User Study on the Trustworthiness, Usability and Explainability of Intent-based and Large Language Model-based Career Planning Conversational Agents

Thorsten Zylowski
Institute for Intelligent Interaction and Immersive Experience
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
Faculty of Business, Economics and Social Sciences
University of Hohenheim
Stuttgart, Baden-Wurttemberg
Germany
thorsten.zylowski@uni-hohenheim.de

Nathalia Sautchuk-Patricio
Institute for Intelligent Interaction and Immersive Experience
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
nathalia.sautchuk_patricio@h-ka.de

Wladimir Hettmann
Institute for Intelligent Interaction and Immersive Experience
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
wladimir.hettmann@h-ka.de

Katharina Anderer
Institute for Intelligent Interaction and Immersive Experience
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
Karlsruhe Institute of Technology
Karlsruhe, Baden-Wurttemberg
Germany
katharina.anderer@kit.de

Karl Fischer
Institute for Intelligent Interaction and Immersive Experience
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
Karlsruhe University of Education
Karlsruhe, Baden-Wurttemberg
Germany
karl.fischer@ph-karlsruhe.de

Matthias Wölfel
Institute for Intelligent Interaction and Immersive Experience
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
Faculty of Business, Economics and Social Sciences
University of Hohenheim
Stuttgart, Baden-Wurttemberg
Germany
matthias.woelfel@h-ka.de

Peter Henning
Karlsruhe University of Applied Sciences
Karlsruhe, Baden-Wurttemberg
Germany
peter.henning@h-ka.de

Abstract

Choosing a career and educational path is a challenging decision for young people. Career planning conversational agents (CAs) can assist by identifying suitable occupations and educational paths. Trustworthiness is an important dimension for the acceptance of a career planning CA and is influenced by several factors. We conducted a user study with n=114 participants across three schools in Germany to explore the trustworthiness of different career planning CAs. We examined the correlation between trustworthiness and perceived competence, autonomy, and social relatedness from self-determination theory (SDT), as well as the explainability of interactions and several usability dimensions of the assistants. These dimensions included the ability to guide the conversation, onboarding quality, error tolerance, and information relevance. We tested three different variants of the career planning assistant: a form-based assistant, an intent-based CA, and a large language model (LLM)-based CA. The results showed that the LLM-based CA was on

ICETC 2024, September 18–21, 2024, Porto, Portugal

ACM ISBN 979-8-4007-1781-9/24/09
https://doi.org/10.1145/3702163.3702409

average significantly more trustworthy and was perceived as more explainable than the intent-based CA. Key trust factors included conversation flexibility, chatbot credibility, intent recognition, and maintenance of a secure conversation. Additionally, perceived autonomy was crucial for trust across all types of assistants and perceived relatedness for the two CAs. Our findings highlight key areas essential for developing trustworthy CAs.

CCS Concepts

• **Applied computing** → Education; • **Human-centered computing** → User studies; Natural language interfaces.

Keywords

Trustworthy Artificial Intelligence, Self-Determination Theory, Usability, Explainable Artificial Intelligence, Career Planning Conversational Agents

ACM Reference Format:

Thorsten Zylowski, Nathalia Sautchuk-Patricio, Wladimir Hettmann, Katharina Anderer, Karl Fischer, Matthias Wölfel, and Peter Henning. 2024. User Study on the Trustworthiness, Usability and Explainability of Intent-based and Large Language Model-based Career Planning Conversational Agents. In *2024 the 16th International Conference on Education Technology and Computers (ICETC) (ICETC 2024), September 18–21, 2024, Porto, Portugal*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3702163.3702409

1 Introduction

Deciding on a professional career and the associated educational path is an important and difficult decision for young people. It has far-reaching consequences for their entire life and should be well thought out. Once an educational path has been taken, it is anything but trivial to take a different direction. Career decisions are made even more difficult by the fact that young people do not yet know exactly what they would like to do in the future. It is possible that some individuals may already be aware of their future career path, although they may hold a misconception about it. There are also external influencing factors such as parental expectations, advice from friends or role models experienced in the media. When young people try to gather information about their future profession, they are overwhelmed by the immense range of offers.

One solution could be the use of career planning conversational agents (CAs), i.e., text-based chatbots, that help to identify the best occupation and associated educational path. Career planning CAs could have access to potentially a lot of career data, such as occupations, degree programs, apprenticeships, etc., and could navigate a student through the many pathways of the educational system.

CAs are in general perceived less judgmental than a human interlocutor. This increases the willingness to disclose towards a CA [1]. It further makes a conversation about future educational paths possible that parents, for instance, will not understand. Additionally, when the interaction with a CA feels anonymous, this leads to increased self-disclosure [2]. Regarding career planning, students are encouraged to engage in open dialogue with the CA, exploring fundamental values and ideas, leading to better recommendations for possible careers and educational paths. In addition, a social connection can be created between the human and the chatbot, which has a positive effect on trust [3].

