Universiteit Leiden
The Netherlands

Opleiding Informatica

The effect of humanoid embodiment on the perceived intelligence, usability, likability and animacy of reception robots

Luuk van Beusichem

Supervisors:
Dr.ir. D.J. Broekens & Prof.dr.ir. F.J. Verbeek

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl 02/07/2023

Abstract

The use of robots in daily life has become increasingly popular in recent times. Robots are nowadays used in different fields of work such as education and elderly care. A lot of research has already been done within the research field of Human-Robot Interaction (HRI) to improve the interaction. Anthropomorphism is for example an important factor. Research has also been conducted on the physical and virtual appearance of robots, as well as on robots with different personalities.

In this study, we investigate the effect of using different humanoid robots in a receptionist setting. We investigate the effect on the following aspects: Animacy, perceived intelligence, likeability, and usability. To investigate this, we divide this study into three parts.

First, we conducted a requirement study, where we attempted to determine, through questioning students, what the key tasks should be that the reception robot should be able to perform.

Additionally, we conducted a pilot study. The purpose of this pilot study was to test the interaction we built based on the key tasks identified in the requirement study. The findings from the pilot study were used to enhance the final interaction used for the effect study.

Lastly, we conducted an effect study. The purpose of this effect study was to examine the impact of various robot embodiment on animacy, likeability, perceived intelligence, and usability.

The results show that only in terms of animacy, a statistically significant effect can be found. The NAO robot had the highest perceived animacy and this difference was significant with the Alpha Mini. For the other three aspects, there was no statistically significant effect found.

Contents

1 Introduction

The use of robots in daily life has become increasingly popular in recent times. Robots are nowadays used in different fields of work such as education and elderly care. This increase in the use of robots results in humans and robots interacting more often.

A lot of research has already been done within the research field of Human-Robot Interaction (HRI) to improve the interaction. For example, research shows the importance of using anthropomorphism [Fin12]. Anthropomorphism is the attribution of human characteristics to non-human artifacts [Fin12]. The main conclusion that can be drawn from this study is that anthropomorphism is a factor to consider in the design of robots. However, anthropomorphism is just one factor in the design of HRI. When designing a robot, one should respect 3 main considerations [CDK02]: "the need to retain an amount of robot-ness so that the user does not develop false expectations of the robots' emotional abilities but realizes its machine capabilities; the need to project an amount of humanness so that the user will feel comfortable engaging with the robot; and the need to convey an amount of product-ness so that the user will feel comfortable using the robot." [CDK02].

Another study shows that depending on the context, humans prefer a robot with a different personality [BTP14]. In this study, the personality is simulated by setting different values for speech parameters such as pitch, volume, and speech rate. The study showed that people appreciate different values for the parameters depending on the task the robot had to perform.

In our experiment, we focus on robots that serve as reception robots. For this experiment, we use three types of robots: Alpha Mini, NAO, and Pepper, which will be discussed in more detail in section 3.1. All three robots will individually participate in a scenario where each robot functions as a receptionist and interacts with a person. All robots can perform the same basic actions such as moving their arms, detecting faces, and producing speech. The main difference between the three robots is their embodiment. The main aim of this study is to examine how user perception changes when different embodiments of the robots are used.

1.1 Motivation and related work

To examine the effect of embodiment, we followed an approach that has been performed similarly in other HRI studies on the embodiment of a robot. The study of Nishio et al [TN21] for example investigated the effect of using a virtual robot versus a physical robot on a conversation task with the elderly. Both robots were able to have the same type of conversation. Namely a conversation about experiences from the past. This type of conversation allows for a deeper relationship to be built between those holding the conversation [LC19]. In the study of Nishio et al [TN21], the participants were split up into two groups. One group used the virtual robot and one group used the physical robot. The physical robot was preferred by the elderly. Another example is the study of Pauwe et al [RAPK15] that follows the same procedure as we do in our experiment by using three different embodied physical robots. These robots acted as physiotherapy assistants. The robots differed in appearance to varying degrees of realism (humanoid, crocodile, and vehicle) and the participants were assigned to one of the three robots. The difference between using a realistic robot and a less realistic robot was minor. More important was the perception of affordance the robots offered. The last example is the study of Bazzano & Lamberti [BL18], similar to the study of Nishio et al [TN21] the effect is investigated of a virtual robot versus a physical robot. This was done in a reception setting. Physical robots were generally preferred to virtual agents.

In short, research has been conducted on the embodiment of receptionist robots, with a focus on their physical presence or absence. Research that has been conducted on different physical embodiments focussed more on different types of embodiment. It has not been investigated what the effect is of using different humanoid-like robots in a receptionist setting. Therefore, the purpose of this paper is to examine three different humanoid-like receptionist robots to shed light on the distinctions between them.

The results that are obtained from our experiment can be used in the design of future reception robots. This study may provide new insights regarding designing the embodiment of a robot, which might improve people's acceptance of robots in everyday life.

2 Research question

In this work, we investigate the effect of different humanoid embodiments on the perceived intelligence, likability, usability, and animacy of reception robots. In an attempt to understand why potential differences in perception occur, we also investigate qualitatively what the user´s experience is using an open-ended question.

2.1 Hypotheses

Our hypothesis is that we expect to see that people prefer interaction with a more humanoid robot, compared to a less humanoid robot, as measured by a higher rating on the four outcome variables mentioned above. The reason for this expectation is the paper of Fink [Fin12] which is described in section 1.1. Fink writes that anthropomorphism could be useful in the design of robots for certain interaction scenarios [Fin12]. The Pepper robot is the most humanlike robot of the 3 robots we investigated. The study of Kontogiorgos [DKG20] shows that people prefer humanlike robots even though it fails the same task as often as a non-humanlike robot. The same study shows that human likeness contributes negatively to the interaction when the robot had to perform a task that has greater consequences for the human. Our reception robot scenario does not contain such consequences for humans. Therefore, it is likely that humanlike robots are preferred.

