Spoonie: A food Recommendation Agent

Jeroen Hofland, Joost Jansen, Jochem van Lith, Zeger Mouw and Jim Vos
EEMCS, Delft University of Technology, The Netherlands
CS4270 Conversational Agents

1 Introduction

Whilst technology keeps getting a bigger role in our lives, human-computer interaction (HCI) still often follows rigid forms of communication. In the research field of Conversational Agents (CA) the goal is to make virtual agents that can have human-like conversations with other humans. Further research in CA's could allow other technology to be implemented more easily in many human-environments, like consultancy or tutoring [8].

Creating an agent that realistically simulates a conversational partner depends on multiple aspects like gestures, expressions, perception of the conversational partner and the agent's memory. In this research the focus lays on the memory aspect of CA. Although computers have a very persistent and well organised memory for computational tasks, its models are not well suited for human-conversations, as storing whole conversations is inefficient and unfeasible over many conversations [12]. As humans are able to interpret conversations they can understand what is relevant to remember and what to forget, creating memory models that define our conversational flow. Memory can be divided in different forms as stated by Sun [19]: Implicit vs. explicit memory, action-centered vs. non-action-centered memory, and episodic vs. semantic. Implicit memory is thoughts that are stored unconsciously, while explicit memory is retrievable. The episodic memory contains the knowledge of events and semantic memory contains facts and general knowledge. For our research we will take a deeper dive into what is important to remember for an explicit action-centered episodic CA to become a more natural conversational partner.

Specifically for this research, we designed a food recommendation agent called "Spoonie" based on the Furhat SDK [2]. Spoonie recommends meals based on the user's preferences which it learns through having conversations. Its memory is divided into short- and long-term[17] and one-shot modules [12]. We are interested in how the long-term and one-shot memory modules affect the user experience. By performing experiments with different memory settings where we measure the user experience, we wish to answer the following research question:

> ***RQ:*** *In what way does long-term and one-shot memory affect the recipe recommendation experience after multiple encounters?*

2 Background

2.1 Memory architecture

Over the past decade, multiple studies have been conducted on different types of memory architectures for CA's. One of the first studies that use long-term memory for social interaction with a robot is Robovie [11]. They studied the interaction of a robot with elementary school children for over a period of two weeks. After that, multiple studies began discussing different memory architectures to see the effects of long-term memory. The study by Elvir et al.[9] proposed a unified episodic architecture with a pipeline approach. They demonstrated that a CA with memory could answer questions more successfully than a CA without memory. Another study created a tutoring robot by leveraging the similarities between episodic memories with a Belief-Desire-Intention (BDI) architecture and a Hierarchical Task Network (HTN)[21]. Richards and Bransky researched the naturalness of forgetting information over a longer period with their memory architecture[16]. Over the three sessions that were held in this study, recall was found to boost satisfaction, belief, and trust. Sánches et al.[17] uses a framework by dividing episodic memory into long-term and short-term memory. Here, short-term memory is only used for within that conversation, and long-term memory is kept over multiple sessions. Kim et al. [12] introduces another type of memory: the one-shot memory. Whereas long-term memory fades away over time. The one-shot memory contains facts that should not be forgotten. They used this type of memory to build an agent who talks about movies. Our study tries to combine these last two works' memory architectures and test their effectiveness in a food recommendation environment.

2.2 Food recommendation systems

Perhaps one of the most similar works, is a study from IBM about a food recommendation system called Foodie [4]. Foodie's purpose is to reduce food waste by optimising grocery usage and to aid families in changing their eating habits with recipe ideas that take into consideration personal context, such as allergies and dietary goals. While Foodie's attributes and goals are very similar to our goals, the study does not research the effects of long term memory and multiple encounters on the likeability of the agent. Another study by Barko-Sherif et al.[6] also investigates a certain food recom-

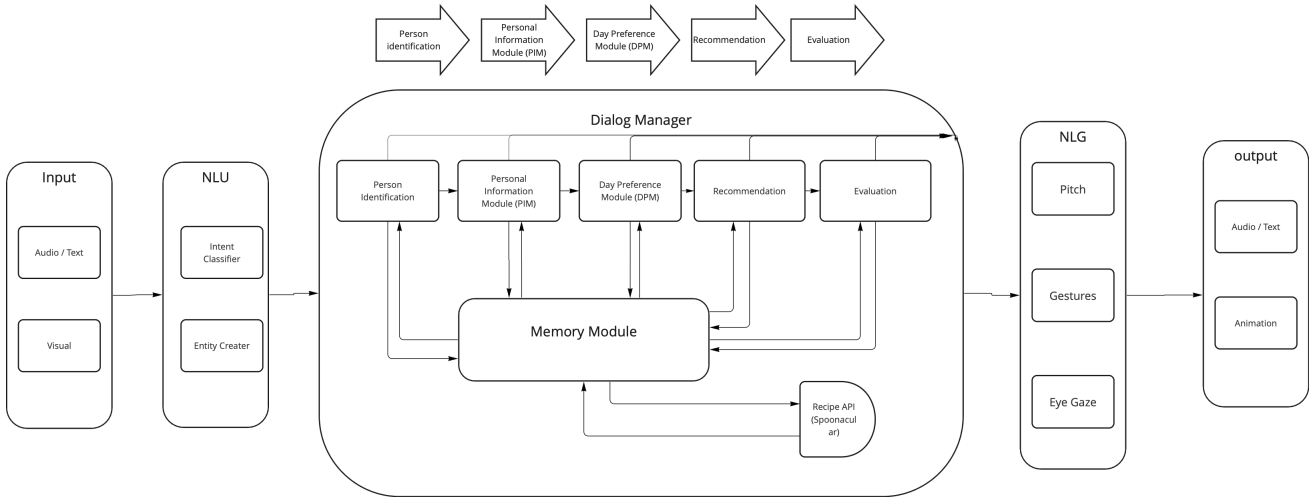


Figure 1: Dialog system

mendation system. Which uses a Wizard-of-Oz style implementation of a conversational agent that recommends food recipes. Their research demonstrates that conversations with a recipe recommendation system may be varied, differing significantly by task, user, and, to some extent, interaction style. However this study also mainly focuses on a one-time interaction with the user and was mostly focused on the conversation framework instead of the memory. As a result, there is still a gap in the research on the impact of long-term memory on food recommendation CA's.