CAs are available in various forms, each with their own advantages and disadvantages. Intent-based chatbots are often based on a machine learning (ML) model that classifies the user input according to a predefined set of user intents and selects the most suitable dialog based on this [4]. The answer of the CA is usually not generated but selected from a set of predefined answers. This type of CA permits a highly structured dialogue and a high degree of control over the output. The major disadvantage is the lack of flexibility in the dialog. If a user makes an input that is not recognized or incorrectly recognized, the conversation usually either stops or a fallback mechanism is activated. With intent-based CAs, it is not possible to have an open conversation, nor can a user response be actively deepened by the chatbot with follow-up questions.

CAs based on a large language model (LLM) offer a solution to this problem. They typically do not use an intent classification component. Instead, the response is generated directly from the input with the help of an LLM. The behavior of the CA can be controlled and adapted by prompt engineering [5]. The strengths of LLM-based CAs are manifold. The conversation is error-tolerant, as in principle a response is generated for every input. Open conversation is also possible, as LLMs have been trained on a huge amount of textual data and can therefore pick up a lot of information from the real world and generate their responses accordingly. It is even possible to change the behavior and personality of the LLM using suitable prompts [6], which makes it possible to tailor responses to the individual needs of the user. Another characteristic is that LLMs try to explain and justify generated answers which makes the conversation even more human-like. Due to this resemblance, we pose the hypothesis that an LLM-based CA is perceived as more trustworthy than an intent-based CA.

The first impression of the CA is of particular importance to convey trustworthiness [7], as the CA is only consulted once or at most a few times by a person. Trust in this case is not built up over a longer period of time and characteristics that are immediately perceptible are of high importance. The work of van der Werff et al. [8] provides a model that describes initial trust in a human-human workplace relationship. In this model, initial trust is determined by three factors:

- The trust propensity of the trustor,
- The perceived trustworthiness of the trustee, and
- The motivation of the trustor to trust the trustee.

The motivation to trust someone is described by the attributes competence, autonomy and relatedness from self-determination theory (SDT) [9]. *Competence* describes how competent people perceive themselves which is related to mastery. *Autonomy* describes the perception of not being dependent on something or someone else and how independent decisions can be made. *Relatedness* is the perception of being connected with others and is related to social support. In the model from van der Werff et al. [8], the motivation to trust someone depends on how competent, autonomous and socially related a person feels towards another person. Applied to the interaction with a CA, this could mean that if a person feels competent in dealing with the CA, can behave autonomously and establishes a social connection with the CA, that person would have a higher motivation to trust the CA. We investigated the role of SDT attributes in the different CAs and how they relate to trust.

For the development of a trustworthy CA the usability is fundamental. Therefore, we conducted comparisons on various usability characteristics including the perceived level of guidance, the quality of the onboarding process, error tolerance, and the relevance of the information provided in the context of a user study. We further investigated the correlations of usability characteristics and trustworthiness of the CA to identify the key areas for the implementation of trustworthy CAs.

Trust is often hindered by the lack of transparency in the decision-making process of artificial intelligence (AI) systems, often referred to as a black box. To address this issue and establish trust, methods are being developed within the field of explainable AI (XAI) to provide insights into how AI systems arrive at their decisions, enabling users to better understand their behavior. LLM-generated responses often contain justifying and explanatory phrases that resemble human reasoning, which is why we hypothesize that LLM-based CAs are perceived as more explainable than intent-based CAs.

2 Related Work

Several career planning CAs have been developed to assist users in making informed educational and vocational decisions, leveraging various technologies and theoretical frameworks. Chatbot E-Orientation utilizes John Holland's RIASEC typology, analyzing the answers to questions categorized using AI into six personality types to guide students entering the job market [10] [11]. D'Silva's chatbot offers psychometric tests (the "big 5 test" and the RIASEC) and emotional state assessments, providing personalized job recommendations and skill-building resources recorded in an E-portfolio [12]. Nair's chatbot, based on the Rasa framework [13], predicts user interests and recommends courses and colleges through a psychometric engine [14]. The Intelligent Career Counseling Chatbot (ICCC), designed for 10th and 12th-grade students, utilizes various algorithms to provide career guidance and educational support to the ones interested in Computer Science and Information Technology, focusing on a very user-friendly interaction [15]. In the same direction, ITCareerBot integrates a context-aware knowledge base and a recommendation method to offer personalized career advice, tailored learning paths, and course resources, enhancing users' career development and educational journeys in the IT area [16]. Talib's team implemented an AI-oriented career guidance system using LLM to offer personalized advice based on academic performance, interests, and career aspirations. Some initial tests were done with Moroccan students and showed high user engagement, positive feedback, and high satisfaction [17]. Sharma et. al. presents an AI-based deep learning chatbot for career and personal mentorship that offers anonymous, tailored guidance to students but requires ongoing improvements due to its dependency on small, pre-defined datasets and lack of emotional intelligence [18]. From the papers mentioned, it was noted that the majority of career planning CAs do not use LLMs, and most of these assistants have not been tested with users yet, especially on trustworthiness. A Wizard of Oz study has already shown that a high level of explainability in conversational agents has a positive effect on trust and acceptance, although in this case it was not studied specifically a career planning CA [19].