A pitfall could be that the robot would be too human-like. As mentioned in Section 2, a robot should always retain some sort of robot-ness to prevent the user from developing false expectations [CDK02]. This is the case for all three embodiments we use in the study. Another aspect relevant to our hypothesis is the uncanny valley [MMK12] (Figure 1). The uncanny valley refers to the perceptual phenomenon that near, but not perfect, human realism results in a sudden drop in perceived familiarity, effectively making the robot look “strange” or “weird”. However, the embodiments used in this study do not resemble humans close enough for the uncanny valley to occur.

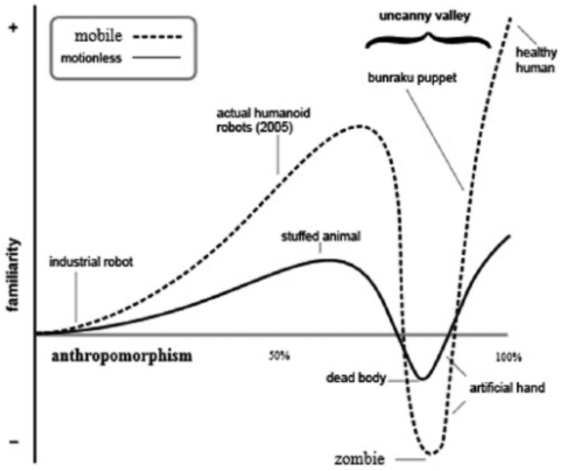

Figure 1: Uncanny valley

3 Research design

This research consists of three parts. For the first part, we performed a requirement study in which we investigate what people would ask a receptionist at the Snellius building. Based on the results of the first part we designed an interaction scenario that was used for parts two and three of this study.

The second part was a pilot experiment. In this experiment, we tested the robustness of an interaction scenario, we created based on the results of the first part. We tested our interaction scenario and noted (unexpected) errors in the interaction. After we finished our pilot experiment, we updated our interaction scenario.

The third and final part was an effect study where we investigate the effect of different humanoid embodiments on the perceived intelligence, usability, likability, and animacy of reception robots. For this, we used the updated interaction scenario.

In the following three chapters, we will take a closer look at each part.

4 Requirement study

To find out which tasks the robot should perform and how the interaction should exactly look like, we interviewed primary and tertiary users of the reception robot. Primary users are users who will interact with the robot. This type of user can give feedback on how they envision the interaction with a reception robot. Some aspects are: Which tasks should it perform? What should be the flow of the interaction?

Tertiary users, in our case, are people who currently work as a receptionist. Their work is affected when a reception robot (partially) takes over their tasks. This type of user can give information about what the current interaction between visitor and receptionist looks like and which tasks a receptionist performs and a reception robot eventually could take over.

4.1 Materials

For the requirement study, we used a google forms document where primary users could enter their answers.

4.2 Experimental Setup/approach

As previously mentioned, we interviewed primary and tertiary users of the robot receptionist. In our case, the primary users were students at Leiden University. 23 students received a google forms questionnaire and were asked which questions they have asked the receptionist at the Snellius building. The primary users are also asked about their gender and age. All primary users belonged to the 19-25 age group. Figure 2 shows the gender distribution.

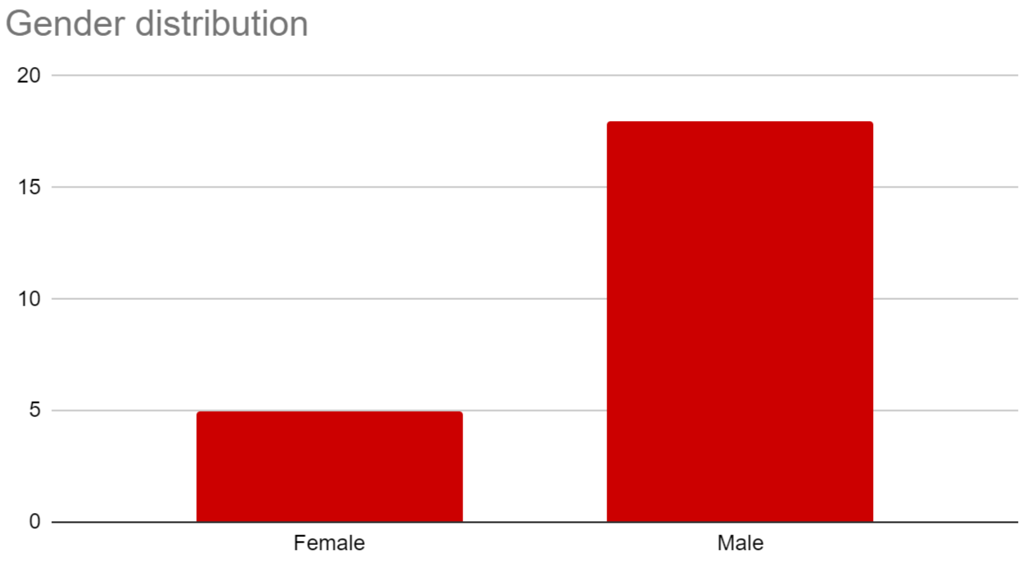

Figure 2: Gender distribution requirement study

The tertiary user we interviewed, was someone who currently works as a receptionist at the Snellius building. A conversation was held with one of the receptionists of the Snellius building to give inspiration for designing an interaction flow.

4.3 Measures

Each primary user received a google forms document with the question: What questions have you ever asked the reception of the Snellius building? The participant could then write down all his questions.

From the conversation with the receptionist, we noted the key points of the conversation.

4.4 Results

Table 1 shows the answers the primary users gave to the question: What questions have you ever asked the reception of the Snellius building? We see that the location of a room and the location of a person are frequently asked questions, so therefore we decided to add these two tasks to the capabilities of our receptionist robot. Other questions that were frequently asked were practical information, key delivery, and the presence of a person. The last two were not possible due to the limitations of the robots, so, therefore, we only added the practical information part.

