at least technically savvy enough to be able to use computers and smartphones, and no one is afraid of new technology. However, the two older participants in the second requirements workshop also expressed no worries. The experts with decades of experience in nursing homes described older people in retirement homes as “perhaps sometimes initially skeptical about new things, but generally curious and open-minded.”

The opportunities for interaction during the workshops and the way the question in the assessment workshop was asked probably had an influence on the participants’ responses. Some concerns expressed by participants in previous studies and surveys could be interpreted in part as skepticism about unknown new situations and in part as fear of hypothetical scenarios that lie far in the future. However, this does not mean, of course, that there are not many legitimate worries related to artificial intelligence, such as worries about possible manipulation, surveillance, etc.

In the above-mentioned survey in Germany, concerns about possible dehumanization if artificial intelligence replaces human contact have also been raised. [TV21] During the second requirements workshop, as mentioned above in connection with motivation, loneliness of many older people was a topic. The consensus among the experts here was that it is better if lonely people can talk to machines than to no one.

As already indicated, in the context of this work, artificial intelligence primarily means the intelligence perceived by users and not (necessarily) real intelligence. Current LLMs (and multimodal models) have clear limits, which become increasingly apparent to users the longer they interact with them. Technical limits in general are subject of a subquestion which was formulated in the Introduction chapter.

- What technical problems and limits arise for the developed application?

Sphero Bolt in its current form is not designed to be thrown or caught. According to the participants, Spheros’s cover should be replaced with a rubber cover, the smooth surface should be replaced with a rougher one, and the electronics should be strong enough to withstand falls. The current version of Sphero also does not have speakers or microphones. For interactive use in the context of fitness and cognitive exercises, these components could be integrated to enable direct speech input and direct speech output or music output. The volume level and clarity of the audio outputs are critical factors, especially when the target audience includes older people who may have limited hearing. As a language model, GPT-4 can repeat sentences on demand, but it cannot repeat them louder. One possibility would be to develop a volume tool that allows the model to

control the volume. For exercises such as the dumbbell exercise or the swing exercise, the weight of Sphero would have to be increased, especially if it is intended as a fitness device for middle-aged people.

A problem is that LLMs are prone to hallucinations and they are not factual, as mentioned. For example, merely 21% of the answers of GPT-3.5 and 41% of the answers of GPT-4 were consistent with known answers to medical questions, a study by Dash et al. revealed. [DTB<sup>+</sup>23] Using tools, the model can search for medical and physiotherapy topics in knowledge databases. The aim was to increase the accuracy of the answers as much as possible, as mentioned.

During the assessment workshop, the participants felt that the prototype application was too slow for a commercial product. One reason why the application was slow was that the ReAct prompt engineering technique [YZY<sup>+</sup>22] was used to improve the quality of the model's output. However, this technique is not necessary for simple tasks, such as starting exercises or searching databases. Another reason why the application was slow is that the version of GPT-4 used during the assessment workshops is rather slow. Several months after the assessment workshop, OpenAI released a faster version of GPT-4 called GPT-4o. [Opea] In addition, the response time of the prototype application could be (significantly) reduced by streaming. An audio stream could already be generated, while the LLM generates its output.

If an application is considered a medical device, it should be noted that very high requirements must be met. However, even if no hard criteria must be met under medical device laws, reliability still plays a major role when it comes to the fact that people would use such a product regularly. The application should have a motivating effect and should not lead to frustrated users. The participants noted that for a commercial product, the accuracy in recognizing movement patterns needs further improvement. One possibility of improvement could be to use other and better sensors. Recognition of more complex movement patterns would have to be done within the visual domain.

A further problem with the prototypical agentic application, which was developed as part of this work, is that three different models have to work together for verbal communication. The text-to-speech and speech-to-text models used are less powerful than GPT-4 and represent bottlenecks. However, multimodal models that also understand audio could lead to improvements in the next years.

In addition to the question of possible technical problems and limits, the question of limits that could arise due to the physical and cognitive restrictions of older people was also formulated in the Introduction chapter.

- What possible limitations could arise with respect to the physical and cognitive restrictions of older people?

As mentioned, the participants had the opportunity to test the dumbbell exercise, the squat exercise, the swing exercise, catch games with tennis balls, and to interact with GPT-



---

4 verbally during the second step of the requirements workshop. Afterward, discussions were held about possible physical and cognitive problems that could arise with these exercises, exercises that are similar to those and other exercises that nursing professionals perform with the elderly in retirement homes.