3 Methodology

In the following subsections, we delineate our methodology concerning the establishment of the experiment and the execution of measurements utilizing structured questionnaires focused on behavioral observations and self-assessment.

3.1 Experimental Design

A between-subject design, where participants were randomly assigned to one of three groups, was developed. The reference assistant is form-based, yet a typical format for career planning. The second and third assistant are CAs, the second intent-based and the third LLM-based. The three assistants have the task of collecting information about a participant's future career ideas, interests, and skills. Specific characteristics for the assistants are described in Table 1.

All assistants follow the same basic flow of interaction. After a welcome, the participants are asked whether they already have an idea of what a future job might look like. If so, they are asked to describe it in as much detail as possible. If not, they are guided through an assessment process to talk about their strengths and weaknesses. This assessment is based on questions inspired by the RIASEC assessment. In both cases, a user profile is compiled from the interaction. The user profile is then used in the same way by the three assistants to provide two occupation recommendations. The participant can either select one of the occupation recommendations or indicate that none of the recommendations are suitable. The choice is made by clicking on buttons. If none of the suggestions are considered suitable, the participant is asked further questions relating to their skills. The set of skill questions for the control group and for the first treatment group were fixed and set beforehand, while for the second treatment group they were generated in real-time by the assistant. The user profile is then updated with the answers made in response to the skill questions. Based on this information, two new occupations were proposed. Upon selecting a recommended occupation, participants have the option to choose a preferred occupational location. Subsequently, one or more educational pathways are calculated and displayed, outlining actual educational steps at specific universities or general apprenticeships in Germany. The participants completed a questionnaire after the experiment. The questionnaire is described in section 3.3.

The form-based assistant collects data through a form, while both the intent-based CA and the LLM-based CA use a chat, leading to conversational interfaces for the latter two. Error tolerance is high for the form-based assistant as the data collection flow is controlled by buttons, low for the intent-based CA as it uses fallback mechanisms when an intent cannot be determined, and high for the LLM-based CA because it always generates a response. The form-based assistant does not provide dynamic responses, while the intent-based CA offers predefined responses, and the LLM-based CA generates responses. Only the LLM-based CA supports open conversations due to its response generation capabilities. For the

Table 1: Summarized view of the implemented characteristics across the CAs

Characteristic	Form-based Assistant	Intent-based CA	LLM-based CA
Data collection	Form (mandatory fields)	Chat (intents and slots)	Chat (Named Entity Recognition)
Conversational interface	No	Yes	Yes
Error tolerance	High (implicit)	Low (fallback)	High (generative)
Response type	None	Predefined	Generative
Open conversation	No	No	Yes
Adaptive conversation	No	No	Yes
Interests and needs assessment	Predefined RIASEC questions	Predefined RIASEC questions	Guiding RIASEC prompts
Skill Assessment	Predefined skill questions	Predefined skill questions	Guiding skill prompts
End of Conversation	Form	Stories and rules	Fixed number of iterations
Interaction flow	Buttons	Chat input and buttons	Chat input and buttons
Assistant behavior	Form	Intents, stories and rules	Prompts
Actively driven Conversation	No	Yes	Yes
Guidance	Form	Stories and rules	Guiding prompts

same reason it is the only assistant that provides an adaptive conversation, where the response is generated considering the user input (and is sometimes repeated in the response). For assessing interests and needs, the form-based assistant and the intent-based CA use predefined RIASEC questions, while the LLM-based CA is prompted to generate RIASEC questions. The same procedure was chosen for the skills questions. The end of the interaction with the form-based assistant is determined by the end of the form, in the intent-based CA designed stories and rules define the end and in the LLM-based CA the interaction ends after a fixed number of iterations of user input and CA response. The flow of interaction in the form-based assistant is controlled by button clicks while in the intent-based and LLM-based additional chat inputs are made. The intent-based CA and the LLM-based CA actively drive conversations based on predefined dialogs for intent-based and specific prompts that guide the conversation in preferred directions for LLM-based CA. In the form-based assistant there is no active guidance. The structure of the form itself implicitly defines the guidance.

3.2 Behavioral Measurement

Various metrics were collected to investigate the interaction with the CAs. The average length of user input was calculated as the average number of characters entered per input. Every interaction with the assistant was time-stamped. From this the total time of interaction was calculated.

