The influence of a personal conversational recommendation on the user experience in chatbots

Submitted July 8, 2020,
in partial fulfillment of the conditions for the award of the degree BSc Artificial Intelligence by:

Dennis den Hollander[1]
s4776658

Supervisor: M.A. De Sa Siqueira[2]

[1]Department of Artificial Intelligence, Radboud University
[2]Behavioural Science Institute, Radboud University

Abstract

Chatbots are increasingly used in our daily lives, such as a virtual assistant, in e-commerce or in customer service. Although these chatbots can get the job done, customers often feel dissatisfied when, for example, the chatbot speaks out of context or does not remember information given earlier by the user. Also forms or advanced searches on the web are often replaced by chatbots. The question is raised whether this is actually an improvement. In this research is investigated how a personal conversational recommendation influences the user experience of a chatbot whose purpose is to give movie recommendations. A personal conversational recommendation (PCR) is a combination of a setting in the dialogue that enabled a more form-free chat and the personalisation of a chatbot, such that it looks and feels like an actual form. Hypotheses about the user experience were testified using a fully implemented chatbot that is able to give a PCR versus a chatbot that does not have this attribute with user testing. Participants had to answer a questionnaire with questions that are mainly conform with the Technology Acceptance Model (TAM).

The results indicate that both chatbots show similar user experience characteristics: the perceived ease of use, perceived usefulness, social presence and enjoyment were not significantly different. On this basis, it seems that people do not care about how the chatbot gives a movie recommendation. As long as the chatbot actually gives a recommendation, people are satisfied, because they achieved the intended goal, which is related with self-efficiacy in chatbots. Further research is needed to identify how people react to an actual form instead of a chatbot that converses in a form-like unpersonalized manner.

Contents

Chapter 1

Introduction

It is impossible to avoid chatbots in the world as we know it today. A growing number of companies are outsourcing parts of their customer service to chatbots to save a lot of labour costs and time; 30.000 chatbots had been built within Facebook Messenger so far in 2016 (Constine & Perez, 2016). These conversational agents often solve many of the problems these companies face. Nonetheless, it is important to make no mistakes when implementing these chatbots, because they otherwise lead to bad experiences. Examples are situations where the chatbot does not answer the question correctly or it does not remember information that has given earlier on. These aspects of chatbots often lead to emotional frustration (Luger & Sellen, 2016; Jenkins, Churchill, Cox, & Smith, 2007), which is why it is important to study the user experience of conversational agents thoroughly. Similarly, chatbots have been developed with recommendation systems to find a movie to watch or a book to read. A notable example of this is the chatbot "And Chill", where you can find a suitable movie based on your preferences regarding other movies. The question was raised whether these chatbots are considered as a better solution in comparison to web forms (Meng & Khelladi, 2017). This study reported that the web form service outperformed the chatbot in terms of voting, but it was not linked with user experience to find out which aspects could explain this phenomenon.

There are many aspects of user experience which could influence the way a user feels when having a conversation with a conversational agent. In this thesis, there will be a focus on the aspect of utilizing a personal conversational recommendation (PCR) in a chatbot. A conversational recommendation is a setting in the dialogue that enables a chatbot to transform its natural language into a more free-form chat (Li et al., 2018). For a personal conversational recommendation, this setting is intertwined with the personalization of the chatbot, where it has a name and the users are able to find out the way the chatbot converses and other traits. This technique is often used to give users a feel for the authenticity and humanity of a chatbot such that it becomes believable in the eyes of users (Kuligowska, 2015).

The goal of this thesis is to find out what the influence of a PCR is on the user experience in chatbots. The main research question is:

What is the influence of a personal conversational recommendation on the user experience in chatbots?

Put differently, it will be researched whether a chatbot whose dialogue is form-like and with no personality has a better user experience than a chatbot that is personal and has a form-free dialogue. It is not compared to an actual web form in order to keep design purposes the same.

Chapter 2

Preliminaries

This chapter covers a brief history of how chatbots came to be and some background of the Technology Acceptance Model (TAM), User Experience and the two addenda to the TAM: Social Presence and Enjoyment. It also contains a short section about the functionality of DialogFlow, a platform that was used in the chatbots that functions as a Natural Language Processor.

2.1 History of chatbots

The idea of chatbots originates from the Turing Test in the 1950s, where a participant would have to decide whether they were chatting with a computer or a real human (Turing, 1950). This test was designed to meet the intelligence criterion of computers: being able to tell whether a computer could impersonate the conversation techniques of a real human in such a way, that the participant could not tell them apart. This eventually led to a great interest of Joseph Weizenbaum and ELIZA was developed in the 1960s as the first chatbot who could pass the turing test (Weizenbaum, 1966). Although ELIZA passed the turing test, Weizenbaum never claimed that ELIZA truly met the criterion of intelligence. He stated that a large part of ELIZA's elegance may be credited to the fact that ELIZA creates an illusion of understanding the participants messages with a very superficial implementation. The shortcomings of ELIZA eventually led to more chatbots being created.

Along with this uprising of improving chatbots, there have been many attempts to create a chatbot that gives a personal conversational recommendation. Building individual user models was demonstrated to be highly useful and beneficial when creating stereotypes in conversation (Rich, 1979). Based on interaction, the computer would infer how to act and what to say. Close to that is research where the possibility of an adaptable chatbot are proposed and implemented (Thompson, Goker, & Langley, 2004). A big downside of this was the fact that dialogue and personalization had to be weighed up against each other instead of going side by side. Many techniques have been used to cross the bridge for such a recommendation chatbot, such as a belief system with states and reward functions (Sun & Zhang, 2018) or constraint satisfaction (Göker & Thompson, 2000). Although it seemed possible to have both a personalization and a good recommendation, these chatbots are most of the times very brief in dialogue length and do not provide a personalized experience. Others have proposed ways to improve the user experience with a conversational recommendation, but this is merely a strategy and there is no user testing (Narducci, de Gemmis, Lops, & Semeraro, 2018).

Nowadays, chatbots are not only used in research, but also the individual increasingly uses chatbots to improve their lifestyle. Famous examples of chatbots are virtual assistants, such as Google's *Google Assistant*, Amazon's *Alexa* and Microsoft's *Cortana*, but chatbots are also used in *Facebook Messenger*, often via companies, and in other places. The amount of users of these virtual assistants increases quickly; it has been estimated that there will be 1.8 billion users of virtual assistants in 2021 (Richter, 2016). This success is due to the fact that virtual assistants are not only available on smartphones applications, but also in other fields, such as telecommunication, education and healthcare.

2.2 User Experience

User experience (UX) is a collection of emotions, intentions and attitudes from the user when using a given product. This term gives an idea of how people experience technology, which can be both good or bad. UX is a crucial part of modern systems that are built today. When developing a product, the developer cannot always foresee all problems that might come up with new software, which is the primary reason that there are often user tests before the software is being put out to the public. A/B testing is an example of how a difference in website design can influence the user experience.

2.3 Technology Acceptance Model

The Technology Acceptance Model (TAM) is probably the most widely cited theoretical framework that explains why people adapt technology into their daily lives based on the concept of UX (David, 1989; Wu, 2009). The structure of the TAM is shown in Figure 2.1. The model suggest that external variables, which come from the technology that is used, influence two main factors: Perceived Ease of Use (PEoU) and Perceived Usefulness (PU). The PEoU is the degree to which a person believes that using a particular system would be free from effort and the PU is the degree to which a person believes that using a particular system would enhance his or her job performance. These two factors influence the Attitude Towards Using (ATU) of the technology, which is the general impression that people get when using technology. Before the actual system use, there is also a factor Behavioral Intention to Use (BI). However, this will not be measured in the thesis, because this is derivable from PU and ATU to explain usage of the actual system. In the last couple of decades, many studies have been conducted that support the TAM and provide empirical evidence (Ma & Liu, 2004; Chau, 1996).