Type of question	Times asked
Location of a room	13
Location of a person	7
Opening a room	1
Asking for a free room	2
Asking for teaching materials	2
Public transport information	1
Practical information about Snellius building	3
Asking for key	4
Asking for presence of person	3
Where does the workshop take place	1
Delivery package	1
Lost and found	2

Table 1: Frequency analysis of participant answers

The conversation with the human receptionist led to the insight that a receptionist from the Snellius acts more as a host. In addition, the receptionist could provide insight into the practical information questions about the Snellius building. This led to the following four main subcategories:

- Opening hours
- Facilities
- Events/activities
- Contact details

In brief, we can state that our robot will perform the following three tasks during the pilot experiment and the effect study:

- Giving the route to the office of a person
- Giving the route to a room in the Snellius building
- Giving practical information about the Snellius building, divided into four subcategories

For the first two tasks, we made use of the code provided by Leendert van der Plas [vdP20] for his bachelor's thesis. This code already contained a working version of the direction to a room or office. It also contained the web interface. For the last task we created an additional webpage that contained the four main categories of practical information as earlier mentioned.

5 Pilot study

In the pilot experiment, we tested the robustness of the experiment of the effect study. This experiment aimed to check whether there were any technical errors and unclarities in the interaction present.

5.1 Materials

For this experiment, a robot was developed that acted as a reception robot that could perform the three tasks described in the results section of the requirement study. We used three different robots, which will be discussed in more detail in the next subsections. We have chosen these three robots, because these robots differ in appearance in such a way, that we could examine the influence of their embodiment on the interaction.

The software used to program the robots can be found at portal.robotsindeklas.nl/python/javascript. All robots can be programmed through this platform. In the next three subsections, the three robots will be discussed briefly.

5.1.1 Alpha Mini

The first robot we discuss is the Alpha Mini. This is also the smallest robot we used for this experiment. The Alpha Mini is developed by the company UBTECH Robotics and is mostly used for education purposes. Except for studies in the field of education, the alpha Mini is also widely studied in the field of elderly care. HRI plays a big role in both fields. The small, cute appearance of the Alpha Mini robots is very different from the NAO and Pepper. The robot can walk, however only over short distances. The sensors of the Alpha Mini include several touch sensors and a camera that can recognize faces. The Alpha Mini can produce speech and can move his arms, legs, and head. It has no fingers and a square face.

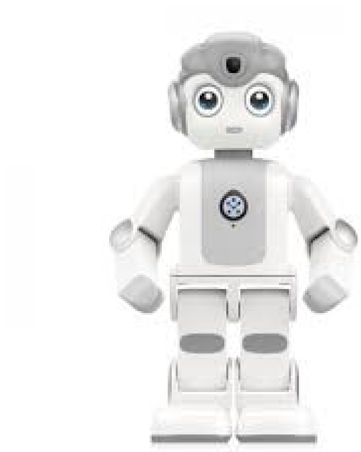

Figure 3: Alpha Mini

5.1.2 NAO

The second robot we discuss is the NAO robot developed by Aldebaran Robotics (later Softbank Robotics). The NAO is mainly used for educational and research purposes, but there are a few cases where the NAO has been used for health care purposes. The robot has four microphones. The NAO can localize sounds. Further, the NAO has two cameras, which can be used together with the NAO software for face and shape reorganization. Like the Alpha Mini robot, the NAO robot can produce speech and can move his arms, legs, and head. It can also walk, but again only over short distances. The NAO robot is a bit bigger compared to the Alpha Mini. In addition, the robot looks a bit more agile compared to the Alpha Mini. It is the only robot of the three where some joints are not fully covered. It has three fingers on each hand. His eyes are relatively small compared to the other robots. It is the only robot that has colorful parts in the embodiment. It has a rectangular face.

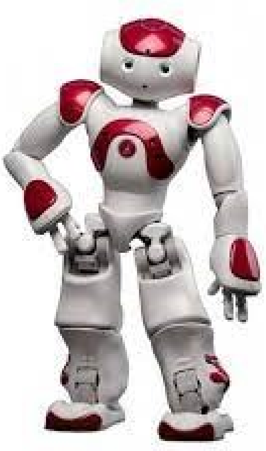

Figure 4: NAO

5.1.3 Pepper

The last robot we discuss is the Pepper robot. The Pepper robot is developed by the same company as the NAO robot, Aldebaran Robotics (later Softbank Robotics). The use of robots in everyday life is mainly aimed at hospitality jobs such as host and most relevant receptionist. Like the previous two robots we discussed, the Pepper robot has been used to study HRI. When we look at the design of the robot, it is immediately apparent that the robot has no legs. Instead, it has a mobile base with three wheels built in. This allows the robot to move over a longer distance. For the same reasons as the NAO robot, the Pepper robot has four microphones and two cameras. Additionally, it has a depth sensor to estimate the depth of objects. The head and hands include touch sensors. The mobile base has three bumper sensors. Like the Alpha Mini and NAO robot, the Pepper robot can produce speech and can move his arms and head. The pepper is the biggest robot of the three with a size of 120 cm. It is the only robot that has 10 fingers. It has a round face. A tablet screen is placed on the torso of the robot that can serve as an interaction tool. However, we will not use this tablet screen since it does not work optimally. Instead for the pilot study, we used the screen of an external laptop that was placed next to the robot.

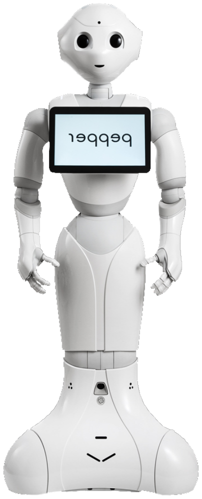

Figure 5: Pepper

5.2 Experimental Setup/approach

For the pilot study, we asked random passersby to have a short interaction with the robot. The experiment was conducted in the Snellius Building. For our experiment, we conducted the following experiment setup that is also used in the effect study: One of the robots is placed next to the desk of the reception of the Snellius Building. People enter the building and face one of the robots.