3.3 Self-Report Measurement

The participants completed a questionnaire after the experiment regarding participants' experiences and perceptions of the CA and covered trustworthiness, SDT, usability and explainability. All items in the questionnaire were assessed using a 5-point Likert scale with value "Strongly Disagree" on the left extremum and "Strongly Agree" on the right extremum without additional descriptions in between. All measured constructs are reported with Cronbach's α.

3.3.1 Trust and Self-Determination Theory. The questionnaire from [20] was used for the self-assessment of trust in the CA (e.g., "The assistant is trustworthy.", $\alpha = 0.64$). The technology-based experience of need satisfaction-interface questionnaire (TENS-Interface) [21] was employed to evaluate how the interface influenced participants' perceived competence (e.g., "I feel very capable and effective using the assistant.", $\alpha = 0.64$) and autonomy (e.g., "I feel under pressure from the assistant.", $\alpha = 0.56$), in using the assistant. Given that the interface did not generate any sense of social relatedness with others, this aspect was excluded from the questionnaire. Instead, the focus was on understanding how the interface altered the perception of the CA as a person, known as anthropomorphization. Anthropomorphization was assessed using a questionnaire adapted from [22] (e.g., "The conversation with the assistant was not artificial", $\alpha = 0.85$).

3.3.2 Usability. The usability of the assistants was assessed using the bot usability scale (BUS) [23] which was reduced to the dimensions *Ease to start a conversation* (Start), e.g., "It was clear how to start a conversation with the assistant" ($\alpha = 0.89$), *Expectation setting* (Expectation), e.g., "It is clear to me early on about what the assistant can do." ($\alpha = 0.68$), *Flexibility and communication effort* (Flexibility), e.g., "I had to rephrase my input multiple times for the assistant to be able to help me." ($\alpha = 0.66$), *Ability to maintain a themed discussion* (Maintain), e.g., "The assistant maintained a relevant conversation." ($\alpha = 0.78$), *Users' privacy and security* (Secure), e.g., "I believe that this assistant maintains my privacy." ($\alpha = 0.75$), *Recognition and facilitation of users' goal and intent* (Recognition), e.g., "The assistant was able to guide me to my goal." ($\alpha = 0.90$), *Relevance of information* (Relevance), e.g., "The assistant provided relevant information as and when I needed it." ($\alpha = 0.90$), *Maxim of quantity* (Quantity), e.g., "The assistant only gives me the information I need." ($\alpha = 0.83$), *Resilience to failure* (Resilience), e.g., "The assistant explained gracefully when it could not help me." ($\alpha = 0.75$), *Understandability and politeness* (Understand), e.g., "I found the assistant's responses clear." ($\alpha = 0.81$), *Perceived conversational credibility* (Credibility), e.g., "I feel like the assistant's responses

Table 2: Statistical distribution values for trust, explainability and SDT variable across the three assistants; SD in brackets

Assistant	Explainability	Trust	Competence	Autonomy	Relatedness
Form-based	0.73 (0.18)	3.48 (0.58)	3.67 (0.80)	3.42 (0.68)	2.87 (0.92)
Intent-based	0.72 (0.17)	3.32 (0.81)	3.74 (0.70)	3.45 (0.66)	2.88 (0.93)
LLM-based	0.80 (0.17)	3.65 (0.72)	3.93 (0.60)	3.46 (0.74)	2.97 (1.14)

Table 3: Correlation summary of the SDT variables on trust and explainability across the three assistants.

Assistant	Variable	Competence	Autonomy	Relatedness	Explainability
Form-based	Trust	**0.52****	**0.34***	0.19	**0.36***
Intent-based	Trust	**0.44****	**0.36***	**0.44***	0.14
LLM-based	Trust	0.25	**0.46****	**0.34***	0.25
Form-based	Explainability	**0.57****	**0.39***	**0.46****	
Intent-based	Explainability	0.2	0.1	**0.63****	
LLM-based	Explainability	**0.42****	**0.43****	**0.74*****	

Note: * p < .05; ** p < .01; *** p < .001

were accurate." (α = 0.85), *Speed of answer* (Speed), e.g., "The time of the response was reasonable." (α = 0.88).

3.3.3 Explainability. The system causability scale (SCS) [24] was used to evaluate the explainability of the CAs (α = 0.87). The scale is inspired by the system usability scale (SUS) and consists of ten questions on a 5-point Likert scale. The final score is calculated by adding up all values of the single items and dividing them by the maximum score possible. The aim of the SCS is to quickly assess the explainability of an interface. One item was removed as it was not relevant to our application.

3.4 Data Collection

The experiment was conducted with three schools in two different federal states in Germany with a total of N=114 participants (46 male, 67 female, 1 not specified). The participants took part in a workshop on AI and were able to test one of the assistants [25]. All participants came from 8th, 9th or 10th grade and were between 14 and 17 years old. The participants were distributed across the three groups according to gender, so that each group roughly consisted of 60% female and 40% male participants.