2.4 Additions to the TAM

In this thesis, two additional constructs will be added to the TAM: Social Presence (SP) and Enjoyment (E). Both concepts will be shortly introduced in this chapter.

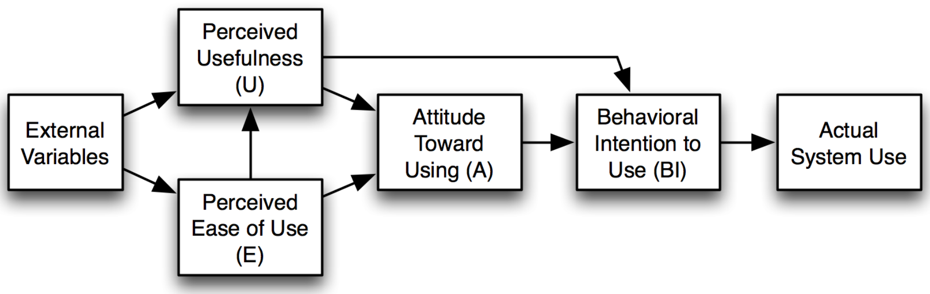


Figure 2.1: The Technology Acceptance Model (David, 1989)

2.4.1 Social Presence

An important aspect of the user experience in technology is social presence. Social presence is the degree of salience of the user of technology when a communication medium is able to create a feel of belonging, intimacy and warmth, such that they help to form personal relationships with the user. In a recent study, social presence was linked with chatbots, where the influence of anthropomorphic design cues were evaluated (Araujo, 2018). Social presence has been previously linked and used as a later addendum to the TAM to explain the dimensions of cultural differences among countries (Straub, 1994).

2.4.2 Enjoyment

Another addition to the TAM will be the perceived enjoyment, which is the degree to which a person finds the conversation with the chatbot enjoyable. A recent study found out that also perceived enjoyment has a significant impact on the usage intention of technology (Yang & Lee, 2019), where this construct is positively influenced by the visual cues of the technology. Enjoyment has also been considered as a critical hedonic and intrinsic motivation to interact with a conversational agent (Venkatesh, 2000). Enjoyment was also added as an addendum to the TAM in similar studies to measure effects as self-efficiacy and trust in technology (Mun & Hwang, 2003; Ha & Stoel, 2009; Teo & Noyes, 2011).

Chapter 3

Design

This chapter contains an explanation of how the conceptual model was constructed, why this design was chosen, how it addresses the problem and how it can answer the research question. It also gives an insight of what is changed in conversation style between the chatbots to determine effects of a Personal Conversational Recommendation in the experiment.

3.1 Hypotheses

To answer the research question, two movie recommendation chatbots will be developed for testing. One of them will give a PCR and the other will not, where the conversation is nor personal nor implemented in a conversational manner. Based on user preferences, the chatbot will give you a movie recommendation. To access the causalities in user experience, the Technology Acceptance Model (David, 1989) will be used as a theoretical framework. The Perceived Ease of Use (PEoU), Perceived Usefulness (PU) and the Attitude Towards Using (ATU) will be tested, along with two extra variables: Enjoyment (E) and Social Presence (SP). Since personalization will be modified between the chatbots, it is important to measure the effects of these constructs on the user experience as well.

Taking the structure of this model into account, 10 hypotheses will be testified in this thesis. H1 to H4 will address the chatbot itself. The chatbot that gives a personal conversational recommendation will be abbreviated as *cPCR* and the chatbot that does not give a personal conversational recommendation as *cNON-PCR* respectively:

H1 *cPCR* is perceived as less easy to use than *cNON-PCR*

H2 *cPCR* is perceived to be more useful than *cNON-PCR*

H3 *cPCR* is perceived as having a higher social presence than *cNON-PCR*

H4 *cPCR* is perceived as more enjoyable than *cNON-PCR*

In addition, H5 to H10 will address the relationship between the Perceived Ease of Use, Perceived Usefulness, Enjoyment, Social Presence and the Attitude Towards Using. These hypotheses are based on the assumption that each of these factors influence each other in a positive manner.

H5 The higher the perceived ease of use of using the chatbot, the higher the people's attitude towards using it

H6 The higher the perceived usefulness of using the chatbot, the higher the people's attitude towards using it

H7 The higher the social presence of using the chatbot, the higher the people's attitude towards using it

H8 The higher the enjoyment of using the chatbot, the higher the people's attitude towards using it

H9 The higher the enjoyment of using the chatbot, the higher the people's social presence of using it

H10 The higher the perceived ease of use of using the chatbot, the higher the people's perceived usefulness of using it

In figure 3.1 can be seen how these hypotheses are structured:

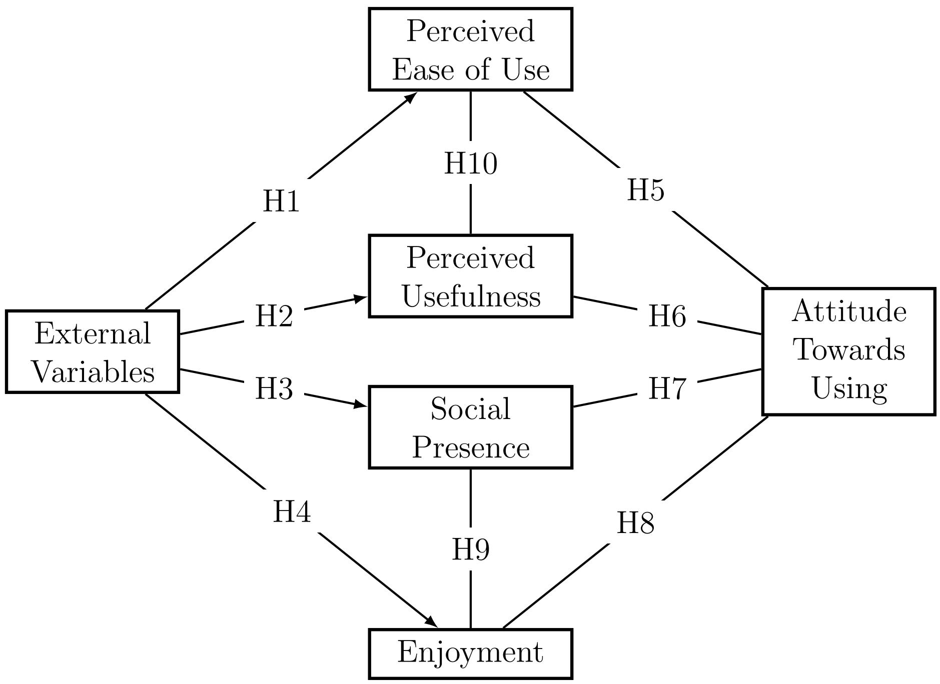


Figure 3.1: The hypothesis diagram

3.2 DialogFlow

DialogFlow (formerly known as API.AI) is a conversational Artificial Intelligence (AI) from Google that will function as a Natural Language Processor in the chatbots that will be developed. In DialogFlow, you can create an agent to which you can communicate with by using an Application Programming Interface (API). DialogFlow uses several agent utilites to be able to understand and act upon what the user is saying. In subsection 3.2.1 to subsection 3.2.4, some of these core functionalities will be briefly addressed.