Only a small number of older people in retirement homes engage in physical exercise or physiotherapy. Nurses perform rather simple movement exercises with residents. These can then be combined with cognitive exercises, as mentioned. The elderly often have problems with shortened ligaments. This means that exercises performed above the head are difficult for many older people. Among the exercises developed, this applies especially to the swing exercise, which is more suitable for middle-aged people. Throwing balls (or other objects) flat is usually not a problem. However, this was not the case for the elderly woman. As she is blind in one eye and has very poor vision in the other because she suffers from glaucoma, it was not possible for her to participate in the catch-game test with tennis balls. The elderly man could easily catch balls. However, he suffers from severe hearing loss. That is why he had difficulty communicating with GPT-4 verbally. The problems many elderly share are knee problems. The developed squat exercise is therefore more suitable for the middle-aged group and less suitable for people in retirement homes. The developed dumbbell exercise is possible without problems for most older people.

Many residents of retirement homes have dementia, which limits short-term memory and temporal and spatial orientation, as experts emphasized. However, they also mentioned a variety of cognitive exercises with the aim of promoting and maintaining mental fitness that they perform with the elderly in retirement homes. For example, an expert said that many residents of nursing homes are surprisingly good at mental arithmetic. GPT-4, on the other hand, is very poor at calculations. The results of additions of large numbers, for example, that GPT-4 delivers, often seem plausible at first, but are basically always wrong when checked with a calculator. The Implementation chapter provides an example of how GPT-4 can use a calculator tool to do calculations. A further point related to cognitive restrictions is that older people's attention spans are often limited, and therefore exercises should not take too long, as nursing experts emphasized.

However, despite the relatively long duration of the second requirements workshop, the elderly woman had no problem maintaining attention. The elderly man struggled to keep up with the discussions during the workshop, primarily due to his mentioned severe hearing impairment. During the workshops, it became clear that the individual abilities and limitations of the two participating seniors are very different. This fact certainly underlines the importance of systems that can be adapted to individual needs. It should be emphasized again that with the help of, for example, an SQL database tool, it would be conceivable to retrieve information stored in databases about individuals to be able to respond to individual needs.

The last subquestion of the research question formulated in the Introduction chapter, which will be discussed and answered at the end of this master's thesis, is whether part of the research should concentrate on less complex robots.

- Should research focus more on less complex robots?

There is a lot of progress in AI research and more and more of our interactions with the digital world will be AI-mediated. However, a robot that can replace the fine motor skills and flexibility of a nurse seems far from this point in time. Currently, there is no system that understands the world itself as humans or animals do, as discussed in the Literature Review chapter. For the foreseeable future, (assistive) robots will remain specialized. However, it is also often very difficult to develop a robotic application that is specialized in one or a few tasks. An important question is the question of reliability.

If in industry a certain percentage of work pieces is wasted, it is maybe a rather small problem, but in the health care sector it depends a lot on the area of application. If, for example, a robot with direct human contact only rarely injures people, it would certainly not be good enough. If such robots were used on a large scale, this would happen constantly somewhere in the world. Today, there is no robot on the market that performs care tasks in direct physical contact with the patient (with the exception of small-scale physical contact, for example, through a robot spoon). [HBG<sup>+</sup>20]

A robot that establishes direct patient contact is the lifting robot ROBEAR. [Rik] Discussions with nursing experts revealed that ROBEAR is too immobile, nursing processes would become more inefficient, there would be no added value for nurses, and it could hardly be an economically viable solution for hospitals or retirement homes. For example, there are much cheaper lifting devices [HBG<sup>+</sup>20] on the market to support nurses and other medical personnel. The development of ROBEAR was eventually stopped, and it will not become a commercial product. The further development of the service robot Care-O-Bot 4 [IPA] has not been discontinued (so far). However, even after 26 years of development, this robot is still just a product vision.

As the thematic analysis revealed, addressing and supporting people emotionally is one key to motivate users, and a robot does not (necessarily) has to be intelligent to provide emotional support. An example of a comparatively less complex and non-intelligent emotional support robot that has been available on the market for years is the robot seal PARO. [oAISA] The participants in a study by Lee et al. liked PARO the most among different robots. Other robots such as Care-O-Bot are too technical and not easy to use, according to them. [LŠC<sup>+</sup>17] A nursing expert who participated in the second requirements workshop has experience with PARO and has a generally positive attitude toward it. However, another nursing expert said that they use therapy dolls for the treatment of patients with dementia in the nursing home, which are even much simpler and much cheaper than PARO.

