“Clicking vs. Writing”—The Impact of a Chatbot’s Interaction Method on the Working Alliance in AI-based Coaching

Vanessa Mai¹ · Caterina Neef¹ · Anja Richert¹

Accepted: 24 December 2021 / Published online: 14 January 2022
© The Author(s) 2022

Abstract

Coaching has become an important didactic tool for reflecting learning processes in higher education. Digital media and AI-based technologies such as chatbots can support stimulating self-coaching processes. For the use case of student coaching on the topic of exam anxiety, the working alliance between a coaching chatbot and a human coachee is investigated. Two coachbot interaction methods are compared: A click-based chatbot (implemented in a rule-based system), where the coachee can only click on one answer, and a writing-based chatbot (implemented in a conversational AI), which allows the coachee to freely type in their answers. The focus is on which coachbot interaction method enables a stronger working alliance between coach and coachee: a click-based or a writing-based chatbot. The working alliance and the technical realization of the chatbot systems were investigated in an exploratory quantitative study with 21 engineering students. The results indicate that the working alliance in both study conditions can be classified as medium to high overall. The results further show higher values for bonding on a writing-based platform than when using a click-based system. However, click-based systems seem to be more helpful as a low-threshold entry point to coaching, as they guide coachees better through the process by providing predefined answers. An evaluation of the technical realization shows that self-reflection processes through digital self-coaching via chatbot are generally well accepted by students. For further development and research, it is therefore recommendable to develop a “mixed” coachbot that allows interaction via clicking as well as via free writing.

Keywords (Digital) Coaching · AI-based Coaching · Conversational AI · Chatbot · Working Alliance

✉ Vanessa Mai
vanessa.mai@th-koeln.de

¹ Faculty of Process Engineering, Energy and Mechanical Systems, TH Köln/University of Applied Sciences, Betzdorfer Str. 2, 50679 Cologne, Germany

„Klicken vs. Schreiben“ – Einfluss der Interaktionsmethode eines Chatbots auf die Arbeitsbeziehung im KI-basierten Coaching

Zusammenfassung

Coaching ist im Rahmen der Hochschulbildung zu einem wichtigen didaktischen Instrument zur Reflexion von Lern- und Arbeitsprozessen geworden. Digitale Medien und KI-basierte Technologien wie Chatbots können dabei unterstützen, Selbstcoaching-Prozesse anzuregen. Für den Anwendungsfall des Studierendencoachings zum Thema Prüfungsangst wird die Arbeitsbeziehung im Coachingprozess zwischen einem Coaching-Chatbot und einem menschlichen Coachee untersucht. Es werden zwei Chatbot-Interaktionsmethoden verglichen: Ein klickbasierter Chatbot (implementiert in einer regelbasierten Systemarchitektur), bei dem der Chatbot eine Antwortauswahl anbietet und der Coachee eine Möglichkeit anklicken kann, und ein schreibbasierter Chatbot (implementiert in einer Conversational AI), der dem Coachee erlaubt, seine/ihre Antworten frei einzugeben. Der Fokus liegt auf der Frage, welche Chatbot Interaktionsmethode eine stärkere Arbeitsbeziehung im Coachingprozess ermöglicht: ein klickbasierter oder ein schreibbasierter Coaching Chatbot. Die Arbeitsbeziehung sowie die technische Realisation der beiden Chatbot-Systeme wurden in einer explorativen quantitativen Studie mit 21 Ingenieurstudierenden untersucht. Die Ergebnisse zeigen, dass die Arbeitsbeziehung in beiden Studienbedingungen (klick- und schreibbasiert) insgesamt als mittel bis hoch eingestuft werden kann. Die Ergebnisse zeigen zudem höhere Werte für das Bonding in einer schriftbasierten Chatbot-Interaktion als bei der Verwendung eines klickbasierten Systems. Allerdings scheinen klickbasierte Systeme als niedrigschwelliger Einstieg in das Coaching hilfreicher zu sein, da sie den/die Coachee durch vordefinierte Antworten besser durch den Prozess führen. Darüber hinaus zeigt die Auswertung der technischen Realisierungsbedingungen, dass Selbstreflexionsprozesse durch digitales Selbstcoaching via Chatbot von den Studierenden grundsätzlich gut angenommen werden. Für die weitere Entwicklung und Forschung ist es daher empfehlenswert, einen „gemischten“ Coachbot zu entwickeln, der sowohl die Interaktion durch Klicken als auch durch freies Schreiben ermöglicht.

Schlüsselwörter (Digitales) Coaching · KI-basiertes Coaching · Conversational AI · Chatbot · Arbeitsbeziehung

1 Introduction

It is impossible to imagine today's society without the use of artificial intelligence (AI): AI enables users to precisely evaluate search queries, receive personalized recommendations on multimedia platforms, make use of the facial recognition algorithm in smartphones—and also interact with natural language conversational user interfaces, so-called chatbots or voice assistants, such as Amazon's Alexa or Apple's Siri (Singh and Thakur 2020). The same technology is increasingly finding its way into other application areas, such as customer service (Barton and Müller 2021), as a digital assistant for physicians (self-diagnosis chatbot) (You & Gui 2020), or in the form of chatbots that have a coaching character and serve to support mental or emotional health (Deepika et al. 2020; de Gennaro et al. 2020).