2.3 Questionnaires

Most studies mentioned above have not been evaluating their CA coherently with each other. As a result, a number of studies have underlined the importance of standardised measurement tools. One of the most frequently used measurements with the perception of CA's is the Godspeed questionnaire [20]. The Godspeed questionnaire [7] aims to evaluate the agent coherently by dividing it in five key concepts: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. The ASA [10] is another questionnaire that captures more than 80% of the questions identified in empirical studies presented in the IVA (Intelligent Virtual Agent) conference 2013-2018. A well-known by origin Japanese questionnaire is the Negative Attitudes towards Robots Scale (NARS) [3]. NARS focuses primarily on the negative perception and likeability of robots. The creators of the Multi-Dimensional Robot Attitude Scale[14] proposed that previous scales, such as the NARS, focused primarily on negative attitudes toward robots, underlining the need for a measure that encompasses a larger range of attitudinal aspects. In our work we'll use the Godspeed questionnaire to make our study easily comparable.

3 Agent Design

Now that we have familiarized with the related work, we will dive into the design of our conversational agent. It is important that the agent has two types of memory settings to be able to perform a comparative study which answers the research question. The agent should also make use of perception. In this section we explain how we designed and implemented both of these requirements, but we begin with describing the architecture of our dialog system.

3.1 Dialog System

Figure 1 shows the dialog system of our agent. The input (audio, text or visual) is passed to the natural language understanding unit. This unit classifies the users intent and creates entities from our defined classes such that the information is represented in a meaningful way to the program. The entities will then be retrieved by the dialog manager, which will first identify the user. Next, the personal information module (PIM) asks new users for diets and allergies. Afterwards, the day preference module (DPM) will explore the user's current wishes through conversation. With all this information along with ratings of previously recommended meals, a new recommendation is proposed to the user. It has been shown that a social explanation can improve the perceived quality of both the system and the interaction [15]. Therefore, our recommendation module also generates an elaborate explanation to improve the quality of the recommendation. When a user rejects the recommendation, the meal receives a negative rating and a new recommendation is made. This is repeated until the user accepts a recommendation. When an existing user returns, they are asked whether they liked their latest recommendation. The reaction is converted into a rating and stored for future recommendations.

3.2 Memory

The dialog manager utilises the memory module in our conversational system. We decided on a temporal perspective of the memory module, because our data can be split up into independent groups based on the period that it should be remembered. As mentioned before, we combine the memory architectures of Sánches et al. [17] and Kim et al. [12]. This allows for three distinct time-based components: the short-term, long-term and one-shot memory modules. The one-shot

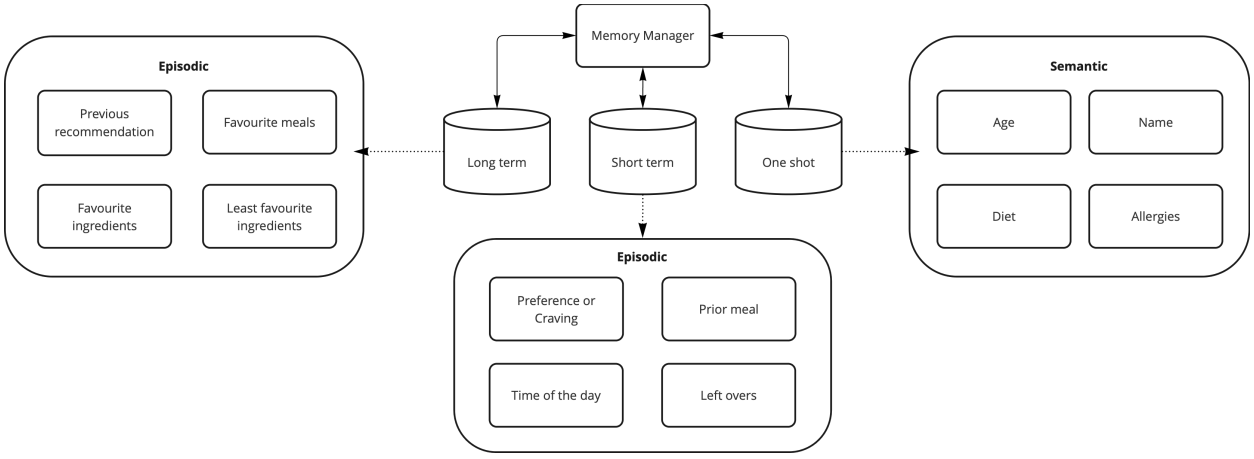


Figure 2: Memory model

memory encapsulates plain facts, so it has a semantic setup. The other two modules are set up in an episodic way to be able to store personal experiences. The architecture is shown in Figure 2.

All of the permanent data, such as a user's name and allergies, are stored in the one-shot memory. This allows the agent to have a natural conversation and show interest by at least remembering their most basic background. In the long-term memory, we store the user's diets and ratings of ingredients, cuisines and specific meals. This module uses a decay function which deletes data after a certain amount of time. We picked the decay time to be 2 months, however the optimal time is very subjective, because it depends on a user's opinion. This allows for exploration of new recipes and accounts for changes in a user's opinion. The short-term memory only captures information which is relevant during one session. Examples of this are left overs and the currently preferred meal type. This ensures that the agent will ask for this information every session, showing interest in the day-to-day activity of a user.

For the purpose of this study, we are able to switch off the long-term and one-shot memory modules. In this setting, all the data is stored in the short-term memory. As a consequence, the agent will treat any user as a new user and request all the general information again in each session. In section 4, we explain how we have used the two settings in the controlled experiment.

3.3 Perception

The robot not only identifies the users intents, but is also capable of sensing the user and its emotions via visuals and text for a more human-like experience. We decided to use visual perception as it is multi-functional and there is always a visual feed which made it a better choice than audio. The robot makes use of the facial recognition library openCV [2] which is capable of identifying the users face position and emotion via a camera feed. The face position parameters are used to keep the agent looking towards the person and the emotion detection is used to mimic the users smiles, in order to generate a feeling of interest.

As the NLU of the Furhat SDK [1] was limited for understanding the users emotion we decided to also use the TweetEval model from Barbieri et al [5]. This model is able to reliably detect the sentiment of text, which gives us another channel for emotion detection. In order to fuse the channels together, both of their output gets translated into either positive or negative emotion. The emotion detected via the text interpretation does weigh twice as heavy as it deemed to be more accurate in detecting the users emotion compared to facial expressions, as expressions are less reliable due to human variance.