Also worth mentioning in connection with comparatively less complex robotic solutions to problems due to demographic developments are vacuum cleaner robots and lawn mower robots. These are specialized in one task, already very widespread, inexpensive to purchase, and can play a role in helping older people live in their own home and avoid institutional care as long as possible.



---

The robot ball Sphero is also cheap to purchase and comparatively less complex. The hardware adjustments mentioned above would probably make it more expensive, but not an order of magnitude more expensive. As mentioned, the nursing experts were confident that the jointly developed (agentic) fitness application with Sphero could be promising for retirement homes.

This promising prototypical implementation, the data obtained in the workshops, and the literature review suggest that some research should focus more on less complex robotic solutions for less complex environments. Such robots could be developed more quickly into actual products that could have the potential to alleviate some of the problems that arise from demographic developments. Robots like Sphero that are already on the market appear to be particularly interesting. The latest developments in the area of large language models and multimodal models and the possibility to combine these models with tools offer new options for the development of companion robots based on simple hardware with which verbal communication is possible.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# List of Figures

1.1	Population development of Italy (top left) [Stad] , Hungary (top middle) [Stah], Japan (top right) [Staf] and age structures of of Italy (bottom left) [Stac], Hungary (bottom middle) [Stag] and Japan (bottom right) [Stae]	2
1.2	A classification of robots according to Heerink et al. [HKEW10]	4
1.3	Care-O-Bot 4 (left) [IPA], PARO (right) [SHP <sup>+</sup> 21]	5
1.4	Delivery robots of the company Aethon [Inca]	6
1.5	The caregiver robot ROBEAR helps a person from a sofa to the wheelchair [Rik]	7
2.1	Encoder and decoder of the transformer architecture (left image) [VSP <sup>+</sup> 17] Pipeline to improve Sparrow (right image). [GMT <sup>+</sup> 22]	14
2.2	Ameca by Engineered Arts [Arta]	15
2.3	Robot used for PaLM-SayCan and PaLM-E (left) [DXS <sup>+</sup> 23], PaLM-E ingests different modalities (right) [DXS <sup>+</sup> 23]	17
2.4	The robot learns to walk in just one hour (first three images) and adapts to withstanding pushes (right image). [WEH <sup>+</sup> 23]	18
2.5	A possible architecture for a future autonomous intelligence [LeC22]	19
2.6	Caregivers lift a person from the bed into the wheelchair using a lifting device. [HBG <sup>+</sup> 20]	20
3.1	Timeline of the master’s thesis	34
4.1	Sphero Bolt [Ince]	38
4.2	Overview of drive mode apps (left), Joystick App (middle), Golf App (right)	39
4.3	Overview of games in game mode (first image), Exile 2 App (second image), Space Drifter (third image), Tour game (fourth image)	40
4.4	A neural network architecture for exercise classification (left), Confusion matrices after applying the neural network on test data of three different exercises (right)	41
4.5	The dumbbell exercise is performed	42
4.6	First example conversation with GPT-4	44
4.7	Second example conversation with GPT-4	45
		85

4.8	Results for the prompts: “What date was 123890 days ago?” (left) and “What is the name of Schwarzenegger’s eldest daughter? Multiply her age by 10” (right) . . . . .	46
4.9	A page from the “Physio Vital” brochure from Novartis [Nov] . . . . .	47
4.10	Illustration of the workflow of the final implementation . . . . .	49
4.11	The LLM decides to look for information in the vector database based on the request . . . . .	50
4.12	The LLM decides to perform a Google Search based on the request . . . . .	51
4.13	The LLM decides to start the dumbbell exercise based on the request . . . . .	51
4.14	Code snippets of the Agent class and the ExcerciseTool class of the implementation . . . . .	52
5.1	The humanoid robot Pepper [Gro] . . . . .	57
5.2	Themes identified by thematic analysis . . . . .	59





Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Bibliography

- [AAA<sup>+</sup>23] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [ABB<sup>+</sup>22] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, et al. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*, 2022.
- [ADB<sup>+</sup>23] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. MusiclM: Generating music from text. *arXiv preprint arXiv:2301.11325*, 2023.
- [ADBB17] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. A brief survey of deep reinforcement learning. *arXiv preprint arXiv:1708.05866*, 2017.
- [Arta] Engineered Arts. Ameca. <https://www.engineeredarts.co.uk/de/robot/ameca/>, Last accessed: December 2023.
- [Artb] Engineered Arts. Ameca conversation using gpt 3. <https://www.youtube.com/watch?v=EWACmFLvpHE>, Last accessed: December 2023.
- [Aus] Statistik Austria. Gesundheitsausgaben. <https://www.statistik.at/statistiken/bevoelkerung-und-soziales/gesundheit/gesundheitsversorgung-und-ausgaben/gesundheitsausgaben>.
- [BC06] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [BC19] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597, 2019.



- [Bea15] Françoise Beaufays. The neural networks behind google voice transcription. *Google Research blog*, 2015.
- [BGJ<sup>+</sup>23] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science*. <https://cdn.openai.com/papers/dall-e-3.pdf>, 2(3):8, 2023.
- [BHK08] Hans-Bernd Brosius, Alexander Haas, and Friederike Koschel. *Methoden der empirischen Kommunikationsforschung*. Springer, 2008.
- [BHvdHdW17] Sandra Bedaf, Claire Huijnen, Renée van den Heuvel, and Luc de Witte. Robots supporting care for elderly people. In *Robotic assistive technologies*, pages 309–332. CRC Press, 2017.
- [Bux11] H Buxel. Wie pflegende am arbeitsplatz zufriedener werden. *Die Schwester Der Pfleger*, 50(5):426–430, 2011.
- [Cap01] Karel Capek. *R.U.R.* Dover Thrift S. Dover Publications, Mineola, NY, August 2001.
- [CGK23] Charlotte Caucheteux, Alexandre Gramfort, and Jean-Rémi King. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, pages 1–12, 2023.
- [CND<sup>+</sup>22] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*, 2022.
- [dACC23] Lucas BV de Amorim, George DC Cavalcanti, and Rafael MO Cruz. The choice of scaling technique matters for classification performance. *Applied Soft Computing*, 133:109924, 2023.
- [DALT22] Ayelet Dembovski, Yael Amitai, and Shelly Levy-Tzedek. A socially assistive robot for stroke patients: Acceptance, needs, and concerns of patients and informal caregivers. *Frontiers in rehabilitation sciences*, 2:793233, 2022.
- [Dav89] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.
- [DBK<sup>+</sup>20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

- [DTB<sup>+</sup>23] Debadutta Dash, Rahul Thapa, Juan M Banda, Akshay Swaminathan, Morgan Cheatham, Mehr Kashyap, Nikesh Kotecha, Jonathan H Chen, Saurabh Gombar, Lance Downing, et al. Evaluation of gpt-3.5 and gpt-4 for supporting real-world information needs in healthcare delivery. *arXiv preprint arXiv:2304.13714*, 2023.
- [DXS<sup>+</sup>23] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.
- [Ede03] Walter Edelman. Intrinsische und extrinsische motivation. *Grundschule*, 4:30–32, 2003.
- [fF] Bundesministerium für Finanzen. Langfristige budgetprognose 2022. <https://www.bmf.gv.at/themen/budget/publikationen/langfristige-budgetprognose.html>, Last accessed: November 2023.
- [FM12] Juan Fasola and Maja J Mataric. Using socially assistive human–robot interaction to motivate physical exercise for older adults. *Proceedings of the IEEE*, 100(8):2512–2526, 2012.
- [GACG17] Nicolas Garcia-Aracil, Alicia Casals, and Elena Garcia. Rehabilitation and assistive robotics, 2017.
- [GDER<sup>+</sup>19] Francisco Gomez-Donoso, Félix Escalona, Francisco Miguel Rivas, Jose Maria Cañas, Miguel Cazorla, et al. Enhancing the ambient assisted living capabilities with a mobile robot. *Computational intelligence and neuroscience*, 2019, 2019.
- [GHKW13] Birgit Graf, Torsten Heyer, Barbara Klein, and Frank Wallhoff. Servicerobotik für den demografischen wandel. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 8(56):1145–1152, 2013.
- [GMHS20] Victoria J Gaesser, William M Maakestad, Erik S Hayes, and Stephen J Snyder. Motivational coaching improves intrinsic motivation in adult fitness program participants. *International Journal of Exercise Science*, 13(5):1167, 2020.
- [GMT<sup>+</sup>22] Amelia Glaese, Nat McAleese, Maja Trębacz, John Aslanides, Vlad Firoiu, Timo Ewalds, Maribeth Rauh, Laura Weidinger, Martin Chadwick, Phoebe Thacker, et al. Improving alignment of dialogue agents via targeted human judgements. *arXiv preprint arXiv:2209.14375*, 2022.
- [Goo] Google. Mediapipe. [https://developers.google.com/mediapipe/solutions/vision/pose\\_landmarker](https://developers.google.com/mediapipe/solutions/vision/pose_landmarker), Last accessed: December 2023.