3.2.1 Intents

Each time a user of the chatbot says something, the chatbot tries to figure out what the users means and wants to achieve with this sentence. An intent is a category that has been predefined to be able to handle a conversation. The agent uses a threshold to decide if it classifies an intent or if it goes back to a fallback intent. A fallback intent is an intent that is called whenever the chatbot does not understand something. This might be because the user of the chatbot told something that was out of context or because the chatbot does not understand that specific type of answer. This threshold is based on the degree of certainty. In Figure 3.2, you can see how this intent classification is done.

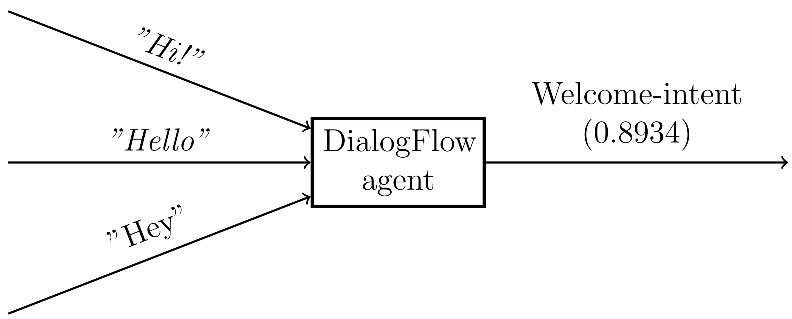


Figure 3.2: The way a DialogFlow agent classifies an intent. The number between brackets is an example of such a degree of certainty. In the developed chatbots for this study, the threshold is set at 0.3, therefore in this specific example, the welcome intent is triggered instead of a fallback intent

3.2.2 Entities

An entity is a data type that deals with variables that could occur in sentences of the user. As an example, a user could provide a time indication for the maximum movie length of *"30 minutes"*, but also *"1 hour"* or *"two days"*. Since it is hard to extract the actual value from these words, DialogFlow has some default entities created for these issues (see Figure 3.3). It is also possible to create custom entities. For the movie recommendation chatbot, we created a movie genre entity that could recognize possible genres that the user would fill in. DialogFlow also uses a bit of Machine Learning in order to also account for possible typing mistakes (i.e. "hrror" will be re-evaluated to "horror" instead of going back to a fallback intent).

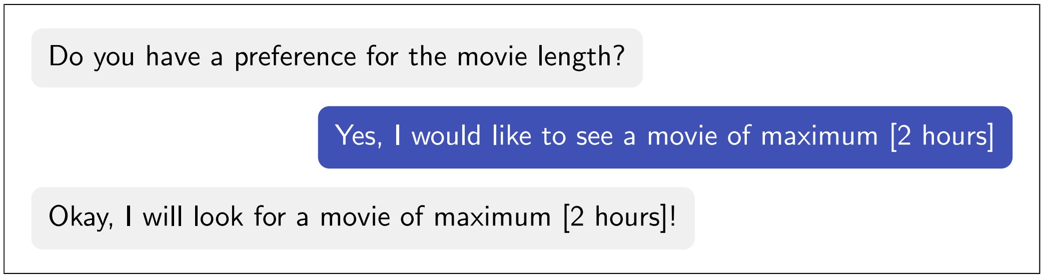


Figure 3.3: An example of entity extraction done by DialogFlow. The part between the square brackets is the entity information that DialogFlow is able to extract as ”movie length” using a combination of pattern matching and machine learning.

3.2.3 Context

To make sure that the chatbot does not end up not knowing what the conversation is about, context is very important. Context is divided into input and output contexts, which are both necessary to maintain a healthy conversation, but also to control the conversation flow, which was necessary to have for the developed movie recommendation chatbot. Context is a setting that is very similar to what happens with natural language. If someone would say *”I like this very much”* out of nowhere, you would not know what *this* refers to. Chatbots operate in the same way, so if some context is provided, they can derive the meaning and act upon this meaning.

3.2.4 Fulfillment

In most cases, a chatbot is able to react with some predefined sentence that is provided beforehand in DialogFlow itself. But sometimes, this response needs to have a custom reaction, for example when a user asks for a movie recommendation and the agent needs to retrieve a customized movie recommendation based on the preferences. This is where fulfillments come into play.

With fulfillments, DialogFlow is able to send the back-end of the webserver a custom request, which is called a webhook. Each time the agent calls the webserver, the back-end is able to perform some action based on the intent that the webhook was called with. If the intent is *movie-recommendation* for example, the back-end acts upon this request by retrieving movies based on the preferences that the user had given earlier. Once the appropriate recommendation has been found, a response will be send back to DialogFlow. Finally, DialogFlow will fill in the correct pieces of information from this recommendation in a predefined sentence.

3.3 Conversation flow of the chatbot

In this section, there will be an overview of how the chatbot conversation was actually implemented in DialogFlow and some of the considerations. Each participant had to communicate with a chatbot that had a certain conversation flow. This conversation starts with a greeting, followed by a back-and-forth question and answer pattern to find out the user's preferences about the movie. Finally, a movie recommendation will be given to the user and then the user is able to ask more questions about this movie, for example what the cast of the movie is. In case the user has no more questions about the movie, the conversation is terminated. To make sure that the chatbot never would stop talking mid-conversation, because it wouldn't know what to say back, we created a fallback intent. In Figure 3.4, you can find an example of how such a fallback is achieved and how this prevents the conversation to drift off to another topic.

Both chatbots were following almost exactly the same traject of topics that they were going to ask. These topics are, in order, *name* (only in the control condition chatbot), *favourite movie*, *genres*, *age*, *favourite actors / actresses*, *favourite director*, *movie duration*, *minimum release year* and *rating*. With these questions both chatbots were able to construct an object of preferences which were going to be used to retrieve a suitable movie.

A more exact diagram of the conversation flow of both the PCR and non-PCR chatbot is found in appendix B.

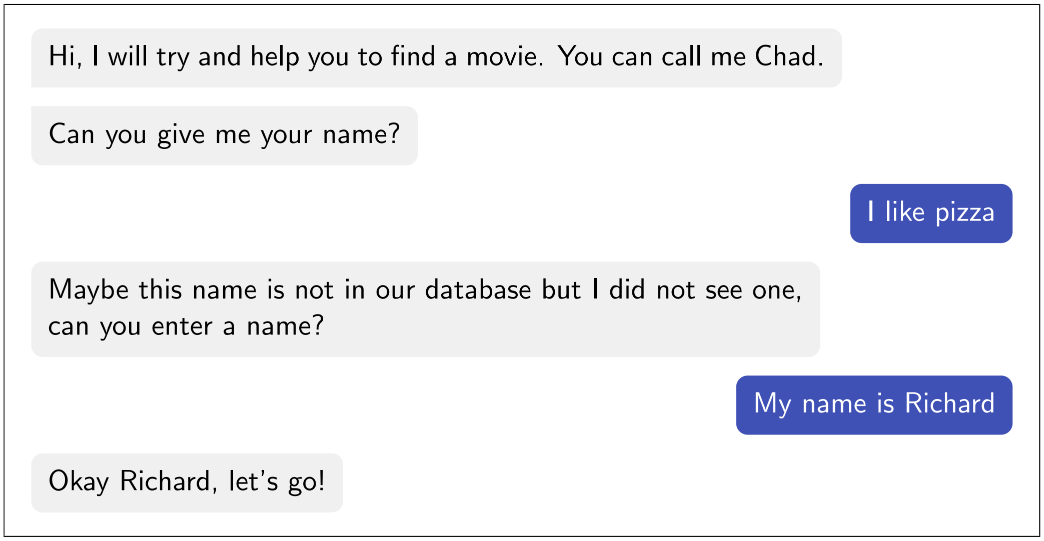


Figure 3.4: An example of how the chatbot maintains the chatbot conversation in the control condition when it does not understand what the user says or when the user fills in something completely irrelevant to the current context. Notice how the chatbots is determined to keep the subject on asking for the name until it detects one

3.4 Differences between the chatbots

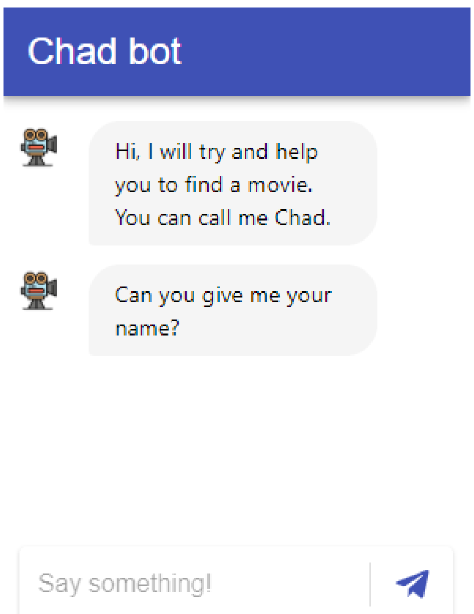
As an indication of what exactly changed between the two version of the chatbot, here is an overview on what is changed and why.