The emotional value that comes out of this calculation is used to scale the users preferences. The ratings of ingredients, cuisines and meals are updated accordingly. These ratings our stored in the long-term memory module. The decay function, described in section 3.2, ensures that ratings can be forgotten over time. This allows a user to change their opinion and makes place for up-to-date ratings.

4 Method

We performed a comparative study between two different agents. The study was done using a between-subject design. This ensures that the users have only seen one agent and that they will not have preliminary information from earlier tested agents. The differences between the tested agents are in terms of memory. Table 1 displays the features of each agent.

		Agent 1	Agent 2
Memory	Short-term	True	True
	Long-term	True	False
	One-shot	True	False

Table 1: Features of the both agents

The memory architecture of agent 1 contains the one-shot, short-term and long-term modules. While agent 2 only has short-term memory. Participants were asked to fill out a questionnaire about their experience with the agent. By analysing the answers to the questionnaire, we wish to answer our research question.

In this section, we describe the methodology behind our experiment in detail. This includes the organisation of participants, a description of the used measures, the procedure and how we dealt with the ethical aspects.

4.1 Participants

For the experiment we gathered 30 participants. We decided to recruit students between the age of 18 up to 25. We evenly divided people with prior knowledge about conversational agents as well as different genders among the two groups. This division was meant to minimize the influence of the participants' bias on our results. Other than that we decided not to ask nor save any personal information, keeping in mind the privacy of the participants.

4.2 Measures

In this study, we want to measure whether the user experience differs between the two agents. User experience is something that we can measure by explicitly asking the users how they experienced the conversation with the agent. A study by Bartneck et al. [7] shows that we have 5 key concepts in Human Robot Interaction (HRI) [18].

- Anthropomorphism
- Animacy
- Likeability
- Perceived intelligence
- Perceived safety

We have used the godspeed questionnaire [7] as the base for our own questionnaire to ask the user what they thought of the experience, using the concepts above. The godspeed questionnaire fits well to this study since human perception of the agent is the main research. We adjusted the godspeed questionnaire terms to questions to clarify the questionnaire for the user. We also inverted a few questions so that the user would keep his focus. The values of the inverse questions were also reversed in the calculation.

4.3 Procedure

The procedure is as simple as possible and focused on isolation of the participants to minimize prevent bias. For an individual evaluation the following steps are conducted.

1. The participant is guided to a quiet place where he/she can interact with the agent without any distractions.
2. The participant is informed about the risks and has to sign the informed consent form described further in the next subsection.
3. The participant is informed that he/she is going to interact with a recipe recommendation agent.
4. When the participant is ready, the researcher launches the first session with either agent 1 or 2 depending on which group the participant belongs to.
5. Once a recipe recommendation is accepted by the participant, the session is over.
6. The researcher restarts the same agent for the next session to simulate a new day.
7. Steps 5 and 6 are repeated for the third and final session.
8. The participant is asked to fill in the questionnaire.
9. The researcher checks whether all steps were done correctly. When this is not the case, the data is not taken into consideration.

4.4 Ethics

In order to fulfil the ethics standards at the TU Delft we have drawn up a Consent form and HREC Risk Analysis which can be found in the appendices. We ask for the age, gender and profession to divide the participants in two groups. This information is not stored. Only the name and the signature on the consent form are saved.

5 Results

5.1 Analysis

We will statistically analyse our data to determine how the different memory settings affect the recommendation experience. The null hypothesis is as follows: there is no difference in user experience between people interacting with an agent that only has short-term memory versus an agent that has short-term, long-term, and one-shot memory.

We will use the independent T-test to determine if the answers to the questionnaires are significant for the different cases. The assumptions of the T-tests are that the data is independent, that the data is normally distributed and that the variances of both groups are homogeneous [13][22]. We should perform pre-tests to verify if these assumptions are met. We know that the data is independent, because we use a between user experimental design.

To test for normality we use the Shapiro-Wilk test. If this test returns a value above 0.05, then our data passed the test. If the data is not normally distributed, then we should use Wilcoxon rank test instead of the T-test [13]. Finally, we will use the F-test to test whether the variances are similar [22]. This test should also return a value larger than 0.05 to pass. If the test fails, then we should use the Welch's T-test.

We can perform the independent T-test if all assumptions are met. The test will return a value that states how significant the means differ between the two groups. If that value is below 0.05, then we consider it significant and we can reject the null hypothesis [13]. This means that users do in fact experience a difference between the agents and that the long-term and one-shot memory (either positively or negatively) influence the recommendation experience. If the value is above 0.05, however, this influence could still be present but our results simply are not reliable enough to draw this conclusion.

Besides taking the means of the entire dataset, we are also able to compare the results on the key concepts in HRI mentioned in section 4.2. The procedure will be the same, but when computing the means and standard deviations one should only consider the data points that belong to the specified concept. In this way, we can get a better understanding of what the agents are particularly good at.

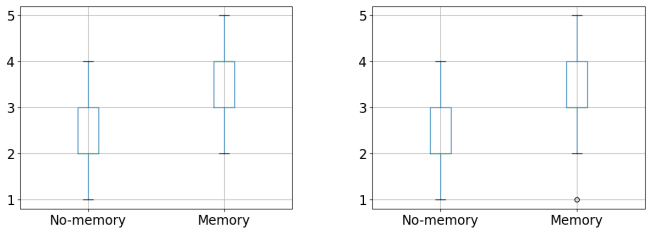
5.2 Results

The experiments show that in 19 of the 22 cases the Shapiro-Wilk test fails, this shows that in most of the investigated

properties the answers are not normally distributed and thus the Wilcoxon test is used. In the 3 cases where the data is normally distributed, the variance or F-test does pass which implies that the variances of those normally distributed answers do not differ. Due to this property, Welch's t-test is kept out of discussion for the rest of the results. In 2 of those 3 cases the t-test rejects the null hypothesis thus implying a significant difference in the two tested groups. This is the case for the properties:

- $Natural \in Anthropomorphism$ (figure 3a)
- $Lifelike \in Animacy$ (figure 3b)

The third parameter Foolish - Sensible is close to but not significant enough (table 2d).