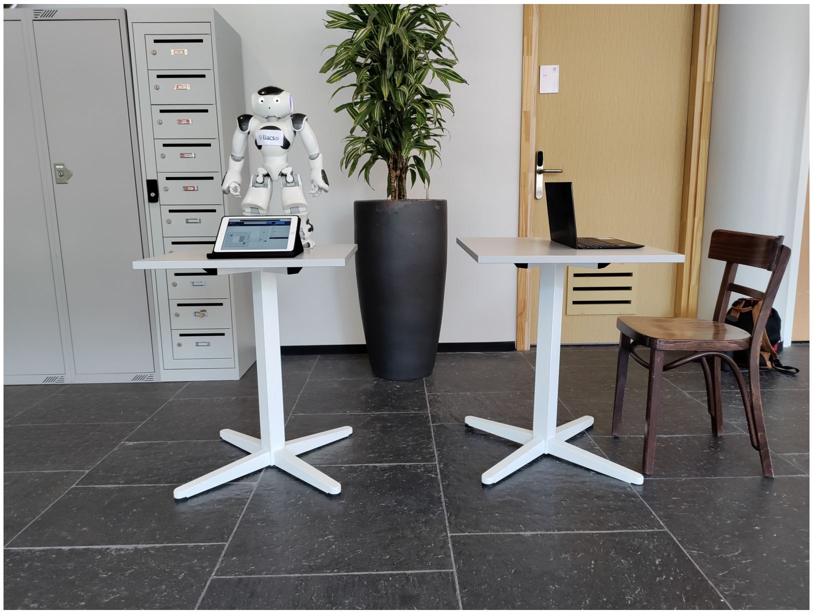

Figure 6: Experimental setting

The interaction can start. The robot introduces itself, tells which tasks it can do, and asks what task it should perform. To interact with the robot, participants can command the robot via speech.

The interaction scenario mainly consists of a dialogue between the robot and the user. proceeds in the following way: The robot asks which task it should perform, and the participant selects the preferred task by saying the corresponding keyword. The robot can perform the three aforementioned tasks. The participant chooses one of the three tasks or he/she chooses to end the conversation by saying stop. Figure 7 shows the interaction flow between the participant and robot for the pilot study.

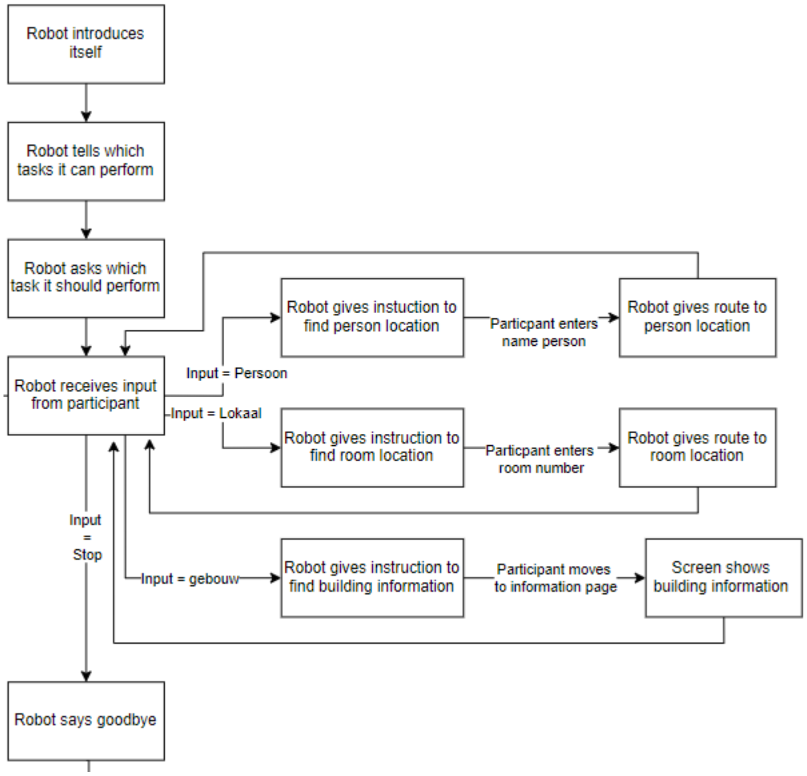
Robot introduces itself
Robot tells which tasks it can perform
Robot asks which task it should perform
Robot receives input from participant
Input = Persoon
Input = Lokaal
Robot gives instuction to find person location
Particpant enters name person
Robot gives route to person location
Robot gives instruction to find room location
Particpant enters room number
Robot gives route to room location
Input = Stop
Input = gebouw
Robot gives instruction to find building information
Participant moves to information page
Screen shows building information
Robot says goodbye

Figure 7: Interaction flow pilot study

Each participant had an interaction with only one of the three robots. For the NAO robot, there were 8 participants. For the Pepper and Alpha Mini robots, there were 5 participants each.

5.3 Measures

We measured in this study which technical errors and issues with the interaction occurred and how often this was the case.

5.4 Results

Table 2 shows the type of errors and how often they occurred for the interaction with the NAO robot.

Error type	Times occurred
Volume too low	2
Searching for people who were not in database	3
Participant did not fully listen to instruction	5
Robot recognized keyword incorrectly	1

Table 2: Frequency analysis of error types NAO

Table 3 shows the type of errors and often they occurred for the interaction with the Pepper robot.

Error type	Times occurred
Speech rate too fast	2
Searching for people who were not in database	2
Participant did not fully listen to instruction	2

Table 3: Frequency analysis of error types Pepper

Table 4 shows the type of errors and how they occurred for the interaction with the Alpha Mini robot.

Error type	Times occurred
Speech rate too slow	1
Searching for people who were not in database	1
Participant did not fully listen to instruction	3
Robot recognized keyword incorrectly	1

Table 4: Frequency analysis of error types Alpha Mini

As can be seen, some errors occurred more than others and not all errors occurred with all robots. We started the pilot experiment with the NAO robot. In the first two trials, the volume of the robot was too low. The place where the robot was standing, was very noisy. We changed the volume two times and after the second time, the volume was perfect for all robots. The second error was that people were looking for the office of people who were not yet in the database. This was a common problem for all robots and was fixed by updating the database. The error that happened the most was that participants did not fully listen to the instructions the robot gave. For example, the robot told a participant to say “persoon” for the location of a person’s room. The participant immediately reacted by saying “persoon”, however, the robot was still giving instructions on how to use the robot. The robot is only able to detect a keyword after giving the instructions completely. This confused the participants as to why the robot had not understood them. This problem was solved by having the robot explicitly state that the participant must wait before the robot finished giving instructions. Another error that occurred was that the robot misunderstood the keyword spoken by the participant. For example, a robot once recognized a keyword, while the participant had not yet said anything. We could not improve the listening skills of the robot, because it is built-in. However, we have expanded the instructions with an extra keyword “instructies”. If the person forgets which keywords, he/she can use, he/she is always able to say “instructies” to receive the keywords again. The final error that occurred was that the speech rate was not set correctly. The speech rate of the Pepper robot is different from that of the NAO and Alpha Mini robots. The first time we held the pilot with the Alpha Mini, the speech rate was still set to that of the Pepper robot.