3.4.1 Chatbot personality

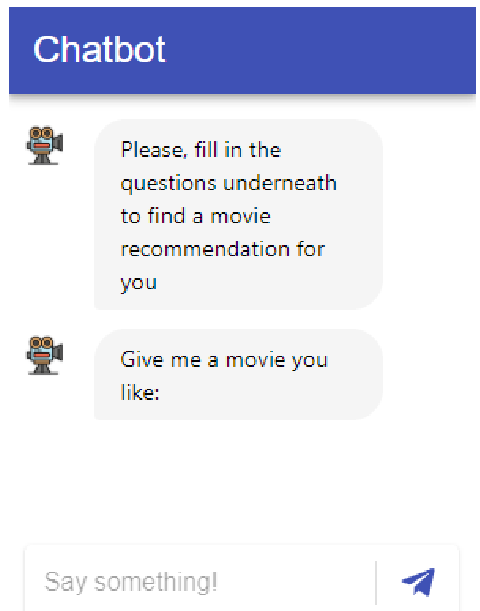
In order to remove the personality of the chatbot, it was decided to remove its name from the conversation and front-end. In addition, the user's name was not asked for in the experimental condition, since it is not necessary to ask in order to give the user a movie recommendation. A visual example is given in Figure 3.5. This way, possible emotional attachments that the user can have with the chatbot while having a conversation are eliminated.

3.4.2 Difference in the use of language

To simulate the effect of filling in a form as much as possible while still using a chatbot format, the chatbot straightly asks the questions and nothing more than that. Where the control condition would use a lot of conjunction words and give confirmation of what the user had filled in, the experimental version will not do this. It will only say something extra if the user has filled in something that it could not process, for example when the user says something out of context or gives a genre that does not exist. This effect is shown in Figure 3.5: the chatbot that does not give a PCR is very straightforward when asking the preference question.



(a) Chatbot interface with a PCR



(b) Chatbot interface without a PCR

Figure 3.5: A difference of the top half of the front-end between the control condition (a) and the experimental condition (b). Notice that the chatbot does not give anything away about his name in the experimental condition.

3.4.3 The recommendation

The way both chatbots give a movie recommendation is also different. The control condition only says the name of the movie, the director and the year at first, and then the user can ask for more information about the movie after that until the participants knows enough and terminates the conversation himself. In constrast, the experimental condition gives everything it knows about the movie in one message and then immediately terminates the conversation. An example of this is shown in Figure 3.6.

Give me a movie recommendation please

URL	https://www.imdb.com/title/tt8305806
Title	The Wretched
Release year	2019
Genres	Horror
Duration	95
Director(s)	Drew T. Pierce, Brett Pierce
Description	A rebellious teenage boy embarks on ...
Cast	John-Paul Howard, Piper Curda, ...
Rating	6.5

Your password is: **MTU5MTYyNjMxODU2Ng==**
Please copy this password in your Qualtrics survey!

Figure 3.6: The way the experimental condition returns movie information. Note how the user is unable to ask anything after the recommendation, because the chatbot immediately ends it. This is done to simulate the effect of pressing the submit button after you fill in a form online. Some descriptions in the recommendation itself are shortened to reduce space.

3.5 Setup

As shown in the sections above, two versions of a chatbot were developed in order to be able to answer the research question. The control condition is a chatbot that has been build with a group that is able to give a personal conversational recommendation. The second chatbot is an experimental condition, where the ability to give a personal conversational recommendation has been removed. There were two restrictions for the participants: they must be able to speak English and they have to be 18 years or older. An age close to 18 years old was preferred, since there is existing evidence that older people have more troubles adapting to technology than young people, and it is possible that this has an influence on the data (Rheingold, 2007). This is generally explained with the fact that younger generations had access to technology from a very young age whereas older generations did not have this opportunity (Brosnan, 2002). Participants engaged in a two-part study that uses a between-subjects design, where each participant interacts with the same chatbot twice, with 1 day delay in between each conversation. This delay was introduced to mitigate the effects of novelty of technology. The novelty effect occurs when an individual tends to have a stronger stress response to the technology, because the first interaction with the chatbot could be perceived as a threatening experience (Gravetter & Forzano, 2018). Over time this effect wears off.

After each conversation, the participant would have to fill in a questionnaire on Qualtrics with questions that follow the structure of the TAM. Each question is based on the likert scale (i.e. 5 options: *strongly disagree*, *disagree*, *neutral*, *agree*, *strongly disagree*. This way, it is possible to operationalize the user experience in the analysis of the thesis. All the questions in the questionnaire can be found in appendix A. These questions are based on articles which also included a questionnaire in technology comparison based on the constructs of the TAM and/or enjoyment and social presence (Chung, Ko, Joung, & Kim, 2018; Lowenthal, 2012; Armentano, Christensen, & Schiaffino, 2015).

Furthermore, the participants also had to fill in information about themselves, for example their age, field of study (if relevant), highest level of education. They also had to provide information on whether they had previous experiences with computers and an e-mail in order to receive the second part of the study a day later. A schematic of the study is provided in figure 3.7.

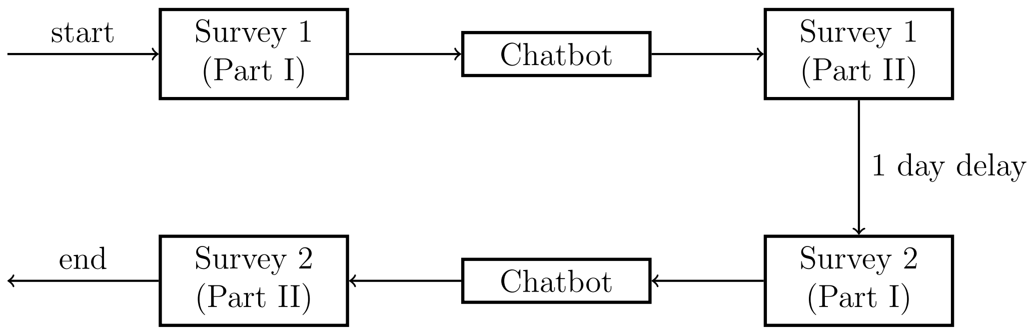


Figure 3.7: Flow of the experiment for one participant

Chapter 4

Results

The ten hypotheses that were introduced in chapter 3, can be split into two groups that are each evaluated differently. Hypotheses H1 to H4 are aimed to provide an insight into which factors of the TAM influence the user experience. These will be statistically analysed using a one-way ANOVA. Additionally, if the ANOVA indicates that the null hypothesis, two means being the same, can be rejected, means comparing will be done to find out which of the two chatbots performs better at the attribute. In this study, for each researched factor, the mean will be taken from each set of questions that corresponds with this factor. Despite the fact that the collected data is Likert-scale ordinal data and that there are resources indicating that evaluating means for ordinal data is not considered a good practise (Jamieson, 2004), arguments were provided to defend the statement that you can use parametric tests like ANOVA and Pearson correlations, even if the data is ordinal (Norman, 2010). Similarly, a lot of research exists that uses a one-way ANOVA and means comparing as analysis techniques for TAM factors, such as the Perceived Ease of Use and Perceived Usefulness (Edmunds, Thorpe, & Conole, 2012; Lu, Zhou, & Wang, 2009; Liu, Liao, & Peng, 2005).