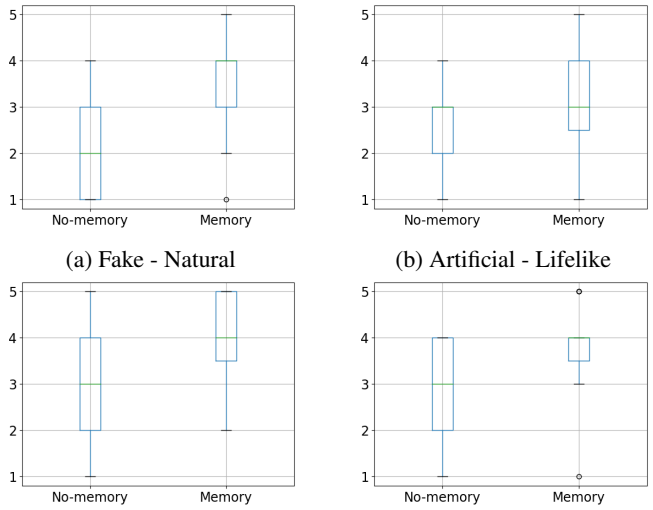
(a) Fake - Natural (b) Artificial - Lifelike

Figure 3: Conversational agent properties that are normally distributed, have equivalent variance, and reject the null-hypothesis using the T-test. Here the scale from 1 to 5 represents the match relative to the two keywords used. Here **(a)** is a property anthropomorphism, and **(b)** of animacy.

The other 19 parameters are tested using the Wilcoxon signed rank test which result in a rejection of the null hypothesis in 4 cases. This is the case for the properties:

- $Humanlike \in Anthropomorphism$ (figure 4a)
- $Lifelike \in Anthropomorphism$ (figure 4b)
- $Nice \in Likability$ (figure 4c)
- $Competent \in PerceivedIntelligence$ (figure 4d)

The other properties tested do not have a significant difference between the two groups tested and thus can not be used to determine the influence of memory with relation to performance (table 2a-2d). For reference and further studies the box-plots of these can be found in the appendix in figure 5-8. Next to agent's properties, the feelings before and after of the participants are also filled in. These do not show a significant difference between the tested groups either (table 2e). The box-plots for these can be found in the appendix figure 9.



(a) Fake - Natural (b) Artificial - Lifelike

(c) Awful - Nice (d) Incompetent - Competent

Figure 4: Conversational agent properties that do not pass normality test and are tested using the Wilcoxon signed rank. Here the scale from 1 to 5 represents the match relative to the two keywords used. Here **(a)** and **(b)** are properties of anthropomorphism, **(c)** of likability, and **(d)** of perceived intelligence.

References

[1] Furhat. https://furhatrobotics.com/.

[2] Opencv. https://opencv.org/opencv-face-recognition.

[3] Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3):437–454, 2006.

[4] Prashanti Angara, Miguel Jiménez, Kirti Agarwal, Harshit Jain, Roshni Jain, Ulrike Stege, Sudhakar Ganti, Hausi A Müller, and Joanna W Ng. Foodie fooderson a conversational agent for the smart kitchen. In *CASCON*, pages 247–253, 2017.

[5] Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421*, 2020.

[6] Sabrina Barko-Sherif, David Elsweiler, and Morgan Harvey. Conversational agents for recipe recommendation. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, page 73–82, New York, NY, USA, 2020. Association for Computing Machinery.

[7] Christoph Bartneck, Elizabeth Croft, and Dana Kulic. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.

[8] Keeley Crockett, James O'Shea, and Zuhair Bandar. Goal orientated conversational agents: Applications to benefit society. volume 6682, pages 16–25, 06 2011.

[9] Miguel Elvir, Avelino J. Gonzalez, Christopher Walls, and Bryan Wilder. Remembering a conversation - a conversational memory architecture for embodied conversational agents. *Journal of Intelligent Systems*, 26:1–21, 1 2017.

[10] Siska Fitrianie, Merijn Bruijnes, Fengxiang Li, Amal Abdulrahman, and Willem-Paul Brinkman. The artificial-social-agent questionnaire: Establishing the long and short questionnaire versions. In *Proceedings of the 22nd ACM International Conference on Intelligent*

Virtual Agents, IVA '22, New York, NY, USA, 2022. Association for Computing Machinery.

[11] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, 19(1-2):61–84, 2004.

[12] Yonghee Kim, Jeesoo Bang, Junhwi Choi, Seonghan Ryu, Sangjun Koo, and Gary Geunbae Lee. Acquisition and use of long-term memory for personalized dialog systems. In *International workshop on multimodal analyses enabling artificial agents in human-machine interaction*, pages 78–87. Springer, 2014.

[13] Kim Tae Kyun. T test as a parametric statistic. *kja*, 68(6):540–546, 2015.

[14] Takumi Ninomiya, Akihito Fujita, Daisuke Suzuki, and Hiroyuki Umemuro. Development of the multidimensional robot attitude scale: Constructs of people's attitudes towards domestic robots. In *ICSR*, 2015.

[15] Florian Pecune. A model of social explanations for a conversational movie recommendation system. *HAI '19: Proceedings of the 7th International Conference on Human-Agent Interaction*, 2019.

[16] Deborah Richards and Karla Bransky. Forgetmenot: What and how users expect intelligent virtual agents to recall and forget personal conversational content. *International Journal of Human-Computer Studies*, 72(5):460–476, 2014.

[17] María-Loreto Sánchez, Mauricio Correa, Luz Martínez, and Javier Ruiz-del Solar. An episodic long-term memory for robots: The bender case. In Luis Almeida, Jianmin Ji, Gerald Steinbauer, and Sean Luke, editors, *RoboCup 2015: Robot World Cup XIX*, pages 264–275, Cham, 2015. Springer International Publishing.

[18] Katrin Schulze and Heidi Krömker. A framework to measure user experience of interactive online products. In *Proceedings of the 7th international conference on methods and techniques in behavioral research*, pages 1–5, 2010.

[19] Ron Sun. Memory systems within a cognitive architecture. *New Ideas in Psychology*, 30(2):227–240, 2012.

[20] Astrid Weiss and Christoph Bartneck. Meta analysis of the usage of the godspeed questionnaire series. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 381–388, 2015.

[21] Zerrin Yumak and Nadia Thalmann. Building long-term relationships with virtual and robotic characters: The role of remembering. *The Visual Computer*, 28:87–97, 01 2012.