- [Gro] Aldebaran United Robotics Group. Pepper. <https://www.aldebaran.com/de/pepper>, Last accessed: December 2023.
- [HBG<sup>+</sup>20] Eva Hornecker, Andreas Bischof, Philipp Graf, Lena Franzkowiak, and Norbert Krüger. The interactive enactment of care technologies and its implications for human-robot-interaction in care. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pages 1–11, 2020.
- [HBKE16] Frank Hänsel, Sören D Baumgärtner, Julia M Kornmann, and Fabienne Ennigkeit. *Sportpsychologie*. Springer, 2016.
- [HKEW10] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. Assessing acceptance of assistive social agent technology by older adults: the almere model, 2010.
- [HLBN19] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019.
- [HLNB20] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*, 2020.
- [HPBL23] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.
- [HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [HW24] Elionardo Feliciano Hanbang Wang. Sphero v2, 2024. <https://github.com/artificial-intelligence-class/spherov2.py>, Last accessed: December 2023.
- [Inca] AETHON Inc. Tug. <https://aethon.com/mobile-robots-for-healthcare/>, Last accessed: August 2024.
- [Incb] LangChain Inc. Langchain framework. <https://python.langchain.com>, Last accessed: December 2023.
- [Incc] Pinecone Systems Inc. Pinecone vector database. <https://www.pinecone.io/>, Last accessed: January 2024.
- [Incd] Sphero Inc. Sphero sensor. <https://sites.google.com/sphero.com/sphero-edu-javascript-wiki/sensors>, Last accessed: December 2023.

- [Ince] Sphero Inc. Spherobolt. <https://sphero.com/products/sphero-bolt>, Last accessed: December 2023.
- [IPA] Fraunhofer IPA. Care-o-bot 4. <https://www.care-o-bot.de/de/care-o-bot-4.html>, Last accessed: November 2023.
- [KFS<sup>+</sup>15] Ralf Kittmann, Tim Fröhlich, Johannes Schäfer, Ulrich Reiser, Florian Weißhardt, and Andreas Haug. Let me introduce myself: I am care-o-bot 4, a gentleman robot. 2015.
- [Lea] Open LLM Leaderboard. Open llm leaderboard. [https://huggingface.co/spaces/HuggingFaceH4/open\\_llm\\_leaderboard](https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard), Last accessed: February 2024.
- [LeC22] Yann LeCun. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62, 2022.
- [LŠC<sup>+</sup>17] Hee Rin Lee, Selma Šabanović, Wan-Ling Chang, Shinichi Nagata, Jennifer Piatt, Casey Bennett, and David Hakken. Steps toward participatory design of social robots: mutual learning with older adults with depression. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 244–253, 2017.
- [MBM<sup>+</sup>16] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937. PMLR, 2016.
- [Mic] Microsoft. Azure ai speech. <https://azure.microsoft.com/en-us/products/ai-services/ai-speech>, Last accessed: December 2023.
- [MKS<sup>+</sup>13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- [Nov] Novartis. Physio vital. [https://seekajak.ch/\\_media/physio-vital-voltaren-app-taschentrainer-a6-160208.pdf](https://seekajak.ch/_media/physio-vital-voltaren-app-taschentrainer-a6-160208.pdf), Last accessed: January 2024.
- [NT05] Marcus B Nicol and Amanda G Thrift. Knowledge of risk factors and warning signs of stroke. *Vascular health and risk management*, 1(2):137, 2005.
- [oAISA] National Institute of Advanced Industrial Science and Technology (AIST). Paro. <https://unit.aist.go.jp/hiiri/topics/03-en.html>, Last accessed: November 2023.