4 Results

This section provides an overview of the analyses of the collected data. Significance tests and correlation tests are used, which are applied to the selected dimensions such as trust, SDT variables, usability and explainability.

4.1 Trust, Explainability, Competence, Autonomy and Relatedness

To identify significant results between the groups for trust and explainability (measured by SCS score) we performed t-tests with a Bonferroni corrected alpha. We confirmed the normal distribution of trust and explainability data using a Kolmogorov-Smirnov test. The measured mean values and standard deviations for the variables in the three groups can be seen in Table 2. Additionally, the mean values and standard deviations for SDT variable were added to the table.

For trust there is a significant difference between intent-based CA and LLM-based CA (t=-2.13, p<0.05), but not between form-based and the other assistants. Therefore, the results are consistent with the proposed hypothesis. The test for explainability (SCS) revealed a significant difference between the intent-based CA and the LLM-based CA (t=-2.26, p<0.05), but not between form-based and the others. Therefore, the results support the proposed hypothesis. We analyzed the correlations of trust and explainability with the SDT variables competence, autonomy and relatedness to investigate the connection between them. The correlations are listed in Table 3.

For trust and the SDT, all variables correlate positively with each other, highlighting the importance of SDT for trustworthy assistants. Competence and autonomy correlate significantly with trust in the form-based assistant, competence, autonomy and relatedness with trust in the intent-based CA and autonomy and relatedness with trust in the LLM-based CA.

Explainability correlates significantly positively with competence, autonomy and relatedness for the form-based assistant and the LLM-based CA. For the intent-based CA the only significant correlation can be seen between explainability and relatedness. The correlations between explainability and relatedness increase from form-based to intent-based and again from intent-based to LLM-based assistants. The only significant correlation between explainability and trust exists for the form-based assistant.

4.2 Usability and Trust

The differences for the usability (measured by the BUS dimensions) of the individual assistants can be seen in Figure 1.

The LLM-based CA has higher median scores than the intent-based CA for ease to start a conversation (Start), expectation setting (Expectation), recognition and facilitation of users' goal and intent (Recognition), relevance of information (Relevance), understandability and politeness (Understand), perceived conversational credibility (Credibility) and speed of answer (Speed), but due to high

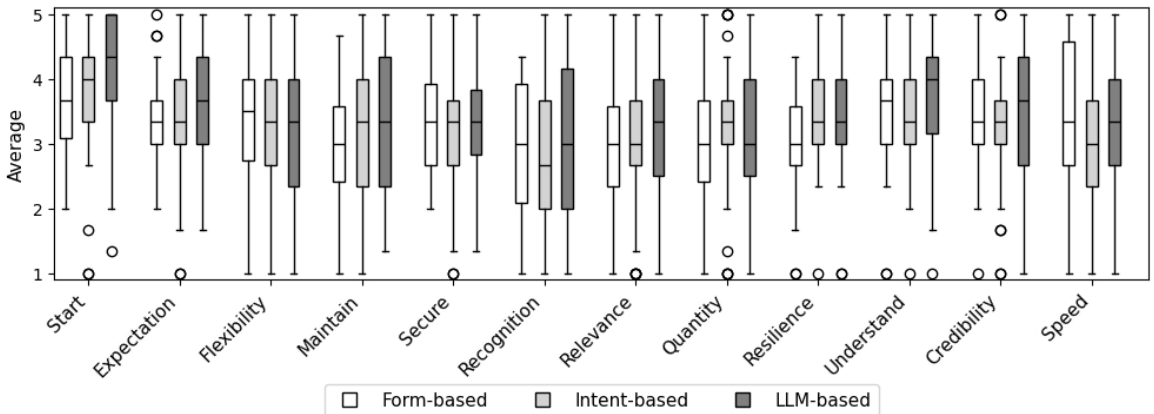

Figure 1: Boxplots for BUS variables across the three assistant versions comparing the median.

Table 4: Summary of correlations between BUS variables and trust across the three assistants.

BUS	Trust (Form-based)	Trust (Intent-based)	Trust (LLM-based)
Start	**0.38***	0.2	-0.25
Expectation	0.16	0.28	0.11
Flexibility	**0.36***	0.23	**0.36****
Maintain	0.23	0.3	**0.44****
Secure	0.19	**0.42****	**0.65*****
Recognition	0.32	0.31	**0.32***
Relevance	0.23	0.3	0.29
Quantity	0.09	0.3	0.18
Resilience	0.16	**0.38***	0.28
Understand	0.2	0.3	0.1
Credibility	0.33	0.3	**0.39***
Speed	0.29	0.14	**0.42****

Note: * p < .05; ** p < .01; *** p < .001

variances these differences are not significant. For the maxim of quantity (Quantity) the median is a bit higher for the intent-based CA. The form-based assistant is perceived inferior with respect to most dimensions compared to the other assistants. For flexibility and communication effort (Flexibility) and speed of answer (Speed) it scores slightly higher though. None of the differences between the assistants are significant. Furthermore, the correlations between the BUS dimensions and trust were analyzed in detail and are summarized in Table 4.