[22] Donald W Zimmerman and Bruno D Zumbo. Parametric alternatives to the student t test under violation of normality and homogeneity of variance. *Perceptual and motor skills*, 74(3):835–844, 1992.

	SW	F	T	W	P-value
Natural	+	+	+		**0.0038**
Humanlike	-			+	**0.0149**
Conscious	-			-	0.4904
Lifelike	-			+	**0.0462**
Moving Elegantly	-			-	0.6718

(a) Anthropomorphism

	SW	F	T	W	P-value
Alive	-			-	0.2505
Lively	-			-	0.4297
Organic	-			-	0.0559
Lifelike	+	+	+		**0.0214**
Interactive	-			-	0.59
Responsive	-			-	0.0558

(b) Animacy

	SW	F	T	W	P-value
Like	-			-	0.2337
Friendly	-			-	0.3049
Kind	-			-	0.5652
Pleasant	-			-	0.1136
Nice	-			+	**0.0186**

(c) Likeability

	SW	F	T	W	P-value
Competent	-			+	**0.0295**
Knowledgeable	-			-	0.1542
Responsible	-			-	0.1191
Intelligent	-			-	0.2149
Sensible	+	+	-		0.0658

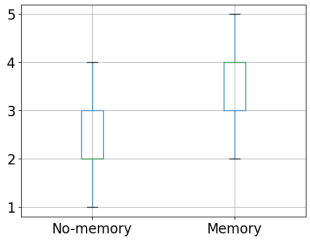
(d) Perceived Intelligence

	SW	T	W	P-value
Relaxed	-		-	0.5839
Calm	-		-	0.3402
Surprised	-		-	0.1628

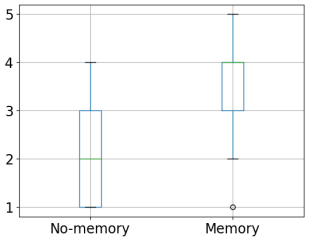
(e) Perceived Safety difference between start and end

Table 2: Godspeed questionnaire categories with their tested properties. A minus (-) represents a failed test, (+) represents a succeeded test, ( ) represents the fact that a test was not executed, and a **bold** value means the test rejects the null hypothesis. Here **SW** represents the Shapiro-Wilk test, **F** the F-test, **T** the T-test, **W** the Wilcoxon test, and **P-value** the value on which the significance is determined.

