Chatbots evaluating business ideas, does the evaluation cloud our judgement of their credibility?

Author: Joost van Loon
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,
In this thesis the impact of a chatbot's evaluation on its perceived credibility is investigated. This with the goal of finding out how to improve people's view on chatbots. Several variables are analyzed through the use of a Spearman's rank correlation model and a simple linear regression model in order to identify possible associations and effects between variables. From this analysis can be concluded that there is no significant and a very weak relationship between the ''preliminary-opinion'' variables, which are confidence, trust in AI and familiarity with AI, and the perceived credibility. There is however a significant strong relationship between the type of evaluation and perceived credibility. This leads to believe that the main hypothesis: ''Type of evaluation is associated with perceived credibility'' can be accepted. This result indicates that it is relevant for markets to investigate further into the opinion of consumers towards chatbots and how to improve adoption rates across the board with the knowledge that such investigations provide.

Graduation Committee members:

First supervisor	**Dr. M.L. Ehrenhard**
Second supervisor	**F. Koefer MSc**

Keywords
AI, marketing, chatbots, credibility, judgement, evaluation, perception.

1. INTRODUCTION

Artificial intelligence is becoming more and more integrated into our society nowadays. We see the ease of use of AI in our daily lives by an ever increasing amount by, to name a few, virtual assistants like Alexa and Siri. While Tesla-users will be able to experience the self-driving car driven by AI in the near future as "all new Tesla cars already have the hardware needed in the future for full self-driving in almost all circumstances" (Tesla, p. 1 2022). Even though artificial intelligence brings many new exciting innovations to the table in the near future, opinions on how we should utilize artificial intelligence are very divided. For instance Bill Gates has said that "humans should be worried about the threat posed by Artificial Intelligence'' (Rawlinson, p. 1, 2015) which could be an explanation for what (Dietvorst et al., p. 3, 2018)has concluded: "Decision makers are often averse to using algorithms, opting instead for the less accurate judgments of humans''. This means that we as people but also businesses miss out on potentially very efficient use cases of AI because we do not trust it enough compared to a possibly less accurate human input.

The reason that we as humans developed such a tendency to look at AI as an unreliable artifact is that they lack certain important human qualities such as diverse speech. For instance in a study by Hill et al. we can see that "human–chatbot communication lacks much of the richness of vocabulary found in conversations among people'' (Hill et al., p. 1, 2015). In addition to this we found that "artificial intelligence cannot yet comprehend ambiguous replies and are not able to hold longer conversations'' according to an experiment conducted by (Mittal et al., p. 3, 2016). What this means is that the average person will for instance, in a conversation with a chatbot, recognize said chatbot in a few exchanged words as the chatbot will not remember past chats and will not always recognize certain words for what they mean.

The earlier mentioned problem is applicable here. "After a chatbot is recognized and the respective task appears subjective, involving intuition or affect, customers are likely even less comfortable with AI'' (Castelo, p. 31, 2019). This will ultimately result in a chatbot that is perceived as less effective.

This introduction shows a glimpse of the potential artificial intelligence has and how it could be possibly applicable. However it is as mentioned easy to limit the use cases of AI and specifically chatbots when nowadays the average customer is less comfortable with AI when the task it performs involves intuition of affect as mentioned in (Castelo, p. 31, 2019). The fact that our beliefs and values about AI and chatbots in specific seem to be so crucial (on the surface) to the adoption rate outlines how important it is to research to what extent their perceived credibility is truly influenced. Only then can we conclude how to adapt chatbots in a way that they are more applicable in fields that require a more "human-like'' input.

1.1 Research Objective

Therefore, the goal of this research is to lay bare what influences the perceived credibility of a chatbot in order to make it clear what needs to be changed to make it more effective and approachable. To do this we need to make clear to what extent external factors from before using the chatbot in our experiment influence the outcome and to what extent the evaluation influences the outcome. This will be done by trying to answer the research question and respective sub question below.

Research question: "Does the nature of a chatbot's evaluation have an impact on its perceived credibility?''

Sub question: "Does the user's opinion on chatbots before interacting with one influence perceived credibility?''

These questions indicate whether there is an association and effect between the evaluation and perceived credibility. This is rather intriguing because it shows how the view on chatbots can be influenced by the output it gives. This contributes to our ultimate research goal which is finding out how to improve the view of people on chatbots.

2. THEORETICAL FRAMEWORK

In this section we discuss the topics that are of relevance to this study and that form the central foundation of this study in collaboration with our own experiment. The framework itself is divided into two sections: Theories that support the sub question and theories that support the research question.

2.1 Lay Beliefs About AI

It is shown in research from (von Walter et al., 2022) that lay beliefs are one of the aspects that affect our initial opinion of chatbots and other artificial intelligence. If our culture and beliefs are strongly against the replacement of human workers with chatbots we may already form a certain judgement of chatbots before we interact with them. The same can be said for the difference between organizational cultures. Where in central Africa there might be a slower rate of adoption of AI (Wairegi et al., p. 1, 2021), in first-world countries most organizations (especially big Tech firms) are looking towards optimizing through employment of chatbots and artificial intelligence in general which makes AI adoption rise rapidly. This leads to people having different views on the usefulness and the actual reliability of said AI. "Ultimately consumers who believe that AI is higher than human intelligence may expect to receive more accurate advice and may therefore expect to make better decisions when using algorithmic advice. Hence, they may be more motivated to use algorithmic advice. In contrast, consumers who believe that AI is inferior to human intelligence may not expect to receive more accurate advice. Consequently, they may be less likely to adopt algorithmic advice.'' (von Walter et al., p. 1, 2022). This explains to a certain extent how "lay beliefs'' may influence the assumed credibility of a chatbot before interacting with it.

2.2 Confidence In One's Own Judgement

As was already suggested in the introduction, initial opinion of our chatbot may already be impacted negatively if a person is overconfident in their innovation idea. Said person could argue, "why would I need to believe what an algorithm thinks of my idea when I already know my idea is solid?''

"One of the earliest explanations of overconfidence was based on the notion that people think of evidence in favor of their chosen answer (and against the alternatives) more easily than they think of evidence in favor of the alternatives (and against their chosen answer).'' (Harvey, p. 1, 1997). The difference in whether people think their own idea is decent enough from the start is what impacts this overconfidence as mentioned above if someone thinks their idea is good enough why should they believe someone or something that criticizes them.