Moreover, hypotheses H5 to H10 are constructed to test how well the TAM aligns with the results of this experiment. The results of these hypotheses will be analysed using Pearson's R in order to find out the relationship between components in the TAM and also the extension with Enjoyment and Social

Presence. Since the data is ordinal by default, to measure the correlation with Pearson's R, the sum will be calculated of each answer of the researched factor questions as was done in a comparable research (Joshi, Kale, Chandel, & Pal, 2015). This is done, because then no ordinal data is used. TAM correlations were also researched in a study where a meta-analysis of 88 TAM studies was performed (King & He, 2006).

4.1 Demographics

A total of 33 participants (23 females and 10 males with an average age of 25.6) participated in the experiments, of which 16 (11 female, 5 male) interacted with the chatbot that gives a PCR and 17 (11 female, 6 male) with the chatbot that does not give a PCR. 75.8% of all participants claimed to have had a conversation with a chatbot before, 18.2% did not and 6% didn't know.

4.2 Analysis

The descriptive statistics for each chatbot type and factor are shown in Figure 4.1. The means of each construct and chatbot type are above or equal to the midpoint 3.00 with a range of 3.00 to 4.27. The standard deviations vary from 0.50 to 1.08. Besides the mean and standard deviation, also skewness and kurtosis are included, in order to check for data normality (Teo & Noyes, 2011). Although data normality on the ordinal data itself was not needed as an assumption, this assumption is needed on the means of this data (Norman, 2010). The skewness is normally ranged with a range of -0.89 to 0.69, but the kurtosis is mostly a bit out of range: -1.34 to 0.74. However, according to Pearson (1931), non-normally distributed means of the data do not influence the robustness of ANOVA. This means that, even though the data violates the assumption for data normality, we can still use ANOVA to get the same relevant results.

An overview of the results of the one-way ANOVA tests for H1 to H4 can

Construct	Type	Mean	SD	Skewness	Kurtosis
PEoU	cPCR	4.27	0.50	0.69	-1.00
	cNON-PCR	4.12	0.84	-0.29	-0.86
PU	cPCR	3.50	0.96	-0.89	-0.48
	cNON-PCR	3.55	0.67	-0.29	-1.24
SP	cPCR	3.47	0.96	-0.27	-1.34
	cNON-PCR	3.28	0.90	-0.08	-1.29
E	cPCR	3.25	0.94	-0.26	-1.27
	cNON-PCR	3.00	1.06	0.36	-1.30
ATU	cPCR	3.59	1.08	0.05	-1.08
	cNON-PCR	3.74	0.83	-0.69	0.74

Figure 4.1: Descriptive statistics per construct (PEoU: Perceived Ease of Use, PU: Perceived Usefulness, SP: Social Presence, E: Enjoyment, ATU: Attitude Towards Using) and chatbot type (cPCR: chatbot that gives a personal conversational recommendation, cNON-PCR: chatbot that does not give a personal conversational recommendation)

be found in Figure 4.2. There was an insignificant effect of the perceived ease of use on the attitude towards using at the $p < .05$ level for the two chatbot conditions [$F(1, 97) = 1.203$, $p = 0.275$]. Therefore, hypothesis H1 is rejected. Similarly, there was no significant effect of the perceived usefulness on the attitude towards using at the $p < .05$ level for both chatbots [$F(1, 163) = 0.161$, $p = 0.689$]. This means that hypothesis H2 was rejected.

Regarding the two addenda to the TAM, there was an insignificant effect of the enjoyment on the attitude towards using at the $p < .05$ level for the two chatbot conditions [$F(1, 97) = 1.455$, $p = 0.231$], meaning that hypothesis H3 is rejected. There was also no significant effect of the social presence on the attitude towards using at the $p < .05$ level for both conditions [$F(1, 130) = 1.167$, $p = 0.282$]. Therefore, hypothesis H4 is rejected.

In addition to the hypotheses for the chatbot, there were also hypotheses that test how "in-line" the results of both chatbots are in this study compared to the Technology Acceptance Model. In order to do this, both the control and experimental condition were grouped and subsequently the correlations were calculated. The results can be found in Figure 4.3 and Figure 4.4.

	Df	**Sum Sq**	**Mean Sq**	**F-value**	**p-value**
ind	1	0.58	0.5803	1.203	.275
Residuals	97	46.77	0.4822		

(a) Perceived Ease of Use (PEoU)

	Df	**Sum Sq**	**Mean Sq**	**F-value**	**p-value**
ind	1	0.12	0.1155	0.161	.689
Residuals	163	117.01	0.7179		

(b) Perceived Usefulness (PU)

	Df	**Sum Sq**	**Mean Sq**	**F-value**	**p-value**
ind	1	1.55	1.546	1.455	.231
Residuals	97	103.00	1.602		

(c) Enjoyment (E)

	Df	**Sum Sq**	**Mean Sq**	**F-value**	**p-value**
ind	1	1.18	1.182	1.167	.282
Residuals	130	131.63	1.012		

(d) Social Presence (SP)

Figure 4.2: Results of the ANOVAs on all four testified hypothesis factors.

Hypothesis 5, which states that the higher the perceived ease of use, the higher the people's attitude towards using it, is rejected, because of its insignificance, $r(31) = .32, p = .073$. The relationship between these two constructs is also visible in Figure 4.4A. Compared to other relationships in this figure expect F, the regression line is somewhat too flat to indicate a strong relationship.

Next, hypothesis 6, stating that the higher the perceived usefulness, the higher the people's attitude towards using it, is supported by the data,

$r(31) = .64, p < .001$. This relation can be found in Figure 4.4B.

Hypothesis 7, which states that the higher the social presence, the higher the people's attitude towards using it, is supported by the data. The correlation between enjoyment and the attitude towards using is very significant and moderately positive, $r(31) = .59, p < .001$. The correlation between these two factors can be found in Figure 4.4C.

Hypothesis 8, which states that the higher the enjoyment, the higher the people's attitude towards using it, is supported by the data. The correlation between enjoyment and the attitude towards using is significant and moderately positive, $r(31) = .53, p < .01$. The correlation between these two factors can be found in Figure 4.4D.

	PEoU	**PU**	**SP**	**E**	**ATU**
PEoU	1				
PU	0.332	1			
SP	0.400*	0.624***	1		
E	0.366*	0.511**	0.844***	1	
ATU	0.317	0.636***	0.588***	0.528**	1

Figure 4.3: Correlation analysis of each construct. *means $p < .05$, **means $p < .01$ and ***means $p < .001$.

Hypothesis 9, which states that the higher the enjoyment, the higher the social presence of using it, is supported by the data. The correlation between enjoyment and the attitude towards using is significant and very positive, $r(31) = .84, p < .001$. The correlation between these two factors can be found in Figure 4.4E.

Finally, hypothesis 10, which states that the higher the perceived ease of use, the higher the perceived usefulness of using it, is not supported by the data. The correlation between enjoyment and the attitude towards using is insignificant and a bit positive, $r(31) = .33, p = .059$. The correlation between these two factors can be found in Figure 4.4F.