6 Effect study

In the effect study, we investigate the effect of using different humanoid embodiments on the perceived intelligence, likability, usability, and animacy of reception robots.

6.1 Materials

For the interaction scenario of the effect study, the same robots are used as in the pilot study. The only difference is the usage of a tablet (see 6.1.1). Once the interaction has ended, the participant receives an online questionnaire.

6.1.1 Tablet

As mentioned in section 5.2, participants communicate with the robot by pronouncing keywords. The robot then performs the task. To add an extra modality to the interaction, the robot uses a screen. For example, the location of a room or person is told by the robot but also displayed on the screen. The Alpha Mini and the NAO robot do not have a built in tablet screen. By using the same tablet for all conditions, we also equalize the three conditions and therefore exclude the possibility of confusion, false expectations, or disappointment. In the pilot experiment, we used the screen of a laptop. In the effect study, we used a tablet screen.

6.2 Experimental Setup/approach

For the effect study we use the same experimental setup as in our pilot study. Figure 8 shows the updated interaction flow. We switched robot types every two and a half hours, or earlier if a robot had interacted with four participants consecutively. This was done to maintain a balanced population of participants across the different robots as much as possible.

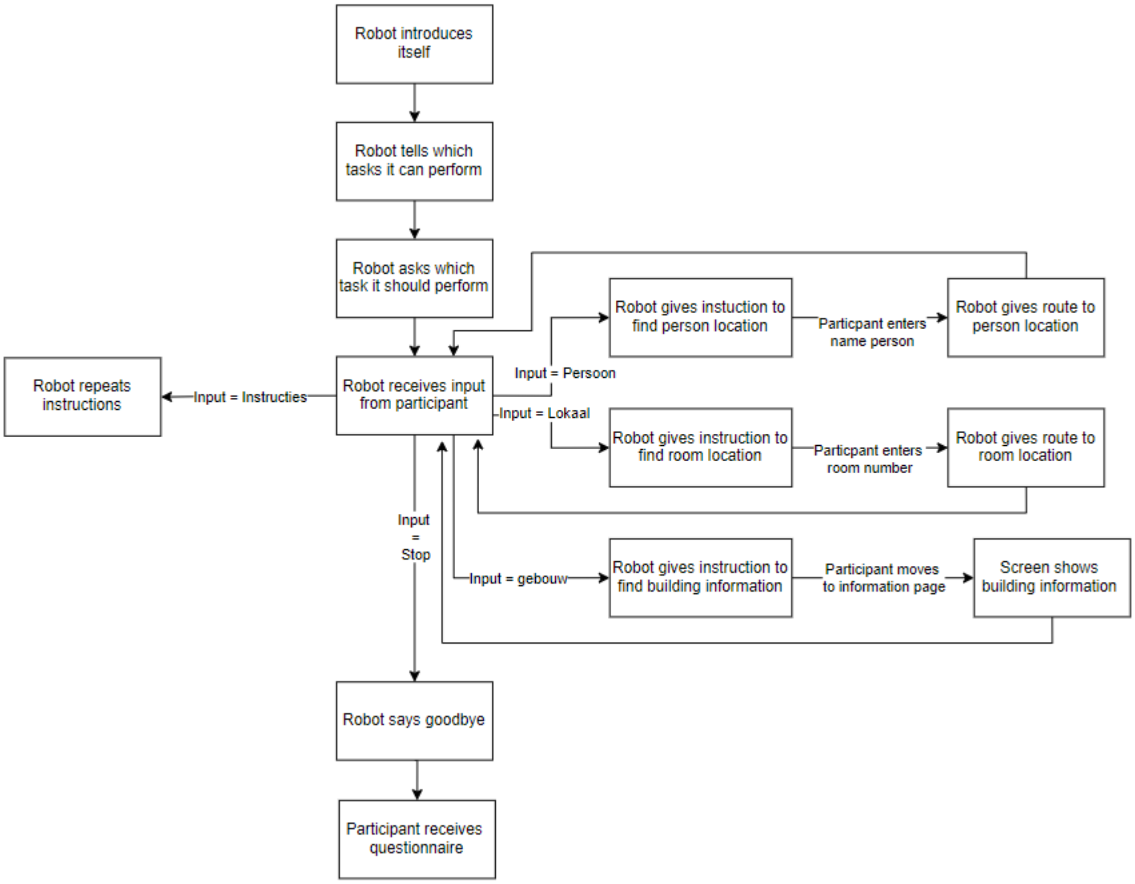
Robot introduces itself
Robot tells which tasks it can perform
Robot asks which task it should perform
Robot gives instuction to find person location
Particpant enters name person
Robot gives route to person location
Robot repeats instructions
Input = Instructies
Robot receives input from participant
Input = Persoon
Input = Lokaal
Robot gives instruction to find room location
Particpant enters room number
Robot gives route to room location
Input = Stop
Input = gebouw
Robot gives instruction to find building information
Participant moves to information page
Screen shows building information
Robot says goodbye
Participant receives questionnaire

Figure 8: Interaction flow effect study

For the effect study, we had 49 participants in total.
Condition 1, Alpha Mini, had 16 participants.
Condition 2, NAO, had 16 participants.
Condition 3, Pepper, had 17 participants.

Figure 9 and Figure 10 show what the population of the subjects looks like.