Moreover an argument could also be made for the opposite side. Hypothetically speaking a person might not have an abundant amount of confidence in their idea. In this case the evaluation a chatbot makes will impact you in the opposite way as someone with overconfidence would be impacted so parts of the evaluation that turn out to be positive may be less trustworthy in the consumers eyes than the negative parts.

Another way that a participant's confidence could influence perceived credibility is when a participant has low confidence. According to (Lewandowsky et al., p. 2, 2000), in the case of a participant with low confidence said participant will move more to the side of automation compared to a human in the case of decision-making. This means that confidence in oneself could influence a person's opinion about the chatbot's perceived credibility before we interact with one.

The introduction of theories from chapters 2.1 and 2.2 already show the importance of clarification, what influences perceived credibility and what does not. We cannot simply assume that a person does not have any opinion towards their own business idea before interacting with a chatbot. Moreover it is would be negligent to assume that they do not have an opinion towards chatbots before interacting with one. When we lay bare what the major drivers are of perceived credibility only then can we decide on to what extent said credibility is actually impacted.

The reason it is important to include pre-chatbot interaction factors is because according to several earlier mentioned studies there are several factors that quite possibly influence the subjective view on our chatbot. Thus these factors could also influence the perceived credibility of our chatbot possibly making our measurements inaccurate if we were to not include these factors.

2.3 Chatbot Credibility

"Credibility is often characterized as a multifaceted concept that has been approached in terms of believability, trust, accuracy, fairness, objectivity, and reliability'' (Shin, p. 2, 2022). These different terms each represent a cornerstone of credibility and when one term is less present then we can assume, in the context of our experiment, that our chatbot is less credible in the eyes of the participant. Firstly objectivity and fairness: in the case of a chatbot these two terms are almost a given in the context of our experiment. Every business innovation is evaluated based on the same criteria and all these criteria are completely objectively analyzed as well.

Accuracy and reliability are terms that are variable in the context of our experiment. As will be mentioned in the methods-paragraph we will divide the experiment into a "placebo'' group and a group with an actually credible chatbot. The credible chatbot is expected to be more reliable as it will more constantly put out accurate evaluations and these accurate evaluations stem from a greater skillset compared to the not-credible chatbot. An argument could be made for the needed specification of accuracy and reliability. We as people deem a chatbot reliable and accurate when comparing the outcome of its evaluation to that of our own. Thus it could be argued that a chatbot is only accurate in the participant's eyes when it matches the participant's judgements of their own idea before interacting such a chatbot. This also shows how important it is to analyze the participant's confidence in their own judgements as stated in chapter 2.2.

Lastly, believability and trust. Both terms are according to Shin a matter of "assurance about issues of neutrality, impartiality, confidentiality, and objectivity'' (Shin, p. 2, 2022), qualities that our chatbot already has. Transparency may play a large role in our specific context as well. If we, as a test, disclose the evaluation process of our chatbot after our participant has given an answer on how credible they think our chatbot is, we can expect to see an increase in perceived credibility. This is because when the process of evaluation is made clear it shows the intentions of our chatbot. This is important because according to Shin and Park "trust is referred to as a user's feeling of confidence that the algorithms will perform actions that are beneficial'' (Shin & Park, p. 277-284. 2019). Thus showing that a chatbot is able to perform actions that are beneficial will increase trust in our AI.

2.4 Transparency

"One way to moderate unrealistically high expectations from users is to provide an explanation regarding virtual AI functionality.'' (Glikson & Woolley, p. 27, 2020). It is assumed to be normal to expect a virtual AI, considering how perfectly they can be tuned for the job, to be outperforming humans in certain tasks. One way to moderate these expectations is to introduce transparency. As we mentioned earlier we assume that transparency can be an important factor to building up trust between the participant and the chatbot and increasing its perceived credibility. This is further backed by (Fan et al., p. 7, 2008). It is demonstrated in their study that when a participant is informed of the reliability of a decision-making agent it would influence the participants' trust and performance for the better.

The limitations of this experiment are also marked by the theory of (Fan et al., p. 7, 2008) as it has become apparent that when sharing the "expert'' status or credibility of our chatbot prior to the experiment that a participant has shown an increase in trust towards AI. A variable that has not been included in this research.

3. HYPOTHESES

Based on our theoretical framework the following hypotheses have been created:

H1.1: Participants' confidence is associated with perceived credibility.

H1.2: Participants' familiarity with AI is associated with perceived credibility.

H1.3: Participants' trust in AI is associated with perceived credibility.

H2: Type of evaluation is associated with perceived credibility

In Hypothesis 2 advice adoption is chosen instead of perceived credibility as according to (Jiang et al., p. 16, 2021) "The quality of information can help users judge the credibility of the information, thereby prompting consumers to consider whether to adopt the information and make corresponding purchase decisions.'' With this we mean that perceived credibility leads directly to the adoption of advice.

3.1 Conceptual Framework

In the framework below the hypotheses are put into perspective to give an organized overview.

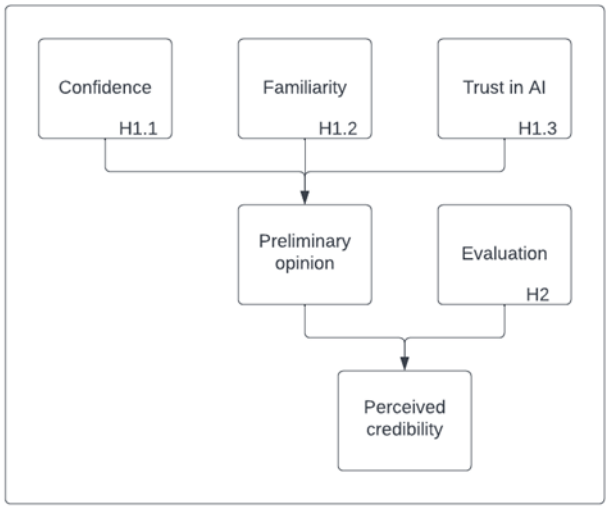


Figure 1. Conceptual framework of hypotheses

4. METHODOLOGY

4.1 Research Design and Data Collection

In this research we will refer to our experiment numerous times. This experiment consists of an actual interaction with our chatbot and a questionnaire either before and/or after. The interaction with the chatbot is based on our 2x2 matrix, in figure 2, which explains how we have a placebo-group which will receive a chatbot of which the credibility was not disclosed and another group will receive a chatbot of which the credibility was disclosed. Said matrix is described below. In our research we focus on the nature of the evaluation, meaning the positive and negative part of the matrix. The reason for this is that we believe the nature of the evaluation has the most impact on the perceived credibility of our chatbot.