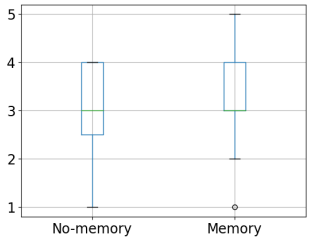
A Appendix



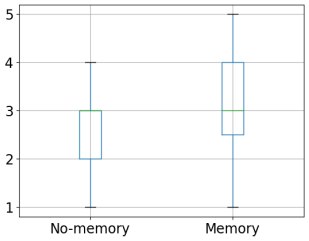
(a) Fake - Natural



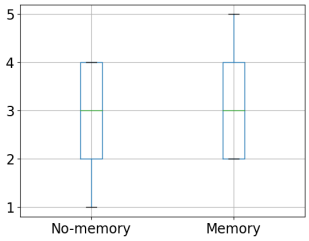
(b) Machinelike - Humanlike



(c) Unconscious - Conscious

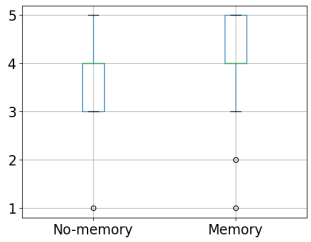


(d) Artificial - Lifelike

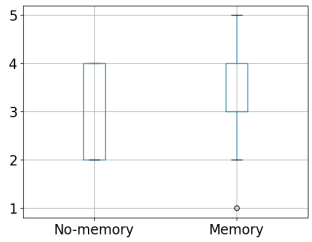


(e) Moving rigidly - elegantly

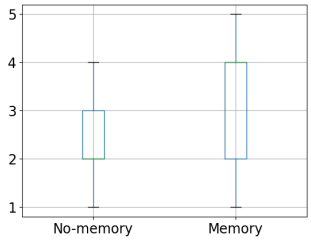
Figure 5: Anthropomorphism



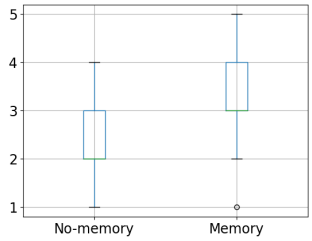
(a) Dead - Alive



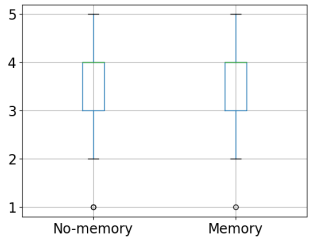
(b) Stagnant - Lively



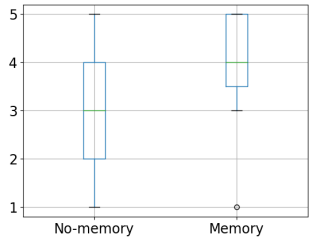
(c) Mechanical - Organic



(d) Artificial - Lifelike

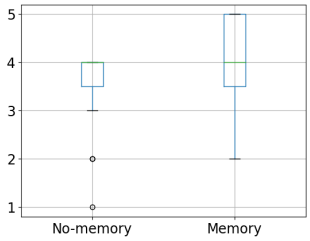


(e) Inert - Interactive

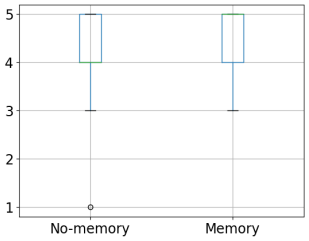


(f) Apathetic - Responsive

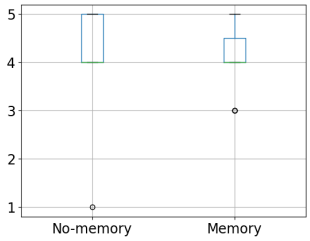
Figure 6: Animacy



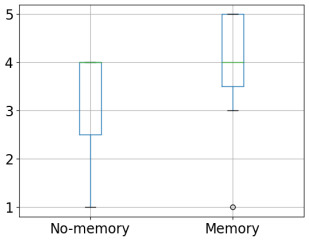
(a) Dislike - Like



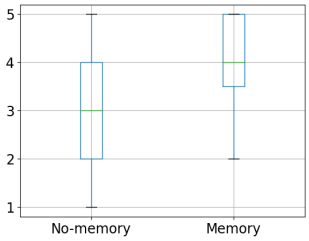
(b) Unfriendly - Friendly



(c) Unkind - Kind

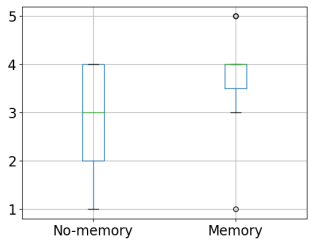


(d) Unpleasant - Pleasant

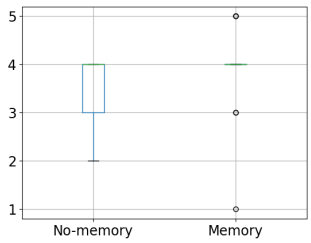


(e) Awful - Nice

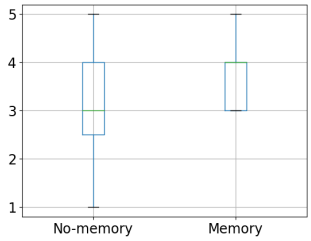
Figure 7: Likeability



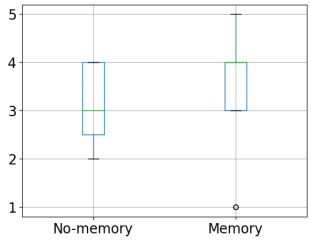
(a) Incompetent - Competent



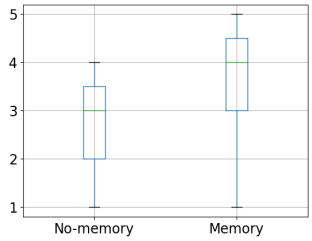
(b) Ignorant - Knowledgeable



(c) Irresponsible - Responsible

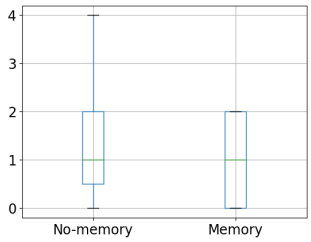


(d) Unintelligent - Intelligent

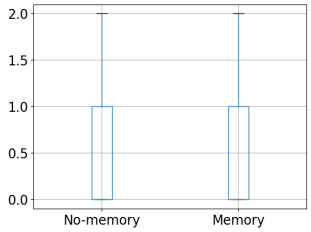


(e) Foolish - Sensible

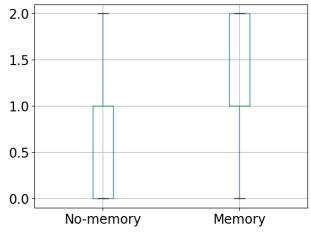
Figure 8: Perceived Intelligence



(a) Anxious - Relaxed



(b) Agitated - Calm



(c) Quiescent - Surprised

Figure 9: Perceived safety

Consent Form for Conversing with Recipe Recommendation Agent

This study will research how the properties of a conversational agent influence the experience of the participant with this agent. The participant will talk with the agent in 3 sessions. During these sessions, your information is saved to test how well the agent can use your personal information to make a better recommendation and conversation. After the 3 sessions, this data is deleted, and a questionnaire is presented to you. This data received from the questionnaire is anonymized so none of your personal information is saved.

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated ____/____/____ or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves a survey questionnaire completed by me (the participant) at the end of the experiments to evaluate the performance of the robot.	○	○
Use of the information in the study		
I understand that information I provide will be used to determine what effects certain traits of conversational agents have on the experience I have. These results will only be used in the course CS4270 Conversational Agents and will not be made public.	○	○
I understand that personal information collected about me that can identify me, such as age, my name, and my facial dimensions (for face identification), will not be shared beyond the study team.	○	○
Future use and reuse of the information by others		
I give permission for the anonymised results of the survey that I provide to be used in the report for the Conversational Agent course so it can be used for future research and learning.	○	○
I understand that the results will be anonymised by removing the participants name and facial identification properties from the data but that the age of the participant might be used to explore the possible difference in outcome based on the age.	○	○
I understand that future research only includes the course CS4270 Conversational Agent and will not be distributed or published and that after the ending of this course all the data will be deleted.	○	○

_____ _____ _____

Name of participant Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.	○	○

_____ _____ _____

Researcher name Signature Date

Study contact details for further information: *Jeroen Hofland, 0649610808, j.l.hofland@student.tudelft.nl*

Delft University of Technology
ETHICS REVIEW CHECKLIST FOR HUMAN RESEARCH
(Version 15.11.2021)

IMPORTANT NOTES ON PREPARING THIS CHECKLIST

1. An HREC application should be submitted for every research study that involves human participants (as "Research Subjects") carried out by TU Delft researchers
2. Your HREC application should be submitted and approved **before** potential participants are approached to take part in your study
3. All submissions from Master's Students for their research thesis need approval from the relevant Responsible Researcher
4. The Responsible Researcher must indicate their approval of the completeness and quality of the submission by signing and dating this form OR by providing approval to the corresponding researcher via email (included as a PDF with the full HREC submission)
5. There are various aspects of human research compliance which fall outside of the remit of the HREC, but which must be in place to obtain HREC approval. These often require input from internal or external experts such as Faculty Data Stewards, Faculty HSE advisors, the TU Delft Privacy Team or external Medical research partners.
6. You can find more guidance on completing your HREC application (including tips for completing this checklist) here
7. Please note that incomplete submissions (whether in terms of documentation or the information provided therein) will be returned for completion **prior to any assessment**
8. If you have any feedback on any aspect of the HREC approval tools and/or process you can leave your comments here