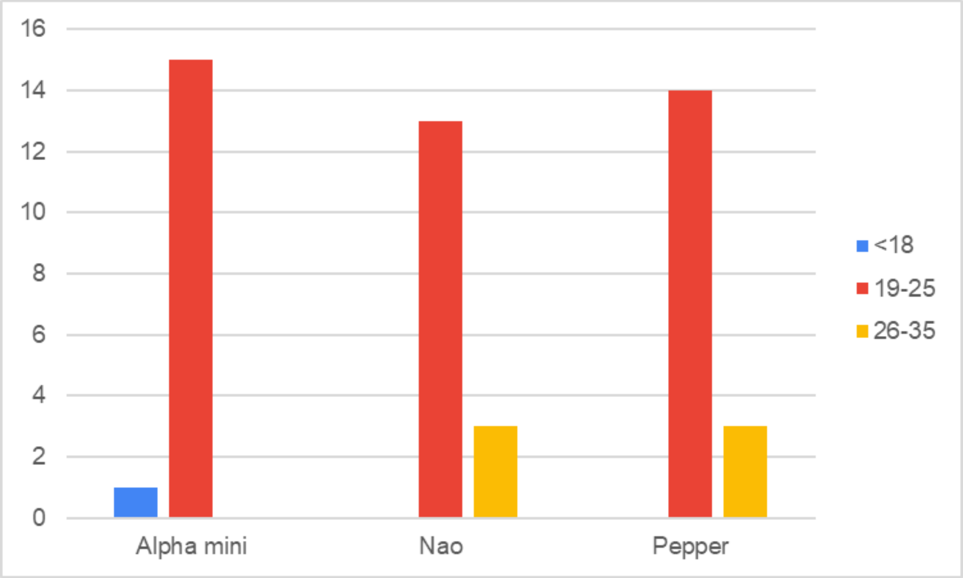

Figure 9: Age distribution effect study

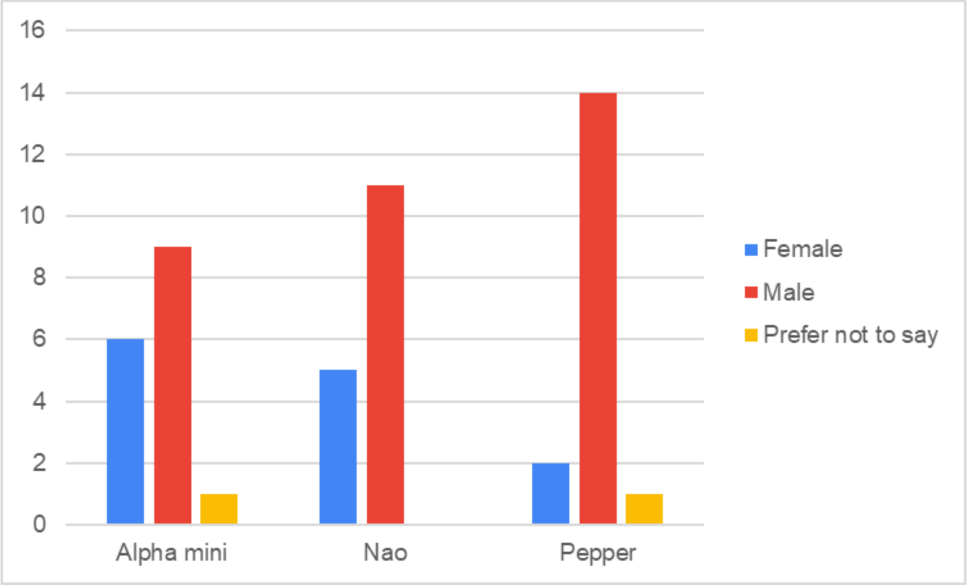

Figure 10: Gender distribution effect study

6.3 Measures

The questions for the questionnaire of the effect study were obtained through the Godspeed questionnaire [CBZ09]. This questionnaire is widely used in HRI research to investigate the users' perception of robots. Table 5 shows the Godspeed questions. Table 6 [RAGP13] shows the questions that correspond to the usability component. These questions are not from the Godspeed questionnaire, but from the System Usability Scale (SUS) [RAGP13]. The SUS questionnaire is a widely used tool for evaluating the usability of a variety of products and systems, including software, websites, mobile apps, and other interactive interfaces. In addition, we collected the age, gender, and familiarity with robots in general of the participants. Finally, participants are asked to characterize the robot in a few sentences. This is an open question to collect qualitative data about the experience.

Animacy						
Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive
Likeability						
Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice
Perceived Intelligence						
Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Table 5: Items for usability

Nr	Statement
1	I think that I would like to use this system frequently
2	I found the system unnecessarily complex
3	I thought the system was easy to use
4	I think that I would need the support of a technical person to be able to use this system
5	I Found the various functions in this system were well integrated
6	I thought there was too much inconsistency in this system
7	I would imagine that most people would learn to use this system very quickly
8	I found the system very cumbersome to use
9	I felt very confident using this system
10	I needed to learn a lot of things before I could get going with this system

Table 6: Items for usability

6.4 Results

6.4.1 Animacy

Table 7 shows the descriptive statistics for each of the three conditions, based on the average of the six animacy questions of the Godspeed questionnaire.

	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Alpha mini	16	2.5729	.63819	.15955	2.2328	2.9130	1.50	3.67
Nao	16	3.3021	.59073	.14768	2.9873	3.6169	2.17	4.33
Pepper	17	3.0392	.55129	.13371	2.7558	3.3227	2.00	4.17
Total	49	2.9728	.65471	.09353	2.7847	3.1608	1.50	4.33

Table 7: Descriptive statistics animacy

Table 8 shows the outcome of the one-way between groups ANOVA test. Based on the results of the one-way ANOVA, there is a statistically significant difference between group means (F(2,46) = 6.199, p = 0.04).

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.368	2	2.184	6.199	.004
Within Groups	16.206	46	.352		
Total	20.575	48			

Table 8: Anova Animacy

As there is a significant difference between the means of the three conditions, we investigate the significant difference between individual groups. Table 9 shows the comparison between individual groups. The only statistically significant difference in mean was between the Alpha Mini and the NAO (p = 0.03).