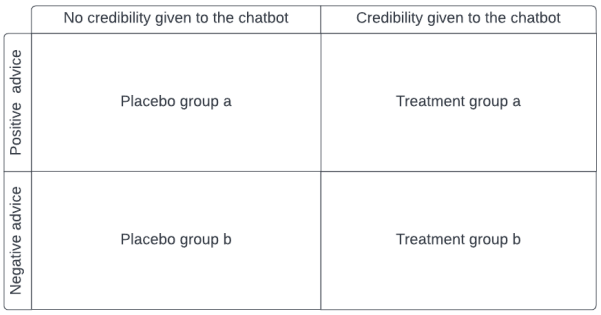


Figure 2. 2x2 matrix experiment

This evaluation will be either positive or negative. Before the interaction with the chatbot we will ask the participants to fill out a questionnaire that will ask several questions about the evaluation itself and the chatbot which can be found in appendix B. The questions depend on the thesis that they refer to, as we have several thesis' based on this experiment different questions will be assigned to different thesis'. In our case we will mostly ask the participant about questions regarding their own opinion on their idea, what they think of chatbots in general such as their trustworthiness and what they think of the evaluation the chatbot gave as well as their confidence and whether they would adopt the given advice. The answers to these questions are analyzed through the means of a Spearman's rank correlation test and a simple linear regression model. The reason for this instead of the conventional Pearson's linear correlation is that "Spearman's rank correlation also works well with ordinal and continuous data.'' (Zou et al., p. 2, 2003). Since the survey is conducted with questions comprising of Likert scale (1-5) answers will be of ordinal nature meaning we should utilize Spearman's rank correlation. "As for the regression methods we are using a simple linear regression model as it is meant for continuous variables.'' (Zou et al., p. 3, 2003). Consequently, our variables will be continuous after having created a dummy variable for each representative variable. This process allows for more reliable analysis

4.2 Selection and Sample

Regarding our sample we decided it would be best to take a sample of double the minimum amount of 30 participants. This with an exception of a few data entries which were invalid due to several reasons. One of which was the incorrect answer to the question "what type of evaluation have you received from the chatbot?''. An extra of 3 invalid entries were removed because of this. These entries were filtered out after comparing all entries to their actual evaluations, which is shown inn figure 3 and 4. In the end out of all entries we are left with 55 people to ensure a good representation of the population. The sample consists of people between the age of 21 and 60 and thus covers a realistic amount of differently aged people that all have had an encounter with a chatbot at least once. The reason we chose this many differently aged people is to prevent data from being skewed a certain way because of the lack of representation of the population in our sample as age represents a vital role in how one could have experienced an encounter with a chatbot.

What type of evaluation have you received from the chatbot? * Actual advice given Combined Crosstabulation

			Actual advice given Combined	
			Positive	Negative
What type of evaluation have you received from the chatbot?	Positive	Count	29	3
		% within Actual advice given Combined	100.0%	10.3%
	Negative	Count	0	26
		% within Actual advice given Combined	0.0%	89.7%
Total		Count	29	29
		% within Actual advice given Combined	100.0%	100.0%

Figure 3. Crosstab evaluation and actual evaluation pre-filter

What type of evaluation have you received from the chatbot? * Actual advice given Combined Crosstabulation

			Actual advice given Combined	
			Positive	Negative
What type of evaluation have you received from the chatbot?	Positive	Count	29	0
		% within Actual advice given Combined	100.0%	0.0%
	Negative	Count	0	26
		% within Actual advice given Combined	0.0%	100.0%
Total		Count	29	26
		% within Actual advice given Combined	100.0%	100.0%

Figure 4. Crosstab evaluation and actual evaluation post filter

4.3 Measurement and Data Analyses

The earlier mentioned questions will mostly be presented in a scale type format where we present statements per question which scale from very negative to very positive (Likert scale). This makes comparing answers between groups easier as the Likert scale is a format which allows you to answer with a number from 1 to 5 with as mentioned before strongly disagree at the first end and strongly agree at the second end. Adding these answers together allows us to reach a total score which then indicates the person's overall result in a certain category like "Familiarity with AI''. These scores are analyzed based on a factor analysis to analyze possible overlap between components. Appendix F shows the rotated component matrix of the factor analysis of our variables. Our survey questions are placed into 4 separate components or "factors''. These factors also match our predetermined grouping of questions in appendix A, which were already validated by other sources, indicating a good pre-selection of questions. Appendix F shows that no factor loading is below 0.5 meaning that with our absolute value set to 0.5 that no item was deleted. In fact our survey questions all have a factor loading of at least 0.7 or higher apart from one. This proves that most factors extract sufficient variance from their variables.

In addition appendix F shows the Cronbach's Alpha for each factor with each of them having a value of at least 0.7 indicating a good internal consistency. The above is true for every factor except one. The Cronbach's Alpha for our confidence-factor is 0.518 meaning that it has low internal consistency. We explain this by the fact that this factor has a rather lackluster amount of questions that make up said factor. Had there been more questions created within the factor of confidence we could have generated a higher internal consistency.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Willingness_to_adopt_AI	.126	55	.030	.920	55	.001
What type of evaluation have you received from the chatbot?	.353	55	.000	.636	55	.000
Pre_Trust_AI	.183	55	.000	.950	55	.024
Pre_Familiarity_AI	.107	55	.174	.967	55	.136
Pre_Confidence_AI	.176	55	.000	.922	55	.002

Figure 5. Shapiro-Wilk test of normality

In order to proceed with our analysis we have also conducted a Shapiro-Wilk test to investigate whether the variables are normally distributed or not. The results of this test will decide upon what tests we will use in our next chapter "Results and Analyses''. The Shapiro-Wilk test shows as seen in table 3 that apart from our variable "Familiarity with AI'' all variables have a respective p-value of < 0.05 meaning that we are able to reject the null hypothesis of the Shapiro-Wilk test. This indicates that the variables are non-normally distributed. Despite the fact that not all variables are normally distributed, for the sake of accuracy and consistency within this experiment we will conduct our further analyses based on non-normally distributed variables.