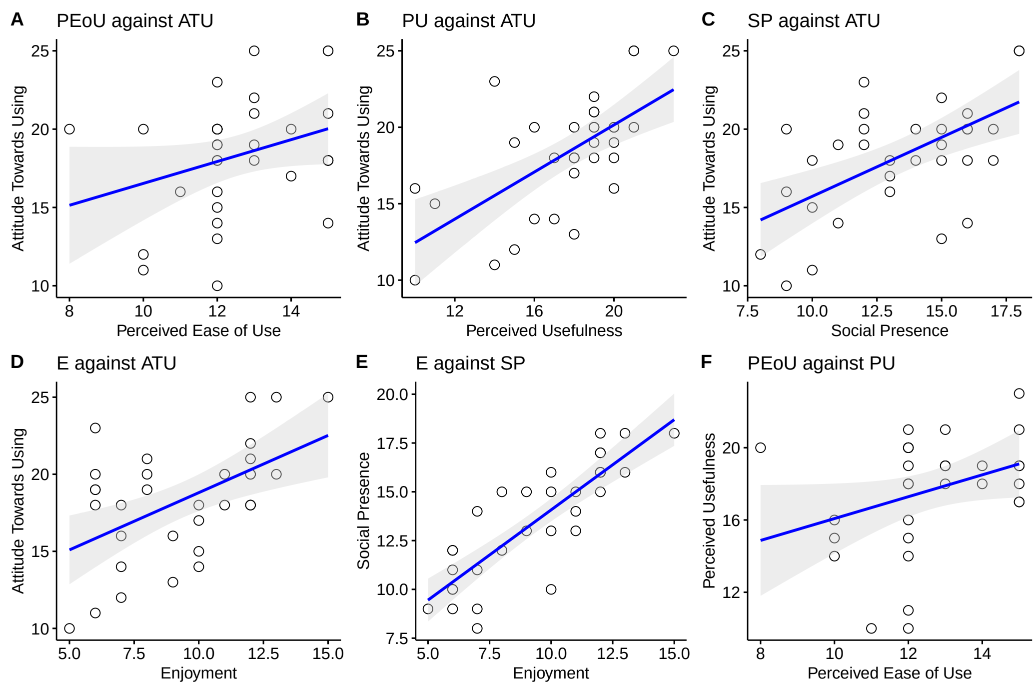


Figure 4.4: Correlation scatterplots of all tested variables in section 6.2 against the Attitude Towards Using.

Overall, hypotheses H1 to H5 and H9 were rejected. Hypotheses H6 to H8 and H10 were accepted.

Chapter 5

Discussion

In this thesis, it was found out whether a personal conversational recommendation influences the user experience in chatbots. In order to provide an explanation of the causalities of how people adapt technology into their daily lives, the TAM was used as a theoretical framework with two addenda: social presence and enjoyment. The results indicate that there were no significant differences between the control (cPCR) and the experimental chatbot (cNON-PCR).

It was hypothesized that the control chatbot would be less easy to use than the experimental chatbot. However, there was no significant difference between them. This effect could be explained because of the fact that many more factors influence the ease of use of a product than initially anticipated. Not only the average length of the messages of the chatbot, but also other things like overall clearity of the chatbot could have an influence on the ease of use. Because the experimental condition communicated in a very concise matter, not too many hints were given to the participant on how to actually use it. After all, we are not comparing a form itself, but rather a conversation where the dialogue of the chatbot is comparable of that of a form. This means that this type of conversation does not necessarily feel more intuitive. In addition, it was stated that the perceived usefulness would be higher in the control chatbot than in the experimental chatbot, because of the fact that such a conversational agent looks like a form, such that participants would not see an added value. But since this is something in between - a conversational form -, participants might have just interpreted this as more useful than a pure form, since it gives a new look to the oldfashioned web form.

Contrary to the hypothesized association that overall enjoyment and social presence would deteriorate when removing chatbot personalisation, this was not the case. A possible explanation for this might be that people prioritize getting a movie recommendation over actually having a fun conversation. As soon as the chatbot would give them a recommendation, they would enjoy it, because the goal has been achieved. This relates to a similar statement that was proposed by Venkatesh (2000), where a more enjoyable experience mitigated the difficulties of Human-Computer Interaction with respect to self-efficiacy.

The measured correlations between the model constructs of the TAM and the two addenda were mainly significantly positive. The correlations that were not significant, the perceived ease of use against the attitude towards using and the perceived ease of use against the perceived usefulness, were on the verge of being significant. This was mainly caused by the datapoints that had a summed score of 12 at the perceived ease of use questions. These participants had extremely varying opinions about the attitude towards using and the usefulness of the chatbots.

5.1 Future work

Possible future work should consist of a study with more participants. It was very hard to draw conclusions from only 33 participants and it would be interesting to see if some relations magnify or diminish. Future studies should also take into account a third version, which is a web form, such that you compare two chatbots (one that gives a PCR and one that doesn't) against a plain form that you can fill in digitally to receive a movie recommendation.

Chapter 6

Conclusion

This thesis aimed to find out what the influence is of utilizing a Personal Conversational Recommendation (PCR) in chatbots. It was found out that participants enjoy a chatbot that gives a PCR as much as a chatbot that doesn't. This is possibly the case because people tend to being okay with chatbots as long as they perform the task correctly (i.e. giving the user a movie recommendation). However, this conclusion should be taken with a grain of salt as it was not tested in the questionnaire whether the participants liked the movie recommendation they received. However, achieving such a goal might mitigate the negative effects of the chatbot.

The fact that there were no significant differences, might be explained by a multitude of limitations. First of all, the reliability of the data is probably impacted due to a lack of participants. Some of the correlations were just out of bound of being significant and it could very well be that extra participants would have solved this problem. This was also noticable when testing for normality, where many times the kurtosis would be outside the -1 to 1 range. In addition there were also some mistakes in the development of the experimental chatbot. The movie recommendation chatbot in the experimental version had still the same avatar, which should actually have no personality at all. Since people could then be able to imagine how the chatbot would look like in real life and make guesses about its behavior, this might have changed the results regarding the enjoyment and social presence, which is why they were almost similar.

Although there were some limitations, the data analysis still got a lot of interesting results. ANOVA on Likert-scale data works well and bridged the gap between having a low participant amount and functional data analysis. However, it was a missed chance to not compare this to a web form as well, such that three forms of a PCR are compared: a pure web form, a conversational agent that talks as a form and a chatbot that gives a personal conversational recommendation. This made it significantly harder to generalize the conclusions of this research to multiple fields.

This research contributed by providing an insight of how people experience a conversational agent who converses similar to a web form. With the results of the experiment, conclusions can be drawn that people do not necessarily need a personalized chatbot with human capabilities to have an overall good experience. In addition, even though a chatbot might use a setting in dialogue where it deviates from a form-like chat, this did not influence the perceived ease of use. This research also provided a new perspective on the situation where the ease of use, speed and convenience are the main reasons to use chatbots (Brandtzaeg & Følstad, 2017). In conclusion, conversational agents do not necessarily improve the user experience if they talk more like humans do. There might be other factors which cause this similarity.

References

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183-189.

Armentano, M. G., Christensen, I., & Schiaffino, S. (2015). Applying the technology acceptance model to evaluation of recommender systems. *Polibits*, *51*, 73-79.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. *International Conference on Internet Science*, 377-392.

Brosnan, M. J. (2002). Technophobia: The psychological impact of information technology. *Routledge*.

Chau, P. Y. (1996). An empirical assessment of a modified technology acceptance model. *Journal of management information systems*, *13*(2), 185-204.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2018). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*.