Except for the ease to start a conversation (Start) in the LLM-based CA, all BUS variables correlate positively with trust in all assistants, highlighting the importance of usability for trustworthy assistants. The LLM-based CA has noticeably more significant correlations between BUS dimensions and trust than the other two assistants. The ease to start a conversation (Start) and the flexibility and communication effort (Flexibility) show significant correlations with trust for the form-based assistant. For the intent-based CA users' privacy and security (Secure) and the resilience to failure (Resilience) show significant correlations, and for the LLM-based CA they flexibility and communication effort (Flexibility), the ability to maintain a themed discussion (Maintain), the users' privacy and security (Secure), the recognition and facilitation of users' goal and intent (Recognition), perceived conversational credibility (Credibility) and speed of answer (Speed) correlate significantly with trust. The correlation between trust and the users' privacy and security (Secure) in the LLM-based CA is particularly high.

4.3 Behavior Analysis

The average interaction time with all assistants and the average length of user input are listed in Table 5.

An increased average interaction time and average user input can be seen for the LLM-based CA. The standard deviation is relatively high, as the individual entries differ largely from one another. There are very short answers, but also very long descriptive answers.

5 Discussion and Conclusion

The significant difference for trust between the intent-based CA and the LLM-based CA and the non-significant difference between the form-based assistant and the two CAs have several possible reasons. The form-based assistant is a very familiar interface that

Table 5: Statistical distribution values of average user interaction time and input length; SD in brackets

Assistant	Average Interaction Time in Seconds	Average User Input Length in Characters
Form-based	402 (438)	20 (27)
Intent-based	483 (267)	38 (23)
LLM-based	906 (699)	83 (47)

is easily understood and can be used effectively. This may have a positive effect on the perceived competence and autonomy in using the assistant. Both variables are significantly correlated with trust. The perception of the form-based assistant as being human-like (relatedness) is not important for the correlation with trust, which is reasonable, as there is no human-like conversation with the form.

The intent-based CA is perceived as less trustworthy than the other two assistants because it essentially represents the form-based assistant as a CA. In times of generative CAs, such as ChatGPT [26], expectations for their performance are very high, potentially leading to reduced trustworthiness for less powerful chatbots like intent-based ones. With regard to the SDT variables, it is crucial for the trustworthiness of the intent-based assistant that users feel competent and have a sense of autonomy when interacting with the assistant. In contrast to the form-based assistant, it is also important that the assistant is perceived as human-like. Since these three properties are difficult to implement in an intent-based CA, trust is possibly reduced.

Conversely, establishing a human-like interaction is easier with the LLM-based CA. The correlation analysis with SDT variables also shows for the LLM-based assistant that this property is important for trust. Perceived competence in the LLM-based CA is higher than for the other two assistants. One possible explanation could be that the conversation is conducted as if with a human being. The information exchanged is much more detailed and individualized. As a result, the needs of the person can be addressed more precisely. This is reflected by a higher average interaction time and higher average length of user input with the LLM-based assistant compared to the other two assistants. One explanation for a missing significant correlation between competence and trust could be that there was a discrepancy expressed by the participants between the conversation that was very detailed and specific and the career recommendations that were given in a very generic way without showing connections to the conversation. The significant correlation of trust and perceived autonomy may be due to the fact that the conversation with the LLM-based CA is more flexible and thus also the main topics are determined by the human.

Perceived autonomy is important for all three assistants, which could be since all three assistants have free text input and decision-making options were implemented in the occupation and educational path recommendations. The correlation analysis of SDT variables and trust suggests that for trustworthy CAs, it is reasonable to focus on developing interaction patterns and components that enhance perceived competence, autonomy, and foster a human-like connection between the CA and the user. However, the relationships do not seem to be quite as clear as in trustworthy human-human relationships as described by the model of van der Werff et al. [8].

The significantly higher perceived explainability of the LLM-based CA compared to the intent-based CA can be explained by the fact that the LLM inserts justifying and explanatory sections in most utterances by default. In addition, the reason why a certain question is asked is explained by the CA (e.g., "To be able to assess your strengths in programming even more precisely, I would like to talk to you about your projects."). The form-based and intent-based assistants do not differ from each other in terms of explainability, which is reasonable because the intent-based CA uses almost the same utterances as the form-based assistant. However, an SCS value of 0.73 for the form-based assistant and 0.72 for the intent-based assistant indicate that these two assistants also have a certain explanatory power. Since no generative but only predefined answers are given, the reason is probably that the predefined responses already contain an explanatory guidance of the dialog, as in the LLM-based CA.