5. RESULTS AND ANALYSES

5.1 Testing hypotheses

As mentioned earlier in this thesis the goal of our research has been to point out the association between the nature of the evaluation and the perceived credibility of a chatbot along with the effects thereof. In order to identify these associations and effects we have decided to formulate several hypotheses with the expectation that both preliminary opinion variables and the evaluation will be associated with perceived credibility.

5.1.1 Participants' confidence is associated with perceived credibility

Due to the earlier mentioned non-normal nature of our variables we have conducted a non-parametric test to determine the association between participant's confidence and the perceived credibility. The test of choice in this case was Spearman's rank-order correlation test. The results as seen in figure 4 indicate a weak positive relation which is non-significant due to the p-level being far above 0.05 (0.655>0.05).

Correlations

			Pre_Confidence_AI	Willingness_to_adopt_AI
Spearman's rho	Pre_Confidence_AI	Correlation Coefficient	1.000	.062
		Sig. (2-tailed)	.	.655
		N	55	55
	Willingness_to_adopt_AI	Correlation Coefficient	.062	1.000
		Sig. (2-tailed)	.655	.
		N	55	55

Figure 6. Spearman's rho participant's confidence

In addition to measuring the association between our two variables we are also determining the effect between the variables. This is done through a linear regression model. In figure 5 below we can see the model split up into the constant value and our independent variable. Both values are assumed to have a significant effect when the null hypothesis is rejected with the p-value being lower than 0.05. In this case we are not able to reject the null hypothesis due to our p-value being 0.314>0.05 and the low unstandardized beta of 0.390. Thus a participant's confidence does not have a significant effect on the perceived credibility or also known as willingness to adopt AI, in our case. This means we will reject our initial hypothesis H1.1.

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	5.325	2.575		2.068	.044
	Pre_Confidence_AI	.350	.345	.138	1.015	.314

Figure 7. Linear regression model participant's confidence

5.1.2 Participant's familiarity with AI is associated with perceived credibility

With this hypothesis along with the other hypotheses we are using Spearman's rho rank-order correlation model to determine the correlation between our variables and the linear regression model to determine the regression. The correlation between our variables is as follows in figure 6.

Correlations

			Willingness_to_adopt_AI	Pre_Familiarity_AI
Spearman's rho	Willingness_to_adopt_AI	Correlation Coefficient	1.000	.045
		Sig. (2-tailed)	.	.747
		N	55	55
	Pre_Familiarity_AI	Correlation Coefficient	.045	1.000
		Sig. (2-tailed)	.747	.
		N	55	55

Figure 8. Spearman's rho participant's familiarity with AI

From figure 6 we can deduce a weak positive association between our two variables which is non-significant due to our p-value being higher than 0.05. As for regression between these two variables we can speak of a weak positive but non-significant effect, in figure 7, which is indicated by the high p-value of 0.556>0.05 and the low unstandardized beta of 0.078. This results in no effect between participant's familiarity with AI and Willingness to adopt AI. Thus we are able to reject hypothesis 1.2.

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	6.542	2.333		2.804	.007
	Pre_Familiarity_AI	.078	.132	.081	.592	.556

Figure 9. Linear regression model participant's familiarity with AI

5.1.3 Participant's trust in AI is associated with perceived credibility

In relation to this hypothesis we conducted a Spearman's rank-order correlation test to find the association between these two variables. In figure 8 becomes apparent through weak positive but non-significant correlation that there no is association of significance between trust in AI and Willingness to adopt AI.

Correlations

			Willingness_to_adopt_AI	Pre_Trust_AI
Spearman's rho	Willingness_to_adopt_AI	Correlation Coefficient	1.000	.160
		Sig. (2-tailed)	.	.243
		N	55	55
	Pre_Trust_AI	Correlation Coefficient	.160	1.000
		Sig. (2-tailed)	.243	.
		N	55	55

Figure 10. Spearman's rho participant's trust in AI

As for regression which is measure with the linear regression model we can speak of a weak positive and non-significant effect between the independent variable trust in AI and the dependent variable willingness to adopt AI. Figure 9 shows the unstandardized beta of 0.256 and the p-value of 0.215. As 0.215>0.05 we are not able to reject the null hypothesis and we can conclude there is no significant effect between participant's

trust in AI and the willingness to adopt AI. This means we are able to reject our initial hypothesis 1.3.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	5.308	2.118		2.506	.015
	Pre_Trust_AI	.256	.204	.170	1.254	.215

Figure 11. Linear regression model participant's trust in AI

5.1.4 Type of evaluation is associated with perceived credibility

As for the last association we use Spearman's rank-order correlation model to assess the association between the variables type of evaluation and willingness to adopt AI. Figure 10 shows a moderately negative significant association between the two variables as the correlation coefficient is -0.304 and the p-value is 0.024 which is lower than our alpha of 0.05. This means we are able to reject the null hypothesis and consider the association between these two variables significant.

Correlations

			Willingness_to_adopt_AI	What type of evaluation have you received from the chatbot?
Spearman's rho	Willingness_to_adopt_AI	Correlation Coefficient	1.000	-.304*
		Sig. (2-tailed)	.	.024
		N	55	55
	What type of evaluation have you received from the chatbot?	Correlation Coefficient	-.304*	1.000
		Sig. (2-tailed)	.024	.
		N	55	55

Figure 12. Spearman's rho type of evaluation

Regarding the linear regression model we have used in figure 11 we can see that our independent variable has a p-value of 0.021 which is lower than our alpha and a unstandardized beta of -2.273. This translates into a rejected null hypothesis and thus a strong negative effect between evaluation type and willingness to adopt AI. This means we are able to accept the hypothesis of H2.

6. DISCUSSION AND IMPLICATION

6.1 Findings and Implications

The objective of this thesis was to research the association between the nature of the evaluation and the perceived credibility or in the context of this experiment willingness to adopt AI. Research from (von Walter et al., 2022) shows that from a cultural perspective there is naturally, within some environments, a resistance to the further introduction of AI. This led us to believe there is an association between our preliminary opinion of AI and the willingness to adopt it. This however was not the case which is shown in the results and analyses section of this thesis through the rejection of multiple hypotheses.