I. Applicant Information

PROJECT TITLE:	**Conversation Agent Food Recommendation**
Research period: *Over what period of time will this specific part of the research take place*	**Q1 of TU Delft**
Faculty:	**Electrotechniek, wiskunde en informatica**
Department:	**Computer Science**
Type of the research project: *(Bachelor's, Master's, DreamTeam, PhD, PostDoc, Senior Researcher, Organisational etc.)*	**Master course**
Funder of research: *(EU, NWO, TUD, other – in which case please elaborate)*	**-**
Name of Corresponding Researcher: *(If different from the Responsible Researcher)*	**-**
E-mail Corresponding Researcher: *(If different from the Responsible Researcher)*	**-**
Position of Corresponding Researcher: *(Masters, DreamTeam, PhD, PostDoc, Assistant/ Associate/ Full Professor)*	**-**
Name of Responsible Researcher: ***Note:** all student work must have a named Responsible Researcher to approve, sign and submit this application*	**Dr. C.R.M.M. Oertel genannt Bierbach**
E-mail of Responsible Researcher: *Please ensure that an institutional email address (**no Gmail, Yahoo, etc.**) is used for all project documentation/ communications including Informed Consent materials*	**C.R.M.M.Oertel@tudelft.nl**
Position of Responsible Researcher : *(PhD, PostDoc, Associate/ Assistant/ Full Professor)*	**PhD**

II. Research Overview

NOTE:** You can find more guidance on completing your HREC application (including tips for completing this checklist) **here

a) Please summarise your research very briefly (100-200 words)

What are you looking into, who is involved, how many participants there will be, how they will be recruited and what are they expected to do?

Add your text here – (please avoid jargon and abbrevations)
We are researching what effects long-term and one-shot memory have on the performance of a conversational agent recommending recipes for a user/cook. The user is will be talking to a conversational agent created with Furhat, which is a digital face that talks to you. It will try to learn the preferences of the user over 3 sessions. In the first session, the participant will tell the Furhat its name and diets and so on, it will also recommend a recipe for the participant based on these diets/requirements. In the following sessions, Furhat will try to learn how to give better recommendations. After the three sessions, we will evaluate how well Furhat worked in the user's eye by using a questionnaire. When this questionnaire is filled in the other data used by Furhat will be deleted and only the anonymized questionnaire data is saved. Participants will be recruited through the Conversational Agent course, if this is not completely possible participants in the researcher's indirect environment will be contacted.

III. Risk Assessment and Mitigation Plan

NOTE: *You can find more guidance on completing your HREC application (including tips for completing this checklist)* ***here***

Please complete the following table in full for all points to which your answer is "yes". Bear in mind that the vast majority of projects involving human participants as "Research Subjects" also involve the collection of **Personally Identifiable Information (PII)** and/or **Personally Identifiable Research Data (PIRD)** which may pose potential risks to participants as detailed in Section G: Data Processing and Privacy below.

To ensure alighment between your risk assessment, data management and what you agree with your "Research Subjects" you can use the last two columns in the table below to refer to specific points in your Data Management Plan (DMP) and Informed Consent Form (ICF) – **but this is not compulsory**.

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT**	**MITIGATION PLAN**	**DMP**	**ICF**
A: Partners and collaboration						
1. Will the research be carried out in collaboration with additional organisational partners such as: • One or more collaborating research and/or commercial organisations • Either a research, or a work experience internship provider[1] [1] *If yes, please include the graduation agreement in this application*		x				
2. Is this research dependent on a Data Transfer or Processing Agreement with a collaborating partner or third party supplier? *If yes please provide a copy of the signed DTA/DPA*		x				
3. Has this research been approved by another (external) research ethics committee (e.g.: HREC and/or MREC/METC)? *If yes, please provide a copy of the approval (if possible) and summarise any key points in your Risk Management section below*		x				
B: Location						
4. Will the research take place in a country or countries, other than the Netherlands, within the EU?		x				
5. Will the research take place in a country or countries outside the EU?		x				
6. Will the research take place in a place/region or of higher risk – including known dangerous locations (in any country) or locations with non-democratic regimes?		x				
C: Participants						

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT**	**MITIGATION PLAN**	**DMP**	**ICF**
7. Will the study involve participants who **may** be vulnerable and possibly (legally) unable to give informed consent? (e.g., children below the legal age for giving consent, people with learning difficulties, people living in care or nursing homes,).		x				
8. Will the study involve participants who **may** be vulnerable under specific circumstances and in specific contexts, such as victims and witnesses of violence, including domestic violence; sex workers; members of minority groups, refugees, irregular migrants or dissidents?		x				
9. Are the participants, outside the context of the research, in a dependent or subordinate position to the investigator (such as own children, own students or employees of either TU Delft and/or a collaborating partner organisation)? *It is essential that you safeguard against possible adverse consequences of this situation (such as allowing a student's failure to participate to your satisfaction to affect your evaluation of their coursework).*	x		**The participants could be known by our group investigating them. This could be roommates, friends or other people they know.**	**This does not result in a risk for a disturbance of results as they have no incentive to disturb the research. This is because there is no specific favourable outcome.**		
10. Is there a high possibility of re-identification for your participants? (e.g., do they have a very specialist job of which there are only a small number in a given country, are they members of a small community, or employees from a partner company collaborating in the research? Or are they one of only a handful of (expert) participants in the study?		x				
D: Recruiting Participants						
11. Will your participants be recruited through your own, professional, channels such as conference attendance lists, or through specific network/s such as self-help groups	x		**The participants are mostly recruited by asking students that also follow the same course to evaluate our conversational agent. This could impose a risk in evaluating since students may be codependent on each other for grading**	**By having an open research question, a students answer is not favourable in one direction. This means that that there is no correct/incorrect answer.**		
12. Will the participants be recruited or accessed in the longer term by a (legal or customary) gatekeeper? (e.g., an adult professional working with children; a community leader or family member who has this customary role – within or outside the EU; the data producer of a long-term cohort study)		x				
13. Will you be recruiting your participants through a crowd-sourcing service and/or involve a third party data-gathering service, such as a survey platform?		x				
14. Will you be offering any financial, or other, remuneration to participants, and might this induce or bias participation?		x				
E: Subject Matter *Research related to medical questions/health may require special attention. See also the website of the CCMO before contacting the HREC.*						
15. Will your research involve any of the following: • Medical research and/or clinical trials • Invasive sampling and/or medical imaging • Medical and *In Vitro Diagnostic Medical* Devices Research		x				