The correlation analysis with SDT variables and explainability reveals that all variables are positively correlated with each other and most of them significantly. This suggests that explainability and SDT variables are connected to each other. This would mean that either explainability is increased by a higher perceived competence, autonomy and relatedness or the perception of being competent, acting autonomous and feeling related to the assistant is affected by explainability or both, as it is a correlation. The explainability of the assistant only correlates significantly with trust in the case of the form-based assistant, although the SCS is in fact the same as that of the intent-based CA and the SCS of the LLM-based CA is even higher. In the case of the form-based assistant, it could again be due to the familiar interface. For the intent-based and the LLM-based CA, the reason could be that the way of explaining in a conversational interface has to be different than in a form-based assistant in order to generate trust. From the correlation analysis of the BUS variables with trust, it can be seen that the advantages of the LLM-based assistant to drive a thematically appropriate conversation and to always consider the user's concern and goal and to respond quickly are positively related to trustworthiness. Implementing these advantages in an intent-based CA in the same way is almost impossible, which speaks in favor of choosing an LLM-based CA, not forgetting the simplicity and effectiveness of form-based systems.

In summary, it was shown that an LLM-based CA is trusted more on average than an intent-based CA. The results indicate that flexibility of the conversation, credibility of the chatbot, recognizing user's intents and a conversation that feels secure and can be maintained are important factors for building trust in the CA. The LLM-based CA therefore has the potential to facilitate trustworthy interactions, which could also lead to higher acceptance and a better user experience. However, this requires an implementation that emphasizes the advantages of LLMs as user expectations of these systems are very high. The use of a well-known and effective interface like a form should not be underestimated. However, the depth of extracted user information for occupation recommendation could be limited. A feeling of autonomy seemed to be important in all three assistants to build trust, where relatedness was important for the two CAs. Usability differences were not significant and showed a large variability within the same CAs. Further research should concentrate on building CAs that are even more credible and can match the user profile successfully to occupation and education path recommendations to meet the expectations of the user. Although there is a significant difference in explainability between the intent-based and LLM-based CA, the correlation with trust is not significant. Further research could show whether an explicit design of human-centered explanations has a positive effect on trust.

References

[1] Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 37, (2014), 94–100. DOI: https://doi.org/https://doi.org/10.1016/j.chb.2014.04.043

[2] Emmelyn A. J. Croes and Marjolijn L. Antheunis. 2021. 36 Questions to Loving a Chatbot: Are People Willing to Self-disclose to a Chatbot? In Chatbot Research and Design, Springer International Publishing, Cham, 81–95. DOI: https://doi.org/10.1007/978-3-030-68288-0_6

[3] Moez Ltifi. 2023. Trust in the chatbot: a semi-human relationship. Future Business Journal 9, 1 (December 2023), 109. DOI: https://doi.org/10.1186/s43093-023-00288-z

[4] Tanja Bunk, Daksh Varshneya, Vladimir Vlasov, and Alan Nichol. 2020. DIET: Lightweight Language Understanding for Dialogue Systems. DOI: https://doi.org/10.48550/arXiv.2004.09936

[5] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. DOI: https://doi.org/10.48550/arXiv.2107.13586

[6] Thorsten Zylowski and Matthias Wölfel. 2023. An NLP Analysis of ChatGPT's Personality Simulation Capabilities and Implications for Human-centric Explainable AI Interfaces. In Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023), Association for Computational Linguistics, Online, 168–177. Retrieved from https://aclanthology.org/2023.icnlsp-1.17

[7] Trong Nghia Hoang, Andreas Reich, and Matthias Wölfel. 2022. The First Impression Counts! The Importance of Onboarding for Educational Chatbots. In 20. Fachtagung Bildungstechnologien (DELFI); Gesellschaft für Informatik e.V.: Bonn, Germany, 239-240. DOI: https://doi.org/10.18420/delfi2022-050

[8] Lisa van der Werff, Alison Legood, Finian Buckley, Antoinette Weibel, and David de Cremer. 2019. Trust motivation: The self-regulatory processes underlying trust decisions. Organizational Psychology Review 9, 2–3 (2019), 99–123. DOI: https://doi.org/10.1177/2041386619873616

[9] Richard Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, *55*(1), 68–78. DOI: https://doi.org/10.1037/0003-066X.55.1.68

[10] Omar Zahour, El Habib Benlahmar, Ahmed Eddaoui, Hafsa Ouchra, and Oumaima Hourrane. 2020. A system for educational and vocational guidance in Morocco: Chatbot E-Orientation. Procedia Computer Science 175, (January 2020), 554–559. DOI: https://doi.org/10.1016/j.procs.2020.07.079

[11] Omar Zahour. 2020. Towards a system for predicting the category of educational and vocational guidance questions using bidirectional encoder representations of transformers (BERT). IJATCSE 9, 1 (February 2020), 505–511. DOI: https://doi.org/10.30534/ijatcse/2020/69912020