In order to accurately link perceived credibility to our independent variables it is important to understand the underlying meaning of perceived credibility. The past research of (Shin, 2022) shows that credibility is in fact distributed among several factors, one of which is trust. One would expect that trust would then affect perceived credibility but our research shows quite the contrary with both correlation and regression showing no association and effect between the two variables respectively.

Furthermore if we look at research from (Harvey, 1997) we realize that it is not only trust and familiarity that decided the preliminary opinion of a chatbot but that confidence in one's own idea also plays a very large role. As is stated in their research "One of the earliest explanations of overconfidence was based on the notion that people think of evidence in favor of their chosen answer (and against the alternatives) more easily than they think of evidence in favor of the alternatives (and against their chosen answer).'' In our experiment also the contrary is assumed to be quite possible which is explained by (Lewandowsky et al., 2000) wherein it is stated that a person with low confidence in their idea might lean towards automation as the main decision-making agent. However even this assumed to be crucial variable in our research appeared to be non-significant in association with our dependent variable.

According to (Dietvorst et al., 2018) there is a preponderance of evidence demonstrating the superiority of algorithmic judgment in decision making situations. This indicates that the average person is aware of the outperforming that the average AI is capable of. Nevertheless studies from said source suggest that the majority of participants chose the less performing human anyway. This study shows how the evaluations of AI or in this case chatbots can influence perceived credibility as participants think AI are less believable and thus less credible in their evaluations. To some extent our research was able to confirm this suspicion. We detected an association between type of evaluation and willingness to adopt AI as well as an effect between the two. All in all the acceptance of our main hypothesis means markets could apply this knowledge of opinions towards chatbots in their business models to increase engagement between chatbot and person.

6.2 Relevance

6.2.1 Academic Relevance

As already mentioned earlier in this thesis, artificial intelligence is becoming an ever growing part of our lives. This makes it more and more important that we put extra efforts into researching and practicing with the optimal way to integrate AI into our business models and our personal lives. In specific how to integrate chatbots.

In addition to this, research shows that "scientific literature on the impact of chatbots, in general, is lacking, be it considering individuals, certain groups, or society at large'' (Følstad et al., p. 9, 2021), making it all the more important that we keep researching AI as we can see the ever increasing use of such AI already in for instance the marketing field.

In our case making the process behind the evaluating of a chatbot's credibility clearer will enable us to see how we interact with artificial intelligence taking over ''human'' tasks. This puts it into perspective how we are able to integrate chatbots into processes that require a rather large amount of human input. A good example is Facebook. They were one of the first of their kind to use chatbots in their marketing strategy in an effective way. They implemented this on their messenger platform. The reason this was so effective and at all possible is because they were advertisements generated by their chatbots on their own site. This is the same site through which they are able and allowed to retrieved certain data on their users on, which allows them to personalize these advertisements.

In the marketing section specifically there are already chatbots making personalized advertisements. Next to this we see most recently chatbots also functioning as a shopping tool helping the customer pick products that they would like. For research in the future it would be most interesting to see how influential this topic can be in the marketing field. This because, as mentioned before, before being able to create a personalized chatbot experience the participant needs to be understood. Else it will be regarded as a non-reliable and thus non-credible chatbot.

6.2.2 Practical Relevance

Other than benefiting the academic research in the field of chatbots sections of this thesis could also benefit the business sector. We could expect, if our chatbot catches on and is adopted, an increase in the number of valuable innovations that pass through the innovation funnel. By being able to create a credible chatbot that generates reliable evaluations of innovations a smaller number bad quality innovations make it to the market while a higher number of good quality does.

In addition to this we can expect down the road an increase in efficiency across the board. When implementing a chatbot that knows how to interact with its users and knows how to recognize patterns in purchasing much like a "Thinking AI'' (Huang & Rust, p. 2, 2021) and is easy to use more businesses and their customers will be inclined to use them. This could replace the need for possible slower and lower quality human inputs, eventually increasing efficiency.

Lastly, chatbots are durable and in an optimal situation reliable as well. What is meant by this is that when using chatbots to evaluate business ideas the chatbot can evaluate at a continuous rate without the need to take breaks to think as it answers everything objectively and reliably. When being able to reduce the need for human input said input can be used elsewhere in complex situations that a chatbot cannot tackle allowing for a better workforce distribution in most companies.

7. CONCLUSION

In this thesis we looked at several associations between our independent variables and our dependent variable. We did this to be able to answer our research question: "Does the nature of a chatbot's evaluation have an impact on its perceived credibility?'' and our sub question: "Does the user's opinion on chatbots before interacting with one influence perceived credibility?''. In order to be able to answer these questions in an organized manner we stated 4 different hypothesis. Hypotheses 1 to 1.3 resulted, against our expectation, into a rejection while our main and last hypothesis was accepted. With the first three hypotheses rejected we are able to answer our sub question. In the context of our research we found that a user's opinion before interacting with a chatbot does not influence perceived credibility.

We are able to answer our main research question through the acceptance of our fourth hypothesis. With this we can to some extent confidently answer that the nature of the evaluation indeed have an impact on its perceived credibility.

8. LIMITATIONS AND FUTURE RESEARCH

Through the use of several statistical tests we were able to determine that to some extent the "preliminary opinion'' is associated with the "perceived credibility'' but not to a significant extent. While our independent variable ''nature of evaluation'' was determined to have a significant association and effect on our dependent variable.

While this means we were able to accept hypothesis 2, our main hypothesis, this research does have its limitations. While we were able to determine associations and effects of our independent variables on our dependent variable we were not able to determine a significant effect of our independent variables on the ''preliminary opinion''-side. This would mean that there is no significant effect of preliminary opinions on the perceived credibility of our chatbot while (von Walter et al., 2022) in "The impact of lay beliefs about AI on adoption of algorithmic advice'' proves the exact opposite. This could possibly be the consequence of a low Cronbach's Alpha for the "Confidence'' survey questions in appendix F indicating a lower internal consistency leading to lesser reliable data.

We have identified a specific self-selection bias. In the context of our research what is meant with this is that our survey tends to attract people that are actually interested in the development of AI this could lead to a skewness in the results. In addition our answers could be inaccurate due to the believability of the survey. The survey is meant to seem like a chatbot, however if a participant recognizes it as a survey, post-evaluation questions about our ''chatbot'' are less reliable.