(I) Used robot	(J) Used robot	Mean Difference (I-J)	Std.Error	Sig.	Lower Bound	Upper Bound
Alpha mini	Nao	-.72917	.20986	.003	-1.2374	-.2209
	Pepper	-.46630	.20675	.073	-.9670	0.344
Nao	Alpha mini	.72917	.20986	.003	.2209	1.2374
	Pepper	.26287	.20675	.418	-.2378	.7636
Pepper	Alpha mini	.46630	.20675	.073	-.0344	.9670
	Nao	-.26287	.20675	.418	-.7636	.2378

Table 9: Comparison between groups animacy

The result of this category shows that the NAO robot has the highest average. The Pepper robot comes second, and the Alpha Mini has the lowest average. The fact that the Alpha Mini has the lowest average can be explained due to its slow and controlled movements of the arms and head, whereas the NAO and Pepper have faster movements. The difference in mean between the NAO and the Pepper can be explained due the fact that the NAO has legs and Pepper have not. This impacts the lifelikeness aspect.

6.4.2 Likeability

Table 10 shows the descriptive statistics for each of the three conditions, based on the average of the five likeability questions of the Godspeed questionnaire.

	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Alpha mini	16	3.6750	.67676	.16919	3.3144	4.0356	1.60	4.60
Nao	16	4.0625	.57373	.14343	3.7568	4.3682	3.00	4.80
Pepper	17	3.7647	.51592	.12513	3.4994	4.0300	3.00	4.80
Total	49	3.8327	.60187	.08598	3.6598	4.0055	1.60	4.80

Table 10: Descriptive statistics likeability

Talbe 11 shows the outcome of the one-way between groups ANOVA test. Based on the results of the one-way ANOVA, there is no statistically significant difference between group means (F(2,46) = 1.892, p = 0.162). Since there is no statistically significant difference between group means, we do not need to investigate the significant difference between individual groups.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.321	2	.661	1.892	.162
Within Groups	16.066	46	.349		
Total	17.388	48			

Table 11: ANOVA likeability

We see the same result as in the animacy category. The NAO has the highest average score. The Pepper again on the second place and the Alpha Mini with the lowest average. Overall the averages are relatively high which indicates that most people enjoyed the interaction regardless of which robot they were talking to.

6.4.3 Perceived intelligence

Table 12 shows the descriptive statistics for each of the three conditions, based on the average of the five perceived intelligence questions of the Godspeed questionnaire.

	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Alpha mini	16	3.3625	.50712	.12678	3.0923	3.6327	2.40	4.40
Nao	16	3.4875	.57023	.14256	3.1836	3.7914	2.20	4.40
Pepper	17	3.3059	.52018	.12616	3.0384	3.5733	2.40	4.00
Total	49	3.3837	.52732	.07533	3.2322	3.5351	2.20	4.40

Table 12: Descriptive statistics perceived intelligence

Table 13 shows the outcome of the one-way between groups ANOVA test. Based on the results of the one-way ANOVA, there is no statistically significant difference between group means (F(2,46) = 0.497, p = 0.611). Since there is no statistically significant difference between group means, we do not need to investigate the significant difference between individual groups.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.283	2	.141	.497	.611
Within Groups	13.064	46	.284		
Total	13.347	48			

Table 13: ANOVA perceived intelligence

Similar to the previous two categories of the Godspeed questionnaire, the NAO robot again has the highest average. This time the Alpha Mini is in second place, followed by the Pepper robot. The means of all conditions are close to each other, so all robots are estimated to be approximately equally intelligent. All robots were able to perform the same actions and had the same "knowledgebase". This explains why the differences between the conditions are not great. However, there are some small differences. Some participants were surprised at the tasks the Alpha Mini robot could perform. Four participants indicate this in the answer to the last open question. This surprise effect could be beneficial for the robot's score in this category.

6.4.4 Usability

Table 14 shows the descriptive statistics for each of the three conditions, based on the average of the ten usability questions of the SUS questionnaire.

	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Alpha Mini	16	3.7125	.34034	.08509	3.5311	3.8939	2.70	4.10
Nao	16	3.4813	.45785	.11446	3.273	3.7252	2.40	4.20
Pepper	17	3.40000	.43589	.10572	3.1759	3.6241	2.70	4.20
Total	49	3.5286	.42769	.06110	3.4057	3.6514	2.40	4.20

Table 14: Descriptive statistics usability

Table 15 shows the outcome of the one-way between groups ANOVA test. Based on the results of the one-way ANOVA, there is no statistically significant difference between group means (F(2,46) = 2.491, p = 0.094). Since there is no statistically significant difference between group means, we do not need to investigate the significant difference between individual groups.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.858	2	.429	2.491	.094
Within Groups	7.922	46	.172		
Total	8.780	48			

Table 15: ANOVA usability

This is the first time, the NAO robot does not have the highest mean in a category. This time the Alpha Mini has the highest mean, followed by the Nao and Pepper robots. However these differences are not significant.

6.4.5 Open question

Table 16, Table 17 and Table 18 show what the participants for the corresponding robot responded to the last open-ended question; How would you describe the robot? It also indicates the number of people who mentioned each description. Some participants mentioned multiple descriptions. That is why the total amount of descriptions is higher than the number of participants.

Description	Number of times said
Friendly	2
Tedious	5
Interactive	1
Cute	3
Apathetic	1
Funny	3
Small	3
Handy	3

Table 16: Alpha Mini description and frequency of mentions

Table 16 shows that almost one-third of the participants found the robot long-winded. Additionally, the Alpha Mini's small size is multiple times mentioned. The Alpha Mini is called funny and cute both three times. Finally, the Alpha Mini is also described as handy.

Description	Number of times said
Tedious	3
Robotic	2
Helpful	3
Talks too much	1
Easy to use	1
Nice experience	1
Handy	1
Clear	3
Mobile	2
Friendly	3

Table 17: NAO description and frequency of mentions

The Nao robot had received a slightly greater variety of descriptions. Table 17 shows that again tedious is mentioned a few times. Additionally, the NAO was called clear and friendly both three times. Finally, the NAO was also called helpful three times.