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT**	**MITIGATION PLAN**	**DMP**	**ICF**
16. Will drugs, placebos, or other substances (e.g., drinks, foods, food or drink constituents, dietary supplements) be administered to the study participants? *If yes see here to determine whether medical ethical approval is required*		x				
17. Will blood or tissue samples be obtained from participants? *If yes see here to determine whether medical ethical approval is required*		x				
18. Does the study risk causing psychological stress or anxiety beyond that normally encountered by the participants in their life outside research?		x				
19. Will the study involve discussion of personal sensitive data which could put participants at increased legal, financial, reputational, security or other risk? (e.g., financial data, location data, data relating to children or other vulnerable groups) *Definitions of sensitive personal data, and special cases are provided on the TUD Privacy Team website.*		x				
20. Will the study involve disclosing commercially or professionally sensitive, or confidential information? (e.g., relating to decision-making processes or business strategies which might, for example, be of interest to competitors)		x				
21. Has your study been identified by the TU Delft Privacy Team as requiring a Data Processing Impact Assessment (DPIA)? *If yes please attach the advice/ approval from the Privacy Team to this application*		x				
22. Does your research investigate causes or areas of conflict? *If yes please confirm that your fieldwork has been discussed with the appropriate safety/security advisors and approved by your Department/Faculty.*		x				
23. Does your research involve observing illegal activities or data processed or provided by authorities responsible for preventing, investigating, detecting or prosecuting criminal offences *If so please confirm that your work has been discussed with the appropriate legal advisors and approved by your Department/Faculty.*		x				
F: Research Methods						
24. Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g., covert observation of people in non-public places).		x				
25. Will the study involve actively deceiving the participants? (For example, will participants be deliberately falsely informed, will information be withheld from them or will they be misled in such a way that they are likely to object or show unease when debriefed about the study).		x				
26. Is pain or more than mild discomfort likely to result from the study? And/or could your research activity cause an accident involving (non-) participants?		x				

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT**	**MITIGATION PLAN**	**DMP**	**ICF**
27. Will the experiment involve the use of devices that are not 'CE' certified? *Only, if 'yes': continue with the following questions:*		**x**				
• Was the device built in-house?						
• Was it inspected by a safety expert at TU Delft? *If yes, please provide a signed device report*		**x**				
• If it was not built in-house and not CE-certified, was it inspected by some other, qualified authority in safety and approved?		**x**				
28. Will your research involve face-to-face encounters with your participants and if so how will you assess and address Covid considerations?	**x**		**Covid regulations are currently not in place anymore. This could change in the coming weeks as the wave of new covid registrations is starting/growing.**	**If the regulations change we will abide by the regulations provided by the Dutch government. The participant should only talk to a computer, this computer can be disinfected after each use to reduce the chance of infection.**		
29. Will your research involve **either:** a) "big data", combined datasets, new data-gathering or new data-merging techniques which might lead to re-identification of your participants **and/or** b) artificial intelligence or algorithm training where, for example biased datasets could lead to biased outcomes?	**x**		**a. Data will be gathered for training the algorithm which recipes a user likes. It will only store the age and personal name of the user. It will use this to address the user when talking to the agent. As the surname is not taken this is not enough to identify a user.**	**The name and identificational properties of the user are only stored through the 3 sessions the user has with the conversational agent. After these sessions all data related to the user will be deleted and only a anonymized questionnaire is saved.**		
G: Data Processing and Privacy						
30. Will the research involve collecting, processing and/or storing any directly identifiable PII (Personally Identifiable Information) including name or email address that will be used for administrative purposes only? (eg: obtaining Informed Consent or disbursing remuneration)		**x**				
31. Will the research involve collecting, processing and/or storing any directly or indirectly identifiable PIRD (Personally Identifiable Research Data) including videos, pictures, IP address, gender, age etc	**x**		**The participant will be fysically asked to accept the consent form. The name and age will be asked by the conversational agent but will only be stored locally. Also the specifications of a participants face are measured for identification and guessing the participants facial expressions. Videos and pictures are not saved but only translated to data.**	**All information is saved only in the 3 sessions the user has with the conversational agent. After these sessions all the information the agent has gather will be deleted. By doing so the participant can not be identified by the collected data. The participant will also fill in a questionnaire, this will be anomomyzed.**		
32. Will this research involve collecting data from the internet, social media and/or publicly available datasets which have been originally contributed by human participants	x		**We will use a API to get recipes, these are made by humans but have no personal identification in them.**	**The dataset does not contain any properties that can identifie or be linked to humans. Only recipes and ingredients etc.**		
33. Will your research findings be published in one or more forms in the public domain, as e.g., Masters thesis, journal publication, conference presentation or wider public dissemination?		**x**				
34. Will your research data be archived for re-use and/or teaching in an open, private or semi-open archive?		**x**				

H: More on Informed Consent and Data Management

__NOTE:__ You can find more guidance on completing your HREC application (including tips for preparing your Informed Consent materials) __here__

Your research involves human participants as "Research Subjects" if you are recruiting them or actively involving or influencing, manipulating or directing them in any way in your research activities. This means you must seek informed consent and agree/ implement appropriate safeguards regardless of whether you are collecting any PIRD.

Where you are also collecting PIRD, and using Informed Consent as the legal basis for your research, you need to also make sure that your IC materials are clear on any related risks and the mitigating measures you will take – including through responsible data management.

IV. Signature/s

Please note that by signing this checklist list as the sole, or Responsible, researcher you are providing approval of the completeness and quality of the submission, as well as confirming alignment between GDPR, Data Management and Informed Consent requirements.

Name of Responsible Researcher (print)

Signature (or upload consent by mail) Responsible Researcher:

Date:

V. Completing your HREC application

Please use the following list to check that you have provided all relevant documentation

Required:

- **Always:** This completed HREC checklist
- **Always:** A data management plan (reviewed, where necessary, by a data-steward)
- **Usually:** A complete Informed Consent form (including Participant Information) and/or Opening Statement (for online consent)

Please also attach any of the following, if relevant to your research:

Document or approval	Contact/s
Full Research Ethics Application	After the assessment of your initial application **HREC will let you know if and when you need to submit additional information**
Signed, valid Device Report	Your Faculty HSE advisor
Ethics approval from an external Medical Committee	TU Delft Policy Advisor, Medical (Devices) Research
Ethics approval from an external Research Ethics Committee	Please append, if possible, with your submission
Approved Data Transfer or Data Processing Agreement	Your Faculty Data Steward and/or TU Delft Privacy Team
Approved Graduation Agreement	Your Master's thesis supervisor
Data Processing Impact Assessment (DPIA)	TU Delft Privacy Team
Other specific requirement	Please reference/explain in your checklist and append with your submission