Lastly we have identified the lack of inclusion of a variable that takes account of the scenarios with and without credibility disclosure. The absence of this variable affects the trust in our chatbot and thus to some extent, although non-significantly, perceived credibility. The reason this variable was not included in the dataset was that it would alter the normality distribution by a drastic amount making the analysis of the other variables too extensive and difficult leading to possible errors. This decision does however lead to a lower level of reliability of the results.

From this paper becomes clear that due to the limited amount of previous research done in this area there is much to be learned and perfected as is already stated by (Følstad et al., 2021). Nevertheless, research towards opinions on chatbots is a good starting point. From this point future research could build further on this by looking at consumers needs specifically by

Implementing "empathy maps'' as described in (Illescas-Manzano et al., 2021). By looking at consumers needs instead of their opinion towards said chatbots we could be able to tailor a chatbot in a way that it perfectly markets the needs of such a consumer. This because an argument could be made for the negligibility of research on opinions towards chatbots as AI implementation in global markets seems to be more and more inevitable according to (IBM, 2022). This report shows that 35% of all companies globally reported already using AI in their business practices and another 42% reported exploring the implementation of AI. Though analyzing opinions on AI may be beneficial in understanding what it takes to increase adoption rates of AI, it may be more beneficial to look towards perfecting the implementation of AI as adoption seems inevitable.

9. ACKNOWLEDGEMENTS

A major thanks to my first and second supervisor Mr. Ehrenhard and Ms. Koefer for helping with the design of the experiment which is the main building block of this research. I also could not have undertake this journey without the help of Ms. Hormann in the technical sections of SPSS statistics.

Lastly I would be remiss in mentioning my colleagues in this thesis-group who have peer reviewed my work on several occasions and have helped me through several obstacles which proved to be difficult to tackle on my own.

10. BIBLIOGRAPHY

Castelo, N. (2019). *Blurring the Line Between Human and Machine: Marketing Artificial Intelligence*.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Fan, X., Oh, S., McNeese, M., Yen, J., Cuevas, H., Strater, L., & Endsley, M. R. (2008). The influence of agent reliability on trust in human-agent collaboration. *Proceedings of the 15th European Conference on Cognitive Ergonomics the Ergonomics of Cool Interaction - ECCE '08*, 1. https://doi.org/10.1145/1473018.1473028

Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, *103*(12), 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences*, *1*(2), 78–82. https://doi.org/https://doi.org/10.1016/S1364-6613(97)01014-0

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250. https://doi.org/10.1016/j.chb.2015.02.026

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

IBM. (2022). *IBM Global AI Adoption Index 2022*. https://www.ibm.com/watson/resources/ai-adoption

Illescas-Manzano, M. D., Vicente López, N., Afonso González, N., & Cristofol Rodríguez, C. (2021). Implementation of Chatbot in Online Commerce, and Open Innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(2), 125. https://doi.org/10.3390/joitmc7020125

Jiang, G., Liu, F., Liu, W., Liu, S., Chen, Y., & Xu, D. (2021). Effects of information quality on information adoption on social media review platforms: moderating role of perceived risk. *Data Science and Management*, *1*(1), 13–22. https://doi.org/10.1016/J.DSM.2021.02.004

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123. https://doi.org/10.1037/1076-898X.6.2.104

Mittal, A., Agrawal, A., Chouksey, A., Shriwas, R., & Agrawal, S. (2016). IJARCCE A Comparative Study of Chatbots and Humans A Comparative Study of Chatbots and Humans. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*. https://doi.org/10.17148/IJARCCE.2016.53253

Rawlinson, K. (2015, January 29). *Microsoft's Bill Gates insists AI is a threat*. BBC. https://www.bbc.com/news/31047780#:~:text=Humans%20should%20be%20worried%20about,strong%20for%20people%20to%20control.

Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI and Society*, *37*(1), 81–96. https://doi.org/10.1007/s00146-021-01158-4

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Tesla. (2022). *Future of driving*. https://www.tesla.com/autopilot

von Walter, B., Kremmel, D., & Jäger, B. (2022). The impact of lay beliefs about AI on adoption of algorithmic advice. *Marketing Letters*, *33*(1), 143–155. https://doi.org/10.1007/s11002-021-09589-1

Wairegi, A., Omino, M., & Rutenberg, I. (2021). AI in Africa : Framing AI through an African Lens. *Communication, Technologies et Développement*, *10*. https://doi.org/10.4000/ctd.4775

Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and Simple Linear Regression. *Radiology*, *227*(3), 617–628. https://doi.org/10.1148/radiol.2273011499

11. APPENDICES

Appendix A – Variables, items, and their corresponding sources

Appendix B – Pre-interaction survey

Appendix C – Task description + video

Appendix D – Questions asked by the chatbot

Appendix E – Post-interaction survey

Appendix F – Factor analysis rotated component matrix

Appendix A

Variable	*Item*	*Source*
Trust in Technologies	**My typical approach is to trust new technologies until they prove me that I shouldn't**	Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Computers in Human Behavior*, *118*, 106700. https://doi.org/10.1016/j.chb.2021.106700
	I generally give a technology the benefit of the doubt when I first use it	
	I usually trust a technology until it gives me a reason not to trust it	
Familiarity with AI and AI chatbots	**I am familiar with AI**	Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. Computers in Human Behavior, 115, 106607. https://doi.org/10.1016/j.chb.2020.106607
	I am familiar with AI chatbots	Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. Computers in Human Behavior, 118, 106700. https://doi.org/10.1016/j.chb.2021.106700
	I have much knowledge about AI chatbots	
	I am more familiar than the average person regarding AI chatbots	
	I know how to interact with AI chatbots	
Trust in AI algorithms and its advice	**I trust the recommendations by algorithms-driven services (chatbots, predictive personalization agents, virtual assistants, etc).**	Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551
	Recommended items through algorithmic processes are trustworthy.	
	I believe that the algorithm service results are reliable.	
Feelings about being judged by others when telling them about an idea you recently had.	**If I needed to, I would feel at ease when presenting an idea to others**	Siemon, D. (2022). Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration. *Behaviour & Information Technology, 1–19.* *https://doi.org/10.1080/0144929x.2021.2023638*
	I tend to worry about being judged by others when presenting an idea	
Confidence in ability to formulate ideas	**I'm confident in my ability to formulate high quality ideas.**	Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior, 127, 107018.* *https://doi.org/10.1016/j.chb.2021.107018*
	I don't believe that my confidence in my high-quality idea will be affected by a machine response.	