Description	Number of times said
Helpfull	1
Friendly	5
Handy	2
Humanlike	1
Knowledgeable	2
Nice experience	1
Static	1
Easy to use	5
Tedious	2
Annoying voice	1
Mechanic	2
Clear	2
Automated	1
Interesting	1
Inviting to talk to	1
Limited	2
Good speech detection	1
Nice arm/hand movements	1

Table 18: Pepper description and frequency of mentions

The Pepper robot had received the highest number of different descriptions. Table 18 shows that friendly and easy to use were mentioned most frequently. Furthermore, it is noteworthy that tedious was mentioned only twice.

7 Conclusion

The requirement study showed that the route to a person's office, the route to a specific room and practical information about the Snellius building were the three most requested tasks students ask.

The pilot study gave us insight in how we could improve our first version of the interaction scenario. Errors such as speech rate, volume, and choosing an option before the robot was done speaking were fixed and were no longer a problem.

In the effect study, we investigated the effect of different humanoid embodiments on the perceived intelligence, likability, usability, and animacy of reception robots. Our hypothesis was that we expect to see that people prefer interaction with a more humanoid robot, compared to a less humanoid robot, as measured by a higher rating on the four outcome variables (animacy, perceived intelligence, usability, and likeability).

The results indicate that a significant difference was found only for the aspect of animacy. Where the NAO had the highest perceived animacy and this difference was significant with the Alpha Mini. Since this was the only found significant effect , the proposed hypothesis is partially supported. No statistical difference was found for the aspects of perceived intelligence, usability, and likeability. Therefore the proposed hypothesis has been rejected in relation to these three aspects.

Although the experiment did not prove the hypothesis, it still brought useful feedback. Most participants were excited about using the robot. This is reflected in the high likeability score for all robots.

Thanks to the open-ended question, it emerged that some participants found the interaction long-winded. This was most prevalent with the Alpha Mini robot and least with the Pepper robot. Friendliness was mentioned the most with the Pepper robot and the least with the Alpha Mini.

8 Further research

Building upon the present experiment, there are opportunities for enhancements aimed at broadening the experiment's scope and refining the existing experimental setup.

8.1 Multilingualism

The robots used in this study were only able to recognize and speak Dutch. This meant that only Dutch students could participate. It happened relatively often that a potential participant was asked to participate in the research, it turned out that the participant could not speak Dutch. By ensuring that the robot understands and speak Dutch and English, it can assist more people. For the research, it might be interesting to see if there is a difference between Dutch and international students in terms of their preference for a particular humanoid embodiment.

8.2 Improved speech recognition

During the pilot study and the effect study, there were instances where the robot incorrectly perceived a keyword. Subsequently, the robot started executing the corresponding task. There were also a few occasions where nothing had been said, but the robot detected background noise and interpreted it as one of the keywords. Finally, on two occasions with the NAO robot, it correctly understood the keyword but then, when starting to execute the task, it heard the same keyword again. As a result, the robot explained the task twice, partially overlapping. All three of the issues were detrimental to the research and should be improved in a potential actual receptionist robot. Using different speech-to-text software may potentially address this problem. Emerging technologies such as ChatGPT and Google Dialogflow can provide a solution for this.

8.3 Improved instruction delivery

What frequently emerged from the participants' responses to the open question was that waiting during the instructions was often perceived as unpleasant. It took quite a while for the robot to finish delivering all the instructions. Perhaps an interesting follow-up study could focus on determining the most preferred way for participants to receive instructions on how to use the robot.

8.4 Different time period

The effect study was conducted during the exam period, which resulted in relatively quiet conditions in the Snellius building. Typically, there is more foot traffic in and out of the building, creating more background noise for the robot. Additionally, most participants were either heading to their exams or had just finished one. The exam experience or pre-exam stress might play role in their evaluation of the robots. To ensure complete certainty, it would be advisable to replicate this experiment during a non-exam period.

References

[BL18] F. Bazzano and F. Lamberti. Human-robot interfaces for interactive receptionist systems and wayfinding applications. *Robotics*, 7(3):56–56, 2018.

[BTP14] Y. jung B. Tay and T. Park. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, 38:75–84, 2014.

[CBZ09] E. Croft C. Bartneck, D. Kulić and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[CDK02] J. Forlizzi C.F. DiSalvo, F.Gemperle and S. Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 321–326, 2002.

[DKG20] O. Wallberg A. Pereira I. Leite D. Kontogiorgos, S. van Waveren and J. Gustafson. Embodiment effects in interactions with failing robots. *In Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[Fin12] J. Fink. Anthropomorphism and human likeness in the design of robots and human-robot interaction. *International Conference on Social Robotics*, pages 199–208, 2012.

[LC19] V. Wade B. R. Cowan N. Pantidi O. Cooney L. Clark, C. Munteanu. What makes a good conversation? *in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems–CHI'19*, pages 1–12, 2019.

[MMK12] K. F. MacDorman M. Mori and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012.

[RAGP13] P.Kortum R. A. Grier, A. Bangor and S. C. Peres. The system usability scale: Beyond standard usability testing. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):187–191, 2013.

[RAPK15] E. A. Konijn R. A. Paauw, J. F. Hoorn and D. V. Keyson. Designing robot embodiments for social interaction: affordances topple realism and aesthetics. *International Journal of Social Robotics*, 7(5):697–708, 2015.

[TN21] K. Sakai T. Iio M. Chiba T. Asami H. Ishiguro T. Nishio, Y. Yoshikawa. The effects of physically embodied multiple conversation robots on the elderly. *Frontiers in Robotics*, 8:61, 2021.

[vdP20] L. P. van der Plas. The influence of behavioral richness on peoples perception of reception robots. *Bachelor thesis*, 2020.

A Usage of chatGPT

ChatGPT has been used a few times as a translator from Dutch to English for writing this thesis. This includes the translation of some single words and the translation of some blocks of multiple (1-3) sentences from Dutch to English. One or multiple sentences were translated to be sure the translation fits the context. The content remains authentic, since the original Dutch text was created by the author of this thesis and not by ChatGPT or any other form of chatbot. The usage of ChatGPT can be compared by the usage of Google Translate.