[12] Godson D'Silva, Megh Jani, Vipul Jadhav, Amit Bhoir, and Prithvi Amin. 2020. Career Counselling Chatbot Using Cognitive Science and Artificial Intelligence. In Advanced Computing Technologies and Applications, Hari Vasudevan, Antonis Michalas, Narendra Shekokar and Meera Narvekar (eds.). Springer Singapore, Singapore, 1–9. DOI: https://doi.org/10.1007/978-981-15-3242-9_1

[13] Rasa. 2024. Rasa: Conversational AI Platform. Retrieved from https://rasa.com/

[14] Ajitesh Nair, Ishan Padhy, J. K. Nikhil, S. Sindhura, M. L. Vandana, and B. S. Vijay Krishna. 2022. AI-Based Career Counselling with Chatbots. In Intelligent Sustainable Systems (Lecture Notes in Networks and Systems), Springer Nature, Singapore, 599–612. DOI: https://doi.org/10.1007/978-981-19-2894-9_46

[15] Reema Goyal, Navneet Chaudhary, and Mandeep Singh. 2023. Machine Learning based Intelligent Career Counselling Chatbot (ICCC). In 2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE, Coimbatore, India, 1–8. DOI: https://doi.org/10.1109/ICCCI56745.2023.10128305

[16] Duy Cuong Nguyen, Nguyen Hanh Dung Dinh, Cuong Pham-Nguyen, Thang Le Dinh, and Le Nguyen Hoai Nam. 2022. ITCareerBot: A Personalized Career Counselling Chatbot. In Recent Challenges in Intelligent Information and Database Systems, Edward Szczerbicki, Krystian Wojtkiewicz, Sinh Van Nguyen, Marcin Pietranik and Marek Krótkiewicz (eds.). Springer Nature Singapore, Singapore, 423–436. DOI: https://doi.org/10.1007/978-981-19-8234-7_33

[17] Abdelmoumen Talib, Mohamed Housni, and Mohamed Radid. 2023. Utilizing M-Technologies for AI-Driven Career Guidance in Morocco: An Innovative Mobile Approach. Int. J. Interact. Mob. Technol. 17, 24 (December 2023), 173–188. DOI: https://doi.org/10.3991/ijim.v17i24.44263

[18] Muskan Sharma, Anita Kumari, and Jyotsna. 2023. AI-Based Deep Learning Chatbot for Career and Personal Mentorship. In 2023 IEEE 3rd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), 1–6. DOI: https://doi.org/10.1109/TEMSMET56707.2023.10149907

[19] Rutuja Joshi, Julia Graefe, Michael Kraus, and Klaus Bengler. 2024. Exploring the Impact of Explainability on Trust and Acceptance of Conversational Agents – A Wizard of Oz Study. In Artificial Intelligence in HCI, Springer Nature Switzerland, Cham, 199–218. DOI: https://doi.org/10.1007/978-3-031-60606-9_12

[20] Cecilie Bertinussen Nordheim, Asbjørn Følstad, and Cato Alexander Bjørkli. 2019. An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. Interacting with Computers 31, 3 (August 2019), 317–335. DOI: https://doi.org/10.1093/iwc/iwz022

[21] Dorian Peters, Rafael A. Calvo, and Richard M. Ryan. 2018. Designing for Motivation, Engagement and Wellbeing in Digital Experience. Frontiers in Psychology 9, (2018). DOI: https://doi.org/10.3389/fpsyg.2018.00797

[22] Desiderio Gutiérrez-Taño Santiago Melián-González and Jacques Bulchand-Gidumal. 2021. Predicting the intentions to use chatbots for travel and tourism. Current Issues in Tourism 24, 2 (2021), 192–210. DOI: https://doi.org/10.1080/13683500.2019.1706457

[23] Simone Borsci, Alessio Malizia, Martin Schmettow, Frank van der Velde, Gunay Tariverdiyeva, Divyaa Balaji, Alan Chamberlain. 2022. The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents, Personal and Ubiquitous Computing, vol. 26, pp. 95–119. DOI: https://doi.org/10.1007/s00779-021-01582-9

[24] Andreas Holzinger, André Carrington, and Heimo Müller. 2020. Measuring the Quality of Explanations: The System Causability Scale (SCS). KI - Künstliche Intelligenz, vol.34(2), 193–198. DOI: https://doi.org/10.1007/s13218-020-00636-z.

[25] Thorsten Zylowski, Nathalai Sautchuk Patricio, Wladimir Hettmann, Katharina Anderer, Matthias Wölfel, Peter Henning. 2024. Evaluating Trustworthiness, Usability and Explainability of an Educational Pathway Recommendation System that use a Large Language Model. EDULEARN24 Proceedings, pp. 6461-6470. DOI: https://doi.org/10.21125/edulearn.2024.1528

[26] ChatGPT. 2024. ChatGPT: OpenAI. Retrieved from https://openai.com/chatgpt/