Constine, J., & Perez, S. (2016). Facebook messenger now allows payments in its 30,000 chat bots. *TechCrunch*. Retrieved from "https://techcrunch.com/2016/09/12/messenger-bot-payments/"

David, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Edmunds, R., Thorpe, M., & Conole, G. (2012). Student attitudes towards and use of ict in course study, work and social activity: A technology acceptance model approach. *British journal of educational technology*, *43*(1), 71-84.

Gravetter, F. J., & Forzano, L. A. B. (2018). *Research methods for the behavioral sciences*. Cengage Learning.

Göker, M. H., & Thompson, C. A. (2000). Personalized conversational case-based recommendation. *European Workshop on Advances in Case-Based Reasoning*, 99-111.

Ha, S., & Stoel, L. (2009). Consumer e-shopping acceptance: Antecedents in a technology acceptance model. *Journal of business research*, *62*(5), 565-571.

Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical education*, *38*(12), 1217-1218.

Jenkins, M. C., Churchill, R., Cox, S., & Smith, D. (2007). Analysis of user interaction with service oriented chatbot systems. *International Conference on Human-Computer Interaction*, 76-83.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *Current Journal of Applied Science and Technology*, 396-403.

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & management*, *43*(6), 740-755.

Kuligowska, K. (2015). Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professionals Center for Business Research*, *2*.

Li, R., Kahou, S. E., Schulz, H., Michalski, V., Charlin, L., & Pal, C. (2018). Towards deep conversational recommendations. *Advances in neural information processing systems*, 9725-9735.

Liu, S. H., Liao, H. L., & Peng, C. J. (2005). Applying the technology acceptance model and flow theory to online e-learning users' acceptance behavior. *E-learning*, *4*(H6), H8.

Lowenthal, P. R. (2012). Social presence: What is it? how do we measure it? *(Unpublished doctoral dissertation) University of Colorado Denver*.

Lu, Y., Zhou, T., & Wang, B. (2009). Exploring chinese users' acceptance of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory. *Computers in human behavior*, *25*(1), 29-39.

Luger, E., & Sellen, A. (2016). "like having a really bad pa" the gulf between user expectation and experience of conversational agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286-5297.

Ma, Q., & Liu, L. (2004). The technology acceptance model: A meta-analysis of empirical findings. *Journal of Organizational and End User Computing (JOEUC)*, *16*(1), 59-72.

Meng, A., & Khelladi, Y. (2017). Comparative use of web form, sms, and chatbot in social election monitoring of the dominican 2016 general election. *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, 1-5.

Mun, Y. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International journal of human-computer studies*, *59*(4), 431-449.

Narducci, F., de Gemmis, M., Lops, P., & Semeraro, G. (2018). Improving the user experience with a conversational recommender system. *International Conference of the Italian Association for Artificial Intelligence*, 528-538.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, *15*(5), 625-632.

Pearson, E. S. (1931). The test of significance for the correlation coefficient. *Journal of the American Statistical Association*, *26*(174), 128-134.

Rheingold, H. (2007). Smart mobs: The next social revolution. *Basic books*.

Rich, E. (1979). User modeling via stereotypes. *Cognitive science*, *3*(4), 329-354.

Richter, F. (2016). Digital assistants - always at your service. *Statista*. Retrieved from "https://www.statista.com/chart/5621/users-of-virtual-digital-assistants/"

Straub, D. W. (1994). The effect of culture on it diffusion: E-mail and fax in japan and the us. *Information Systems Research*, *5*(1), 23-47.

Sun, Y., & Zhang, Y. (2018). Conversational recommender system. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 235-244.

Teo, T., & Noyes, J. (2011). An assessment of the influence of perceived enjoyment and attitude on the intention to use technology among pre-service teachers: A structural equation modeling approach. *Computers & education*, *57*(2), 1645-1653.

Thompson, C. A., Goker, M. H., & Langley, P. (2004). A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research*, *21*, 393-428.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433-460.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, *11*(4), 342-365.

Weizenbaum, J. (1966). Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45.

Wu, P. F. (2009). User acceptance of emergency alert technology: A case study. *In Proceedings of the 6th International ISCRAM Conference*, 1-9.

Yang, H., & Lee, H. (2019). Understanding user behavior of virtual personal assistant devices. *Information Systems and e-Business Management*, *17*(1), 65-87.

Appendices

Appendix A

Questionnaire

Information sensitive parts like names and e-mail addresses have been replaced by three dots due to privacy reasons.

A.1 Questionnaire introduction

You are invited to participate in a research project in which you will interact twice with a chatbot used for movie recommendation. You will receive a task prior to interacting which you will have to attempt to complete. After each interaction, you will fill in an online questionnaire regarding your experience with the chatbot. This research project is being conducted by six Artificial Intelligence students of the Radboud University.

The procedure involves filling out an online survey twice. The questions concern your experiences during the interaction with our chatbot called Chad. Filling out the survey will take approximately 3 minutes. The interaction with the chatbot will take approximately 5 minutes.

A.2 Terms and conditions

What will happen to my data?
The research data we collect during this study will be used by scientists as

part of data sets, articles and presentations. The anonymized research data is accessible to other scientists for a period of at least 10 years. When we share data with other researchers, these data cannot be traced back to you. Moreover, for the purpose of sending a reminder e-mail with the second link, your e-mail address will be saved for 2 days. It will be saved separate from the data.

Voluntary Participation
Your participation in this research is voluntary. This means that you can withdraw your participation and consent at any time during the research, without giving a reason. All data we have collected from you will be deleted permanently.

Compensation
Participants participating via SONA will be compensated with 0.5 ppu. Participants that do not participate via SONA will not be awarded with any compensation for participating. However, we greatly appreciate the efforts made by participants.

More information?
Should you want more information on this research study, please contact ... (telephone: ...; email: ...).

Ethical assessment and complaints
This research study has been approved by the Ethics Assessment Committee Humanities of Radboud University (EACH file number ECSW-2019-097)

Should you have any complaints regarding this research, please contact the researcher.

You can also file a complaint with the secretary of the Social Science Ethics Committee of Radboud University. For questions on data processing in this research, please contact: ...

Clicking on the 'Agree' button below indicates that: you have read the above information you voluntarily agree to participate you are at least 16 years of age If you do not wish to participate in the research study, please decline participation by clicking on the "I do not want to participate" button.

☐ Agree (proceed to survey)

☐ Disagree

A.3 Part I

One of the chatbots requires the participant to have a Android phone with an English version of Google Assistent, do you have this? (If you have Google Assistent, but not English, you can easily change the language. This will help us out greatly!)

☐ Yes, I have a phone with Google Assistant

☐ No, I do not have a phone with Google Assistant

Keep in all cases the tab of this survey open. Please turn notification on your phone to silent mode, such that you will not be distracted during the conversation. Interact with the chatbot until you have received a movie recommendation. You will only get one recommendation which you need to accept. The chatbot itself can be found via the link on the next page. Ask Chad for some more information about the movie. Make sure you get to know at least 2 things about the movie. Things you can ask are: duration, genre, IMDB rating, cast or a short summary/description. **When you don't want to know anything more about the recommended movie you will receive a password. Copy-paste this password, because you will need it later on. After filling in the password in the corresponding field you can forget and delete it.**

Note: it is important that you finish the tasks in any circumstance. For this, it is necessary to keep continuing the conversation as long as the chatbot is replying. Since you will need to get to the end of the conversation in order to get the password.

For any questions regarding the tasks, please contact the researcher at: ...

A.4 Part II

Questionnaire about the interaction with Chatbot Chad

The last part of today consists of a short questionnaire about your experience with the chatbot. There will follow 20 statements about the chatbot which you need to answer on a scale from 'completely agree' to 'completely disagree'. After the statements a few general questions are asked.