- [OEC] OECD. Working age population. <https://www.oecd.org/en/data/indicators/working-age-population.html?oecdcontrol-f75fe24c3f-var3=1989&oecdcontrol-f75fe24c3f-var4=2022&oecdcontrol-d6804ae080-var1=EU27>, Last accessed: November 2023.
- [Opea] OpenAI. Openai documentation. <https://platform.openai.com/docs/overview>, Last accessed: January 2024.
- [Opeb] OpenAI. Whisper. <https://openai.com/index/whisper/>, Last accessed: December 2023.
- [PSMG<sup>+</sup>19] Jose Carlos Pulido, Cristina Suarez-Mejias, Jose Carlos Gonzalez, Alvaro Duenas Ruiz, Patricia Ferrand Ferri, Maria Encarnacion Martinez Sahuquillo, Carmen Echevarria Ruiz De Vargas, Pedro Infante-Cossio, Carlos Luis Parra Calderon, and Fernando Fernandez. A socially assistive robotic platform for upper-limb rehabilitation: a longitudinal study with pediatric patients. *IEEE Robotics & Automation Magazine*, 26(2):24–39, 2019.
- [Rik] Sumitomo Riko Company Limited Riken. The strong robot with the gentle touch. [https://www.riken.jp/en/news\\_pubs/research\\_news/pr/2015/20150223\\_2/](https://www.riken.jp/en/news_pubs/research_news/pr/2015/20150223_2/), Last accessed: November 2023.
- [Rob05] A Roberts. *The history of science fiction*. Palgrave Histories of Literature. Palgrave Macmillan, Basingstoke, England, November 2005.
- [RYKY<sup>+</sup>23] Nur Lidiya Abdul Rashid, LEOW Yihong, Piyanee Klainin-Yobas, ITOH Sakiko, and WU Vivien Xi. The effectiveness of a therapeutic robot, ‘paro’, on behavioural and psychological symptoms, medication use, total sleep time and sociability in older adults with dementia: A systematic review and meta-analysis. *International Journal of Nursing Studies*, page 104530, 2023.
- [SDYD<sup>+</sup>23] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Tool-former: Language models can teach themselves to use tools. *arXiv preprint arXiv:2302.04761*, 2023.
- [SE09] Bertram Scheufele and Ines Engelmann. *Empirische Kommunikationsforschung*. UVK-Verlag-Ges., 2009.
- [Ser] SerpApi. Search engine results page api. <https://serpapi.com/>, Last accessed: January 2024.

- [SHE<sup>+</sup>99] Rainer Schnell, Paul B Hill, Elke Esser, et al. Methoden der empirischen sozialforschung. 1999.
- [SHK<sup>+</sup>14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [SHP<sup>+</sup>21] Takanori Shibata, Lillian Hung, Sandra Petersen, Kate Darling, Kaoru Inoue, Katharine Martyn, Yoko Hori, Geoffrey Lane, Davis Park, Ruth Mizoguchi, et al. Paro as a biofeedback medical device for mental health in the covid-19 era. *Sustainability*, 13(20):11502, 2021.
- [Sph] Sphero. Sphero bolt. <https://support.sphero.com/category/sx9c3rpowl-bolt>, Last accessed: December 2023.
- [SPL22] Daniel Sonnet, V Perceptron, and D Learning. Neuronale netze kompakt. *Springer Nature*, pages 71–84, 2022.
- [SRS<sup>+</sup>14] Luise Süßenbach, Nina Riether, Sebastian Schneider, Ingmar Berger, Franz Kummert, Ingo Lütkebohle, and Karola Pitsch. A robot as fitness companion: towards an interactive action-based motivation model. In *The 23rd IEEE international symposium on robot and human interactive communication*, pages 286–293. IEEE, 2014.
- [Staa] Statista. Altersstruktur in Österreich im jahr 2021 und prognose für 2030 bis 2100. <https://de.statista.com/statistik/daten/studie/688475/umfrage/prognose-zur-alterstruktur-in-oesterreich/>, Last accessed: November 2023.
- [Stab] Statista. Bevölkerung von Österreich im jahr 2021 und prognose für 2030 bis 2100. <https://de.statista.com/statistik/daten/studie/687869/umfrage/bevoelkerungsprognose-fuer-oesterreich/>, Last accessed: November 2023.
- [Stac] Statista. Italien: Altersstruktur der bevölkerung von 1950 bis 2021 und prognosen bis 2050. <https://de.statista.com/statistik/daten/studie/167721/umfrage/alterstruktur-in-italien/>, Last accessed: November 2023.
- [Stad] Statista. Italien: Gesamtbevölkerung von 1950 bis 2021 und prognosen bis 2050. <https://de.statista.com/statistik/daten/studie/19304/umfrage/gesamtbevoelkerung-in-italien/>, Last accessed: November 2023.