Trust in the AI chatbot	**I trust the advice the chatbot provided me with.**	Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551
	I find the chatbot's advice to be trustworthy.	
	I believe that the chatbot's advice is reliable.	
	I believe that the chatbot was credible during our conversation.	Toader, D. C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D., & Rădulescu, A. T. (2019). The Effect of Social Presence and Chatbot Errors on Trust. *Sustainability, 12(1), 256. https://doi.org/10.3390/su12010256*
Advice utilization	**I am willing to let this chatbot assist me in deciding whether or not to develop my business idea**	Benbasat, I., & Wang, W. (2005). Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems, 6(3), 72–101. https://doi.org/10.17705/1jais.00065*
	I am willing to use this chatbot as an aid to help with developing my business idea.	
	I am willing to use this chatbot's advice recommendations.	
Perceived usefulness of the chatbot	**The evaluation provided by the chatbot would be useful to me.**	Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13(3), 319. https://doi.org/10.2307/249008*
	The evaluation provided by the chatbot would help me to feel at ease when presenting my idea to others.	Siemon, D. (2022). Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration. *Behaviour & Information Technology, 1–19. https://doi.org/10.1080/0144929x.2021.2023638*
	The evaluation provided by the chatbot would help me to worry less about being judged by others when I present my idea.	
	The evaluation provided by the chatbot would help me to be more creative.	
	The evaluation provided by the chatbot would help me to feel encouraged to present my idea to others.	Siemsen, E., Roth, A. V., Balasubramanian, S., & Anand, G. (2009). The Influence of Psychological Safety and Confidence in Knowledge on Employee Knowledge Sharing. *Manufacturing & Service Operations Management, 11(3), 429–447. https://doi.org/10.1287/msom.1080.0233*
	The evaluation provided by the chatbot would help me to have more confidence in my idea.	

Appendix B

What is your year of birth?

Where do you come from?

What gender do you identify as?

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ Prefer not to say

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
My typical approach is to trust new technologies until they prove me that I shouldn't	○	○	○	○	○
I generally give a technology the benefit of the doubt when I first use it	○	○	○	○	○
I usually trust a technology until it gives me a reason not to trust it	○	○	○	○	○

According to Duan et al., (2019), Artificial Intelligence refers to "the ability of a machine to learn from experience, adjust to new inputs and perform human-like tasks". And a chatbot is a tool that is designed to mimic human-like conversations (Kushwaha et al., 2021). The purpose of this new source of advice (Logg et al., 2019) is to provide people with information they can use in their decision-making (Klaus & Zaichkowsky, 2020 as cited in Kushwaha et al., 2021).

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I am familiar with Artificial Intelligence (AI)	○	○	○	○	○
I am familiar with AI chatbots	○	○	○	○	○
I have much knowledge about AI chatbots	○	○	○	○	○
I am more familiar than the average person regarding AI chatbots	○	○	○	○	○
I know how to interact with AI chatbots	○	○	○	○	○

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I trust the recommendations by algorithms-driven services (chatbots, predictive personalization agents, virtual assistants, etc)	○	○	○	○	○
Recommended items through algorithmic processes are trustworthy.	○	○	○	○	○
I believe that the algorithm service results are reliable.	○	○	○	○	○

Please indicate your level of agreement with the following statements.

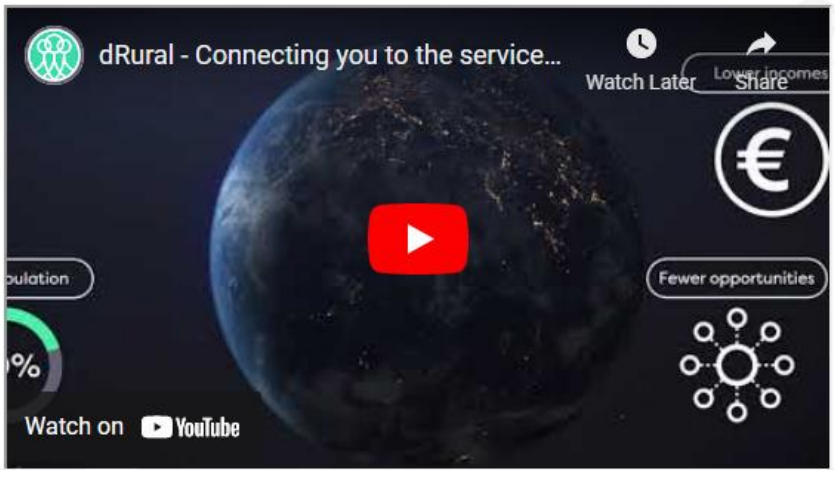
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
If I needed to, I would feel at ease when presenting an idea to others	○	○	○	○	○
I tend to worry about being judged by others when presenting an idea	○	○	○	○	○

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I'm confident in my ability to formulate high quality ideas.	○	○	○	○	○
I don't believe that my confidence in my high quality idea will be affected by a machine response.	○	○	○	○	○