The choices range from **1 (Strongly disagree)** to **5 (Strongly agree)**. **Note**: not the numbers but the corresponding texts were shown to the participant.

A.4.1 Perceived Ease of Use

	1	**2**	**3**	**4**	**5**
The chatbot is ease to use					
It was easy for me to learn how to use the chatbot					
The interaction with the chatbot is clear and understandable					

A.4.2 Perceived Usefulness

	1	2	3	4	5
Using the chatbot helps me find a movie that I like					
Using the chatbot saves me time in finding a movie					
This chatbot makes it easier for me to find a movie					
It is a good idea to use this chatbot to find a movie					
The chatbot understands my preferences regarding movies					

A.4.3 Social Presence

	1	2	3	4	5
Getting to know Chad gave me a sense of belonging during the conversation					
I was able to form impressions of Chad's behaviour					
I felt that my preferences were considered by Chad					
I felt comfortable during the conversation with the Chad					

A.4.4 Enjoyment

	1	2	3	4	5
It is enjoyable to share a conversation with the chatbot					
I was absorbed in the conversation with the chatbot					
The conversation with the chatbot was exciting					

A.4.5 Attitude Towards Technology

	1	2	3	4	5
Using this chatbot to find a movie is a good idea					
Using this chatbot makes it more interesting to find a movie					
I would like to use this chatbot to find a movie					
Using the chatbot to search for a movie seems fun					
I found the chatbot useful to find new movies that I would like to see and therefore I would use it again					

A.5 Questions about participant

These questions were only asked the second time the participant did the experiment after the chatbot questions.

What is your age?

___

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

What is your highest level of education?

- ☐ No education
- ☐ Primary school
- ☐ Secondary school
- ☐ HBO Bachelor
- ☐ HBO Master
- ☐ WO Bachelor
- ☐ WO Master
- ☐ PhD

In which field do you work or study?

- ☐ Artificial Intelligence
- ☐ Computer Science
- ☐ Psychology
- ☐ Communication
- ☐ Other

Before participating in this study, did you ever have a conversation with a computer?

- ☐ Yes
- ☐ No
- ☐ Maybe, I don't know

Do you participate in this study via SONA? If yes, enter your SONA ID code. You can find your SONA identity code under 'My profile' in your SONA account

- ☐ Yes, ____________________
- ☐ No

Appendix B

Conversation Flow

In Figure B.1 and Figure B.2 the conversation flows of both chatbots are found. All red arrows represent fallbacks that are triggered when an invalid input is entered. **Note:** this conversation flow is a simplified version of the actual conversation to get a basic idea of how the participants conversed with the chatbots and which topics were discussed for the preferences.

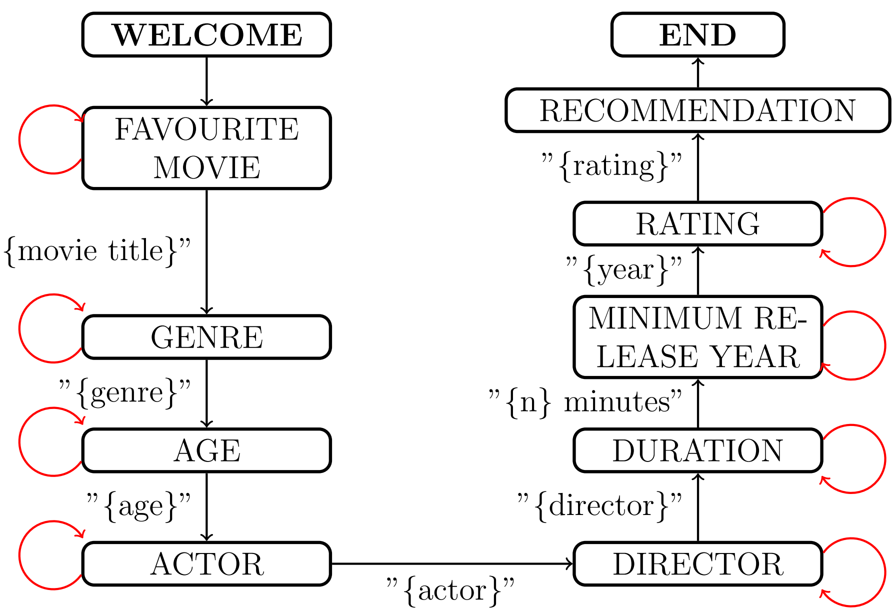


Figure B.1: Conversation flow of the non-PCR chatbot

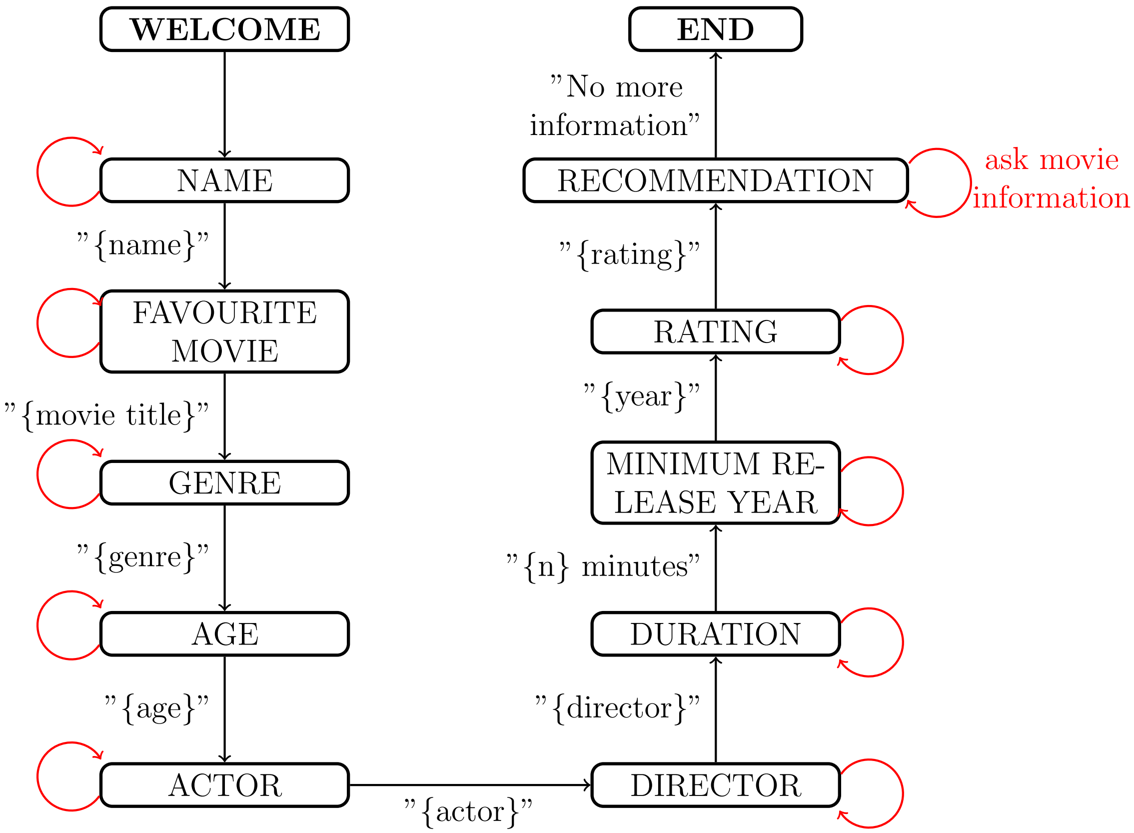


Figure B.2: Conversation flow of the PCR chatbot