- [Stae] Statista. Japan: Altersstruktur von 1950 bis 2021 und prognosen bis 2050. <https://de.statista.com/statistik/daten/studie/165976/umfrage/altersstruktur-in-japan/>, Last accessed: November 2023.
- [Staf] Statista. Japan: Gesamtbevölkerung von 1950 bis 2021 und prognosen bis 2050. <https://de.statista.com/statistik/daten/studie/19305/umfrage/gesamtbevoelkerung-in-japan/>, Last accessed: November 2023.
- [Stag] Statista. Ungarn: Altersstruktur der bevölkerung von 1950 bis 2021 und prognosen bis 2050. <https://de.statista.com/statistik/daten/studie/272447/umfrage/altersstruktur-in-ungarn/>, Last accessed: November 2023.
- [Stah] Statista. Ungarn: Gesamtbevölkerung von 1950 bis 2021 und prognosen bis 2050. <https://de.statista.com/statistik/daten/studie/19301/umfrage/gesamtbevoelkerung-von-ungarn/>, Last accessed: November 2023.
- [SW11] Takanori Shibata and Kazuyoshi Wada. Robot therapy: a new approach for mental healthcare of the elderly—a mini-review. *Gerontology*, 57(4):378–386, 2011.
- [TDS22] Rachel Teng, Yichen Ding, and Kay Choong See. Use of robots in critical care: systematic review. *Journal of Medical Internet Research*, 24(5):e33380, 2022.
- [TV21] TÜV-Verband. Sicherheit und künstliche intelligenz, 2021. [https://www.tuev-verband.de/?tx\\_epxelo\\_file\[id\]=856779&cHash=1af8a3f0e6c845423fdd637c8dbcd080](https://www.tuev-verband.de/?tx_epxelo_file[id]=856779&cHash=1af8a3f0e6c845423fdd637c8dbcd080), Last accessed: February 2024.
- [uMvGM00] Katja Mruck unter Mitarbeit von Günter Mey. Qualitative sozialforschung in deutschland, 2000. <https://www.qualitative-research.net/index.php/fqs/article/view/1114/2465>.
- [VOSK22] Karthik Valmeekam, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. Large language models still can’t plan (a benchmark for llms on planning and reasoning about change). *arXiv preprint arXiv:2206.10498*, 2022.
- [VSP<sup>+</sup>17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.



- [WEH<sup>+</sup>23] Philipp Wu, Alejandro Escontrela, Danijar Hafner, Pieter Abbeel, and Ken Goldberg. Daydreamer: World models for physical robot learning. In *Conference on Robot Learning*, pages 2226–2240. PMLR, 2023.
- [Wora] Worldbank. Fertility rate, total (births per woman). <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN>, Last accessed: November 2023.
- [Worb] Worldbank. Gender data portal. <https://genderdata.worldbank.org/indicators/sp-dyn-tfirt-in>, Last accessed: November 2023.
- [WSMK08] Kazuyoshi Wada, Takanori Shibata, Toshimitsu Musha, and Shin Kimura. Robot therapy for elders affected by dementia. *IEEE Engineering in medicine and biology magazine*, 27(4):53–60, 2008.
- [WYG<sup>+</sup>21] Jianguo Wang, Xiaomeng Yi, Rentong Guo, Hai Jin, Peng Xu, Shengjun Li, Xiangyu Wang, Xiangzhou Guo, Chengming Li, Xiaohai Xu, et al. Milvus: A purpose-built vector data management system. In *Proceedings of the 2021 International Conference on Management of Data*, pages 2614–2627, 2021.
- [YZY<sup>+</sup>22] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.
- [ÖP] ÖPIA. Hochaltrigenstudie. [http://www.oepia.at/hochaltrigkeit/?page\\_id=17](http://www.oepia.at/hochaltrigkeit/?page_id=17), Last accessed: November 2023.