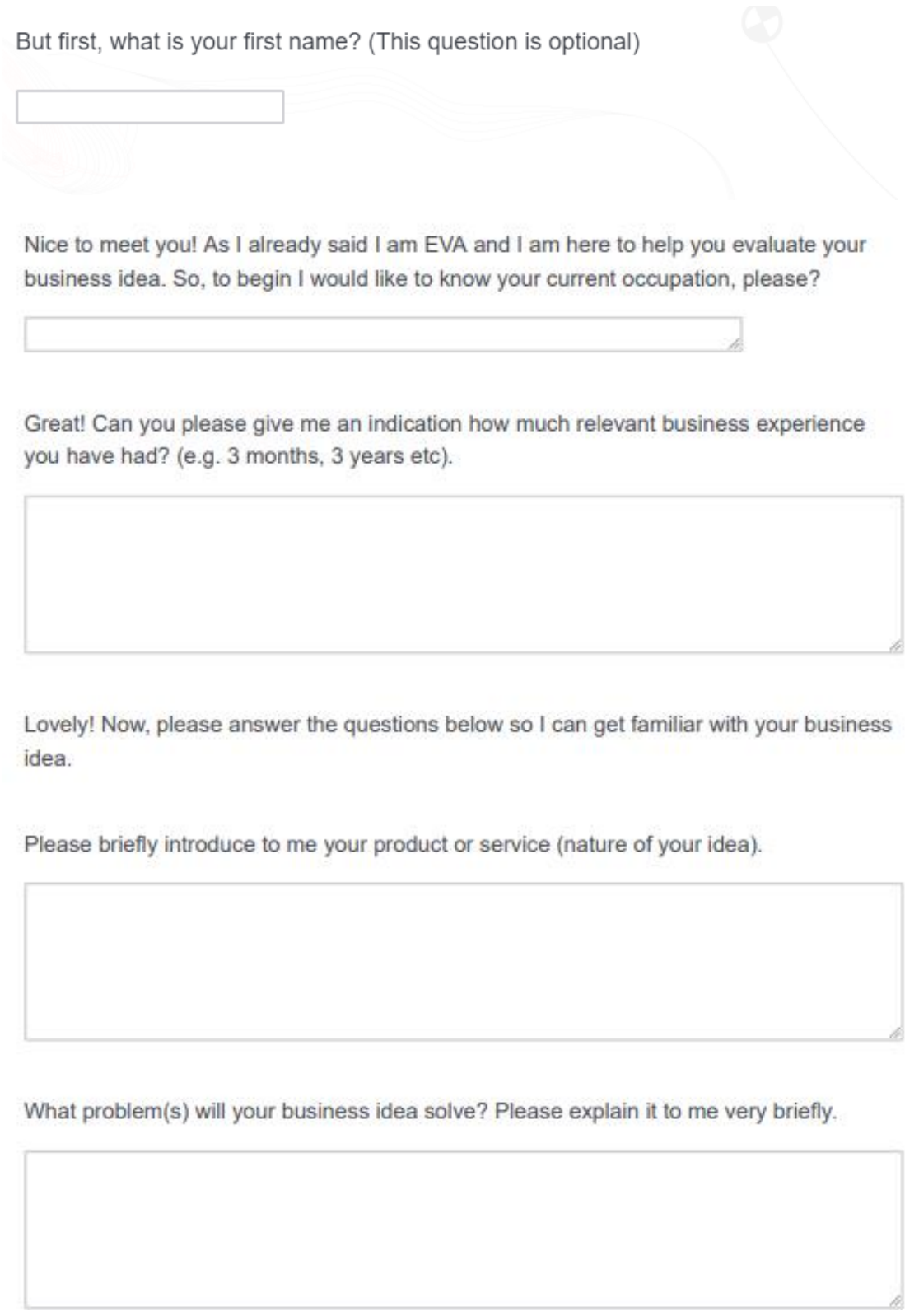
Appendix C

In this study you are asked to imagine that you would like to start a new business on a digital services platform. Below, you will watch a video about a new digital platform for rural services in Europe. Please watch the video carefully and think of potential business ideas for services that could be offered via this digital services platform – try to be as innovative and creative as you can. Next, please choose the business idea that in your opinion would be the most viable, which means that there is a clear customer base that is willing to pay for your service. After the press release, a chatbot will help you evaluate various aspects of your business idea and then provide an overall assessment of the quality of your business idea.



https://www.youtube.com/watch?v=RCQc24UYfeI

Appendix D


But first, what is your first name? (This question is optional)
Nice to meet you! As I already said I am EVA and I am here to help you evaluate your business idea. So, to begin I would like to know your current occupation, please?
Great! Can you please give me an indication how much relevant business experience you have had? (e.g. 3 months, 3 years etc).
Lovely! Now, please answer the questions below so I can get familiar with your business idea.
Please briefly introduce to me your product or service (nature of your idea).
What problem(s) will your business idea solve? Please explain it to me very briefly.

Thank you for all your input so far! To better understand your idea can you please briefly describe who your target customers will be.

Great! Why do you believe your target customers are interested to buy your product/service?

In what country or region are you planning to establish/sell your product or service?

How will your business idea generate revenue? Please briefly describe your strategy, e.g. your pricing strategy

Fantastic! You already provided a lot of informative details about your business idea which will help me to compare it with existing databases.

My final question: What do you think will be the competitive advantage of your business idea in your target region? (for example, the price, uniqueness of the product/service, high social or environmental impact)

Amazing! Thank you! Now please give me a minute to make the evaluation and I will get back to you with my advice! You can move forward.

Appendix E

What type of evaluation have you received from the chatbot?

- ○ Positive
- ○ Negative

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I trust the advice the chatbot provided me with.	○	○	○	○	○
I find the chatbot's advice to be trustworthy.	○	○	○	○	○
I believe that the chatbot's advice is reliable.	○	○	○	○	○
I believe that the online agent was credible during our conversation.	○	○	○	○	○

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I am willing to let this chatbot assist me in deciding whether or not to develop my business idea	○	○	○	○	○
I am willing to use this chatbot as an aid to help with developing my business idea.	○	○	○	○	○
I am willing to use this chatbot's advice recommendations.	○	○	○	○	○

Please indicate your level of agreement with the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The evaluation provided by the chatbot would be useful to me.	○	○	○	○	○
The evaluation provided by the chatbot would help me to feel at ease when presenting my idea to others.	○	○	○	○	○
The evaluation provided by the chatbot would help me to worry less about being judged by others when I present my idea.	○	○	○	○	○
The evaluation provided by the chatbot would help me to be more creative.	○	○	○	○	○
The evaluation provided by the chatbot would help me to feel encouraged to present my idea to others.	○	○	○	○	○
The evaluation provided by the chatbot would help me to have more confidence in my idea.	○	○	○	○	○

Appendix F

Rotated Component Matrix[a]

	Component				Cronbach's Alpha
	1	2	3	4	
My typical approach is to trust new technologies until they prove me that I shouldn't			.876		
I generally give a technology the benefit of the doubt when I first use it			.607		0.768
I usually trust a technology until it gives me a reason not to trust it			.926		
I am familiar with Artificial Intelligence (AI)	.702				
I am familiar with AI chatbots	.775				
I have much knowledge about AI chatbots	.794				0.835
I am more familiar than the average person regarding AI chatbots	.780				
I know how to interact with AI chatbots	.792				
I'm confident in my ability to formulate high quality ideas.				.830	
I don't believe that my confidence in my high quality idea will be affected by a machine response.				.741	0.518
I am willing to let this chatbot assist me in deciding whether or not to develop my business idea		.946			
I am willing to use this chatbot as an aid to help with developing my business idea.		.943			0.936
I am willing to use this chatbot's advice recommendations.		.914			

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 4 iterations.