The latter also holds enormous potential for universities. Students face challenges during their studies, such as changes in their personal lives, facing new tasks, pressure to perform, and decision-making. To support students with these difficulties, coaching is increasingly offered at universities (Wiemer 2012). Targeted questioning techniques are used to stimulate students' self-reflection, making possible actions visible again and promoting solution-oriented thinking (Albrecht 2016). Coaching no longer takes place only face-to-face but is also practiced digitally, including

self-coaching via chatbots (Kanatouri 2020). The interaction with such a coaching chatbot or coachbot cannot replace human-human reflection and feedback conversations, but it offers advantages: In chatbot-based interactions, the willingness to reveal something about oneself may be higher—especially for sensitive topics—because the systems are perceived by users as unbiased (Lee et al. 2020). This may lead to earlier self-identification of counseling concerns by students. In addition, coachbot services are a good preparation for having in-depth conversations with advisors, for which students can then formulate specific concerns. Furthermore, a coaching chatbot offers the advantage of being available to students around the clock and is scalable to large numbers of students.

Building on existing coaching formats at universities (Mai 2020; Reis 2006; Schumacher 2006; Wiemer 2012) reflection conversations are currently being made scalable by implementing a coaching chatbot. An AI-based coachbot is being developed and implemented at the *Faculty of Process Engineering, Energy and Mechanical Systems at TH Köln/University of Applied Sciences* to deepen students' self-reflection in a coaching process (Mai and Richert 2021). In such a process, goals are defined (e.g., dealing with exam anxiety) based on which interventions can be implemented. These interventions consist of a structured coaching conversation that uses open-ended questions to deepen stu-

dents’ self-reflections, as well as materials in the form of summaries, videos, reflection tools such as self-assessment tests, and scales (for interventions in face-to-face coaching see Berninger-Schäfer 2018; Schlippe and Schweitzer 2016; and in AI-based coaching Graßmann and Schermuly 2021; Terblanche 2020).

The coaching processes between the coachbot and students are being researched at the Faculty as part of a project. The focus is on investigating effectiveness factors in chatbot coaching. A central effectiveness factor is the establishment of a functioning working alliance (Graßmann et al. 2019; Künzli 2019; Lippmann 2013; Wasylyshyn 2003). Within the project (Mai and Richert 2020), key factors for an effective working alliance are being developed from the current research discourse on face-to-face and digital coaching (Berninger-Schäfer 2018; de Haan et al. 2016; Grant 2014; Lindart 2016) as well as on AI-based coaching (Graßmann and Schermuly 2021; Terblanche 2020; Terblanche and Cilliers 2020). These factors are further operationalized into intervention strategies.

Within this contribution, we present several aspects influencing the working alliance (see Chap. 2.3) and raise the question of how these aspects can be promoted in a coaching process between chatbot and human coachee. Related work suggests that the interaction method of a coaching chatbot (click-based vs. writing-based) could be an influencing factor (see Chap. 2.2). We, therefore, hypothesize that there are differences in the strength of the working alliance—especially in the outlined use case of chatbot coaching—depending on whether a click-based or writing-based coaching chatbot is used. We investigate which coachbot interaction method enables a stronger working alliance between coach and coachee. In doing so, we compare two chatbot design approaches and interaction methods with each other. Different interaction methods go hand in hand with different chatbot design approaches (see Chap. 2.1): Click-based chatbots are often rule-based, whereas conversational AI chatbots operate as writing-based chatbots with free text input.

In this study, two coaching chatbot systems that differ in their interaction methods (clicking vs. writing) were developed and programmed for the specific use case of student coaching on exam anxiety (see Chap. 3.1 and 3.2). The working alliance and technical implementation conditions of the two chatbot systems were investigated in an exploratory quantitative study with engineering students (see Chap. 3.3). The results of the study are presented and discussed (see Chap. 4 and 5). They contribute to the question of how the choice of interaction method and chatbot design approach influences the establishment of a working alliance in the use case of student coaching.

2 Definitions and Related Work

2.1 Chatbot Classifications

Chatbots are software-based agents that interact with users to provide access to services and information. They are mostly incorporated into websites via chat functions or as virtual assistants in messaging apps to interact with users in text-based conversations (Barton and Müller 2021). Two factors are presumably contributing to the current interest in the use of chatbots (Brandtzaeg and Følstad 2017): On the one hand, the mobile internet and the increase in messaging platforms (such as Facebook Messenger, Skype, Telegram, Slack) are promoting the acceptance of chatbots (Følstad and Brandtzaeg 2017), which means that a large number of potential users can be reached. For companies, this opens up new possibilities e.g., in customer service and automated marketing. On the other hand, developments in communication technology, artificial intelligence, and machine learning enable better interpretation and prediction of natural languages (Shah et al. 2016). For the technical development of chatbots, this means that advances in conversation modeling are expected (Følstad et al. 2019; Hussain et al. 2019; Ramesh et al. 2017).

The willingness to use chatbots is increasing among users. According to a study that examined people’s motivation to use a chatbot (Brandtzaeg and Følstad 2017), the central motivation is productivity: Chatbots are used because they can provide information, assistance, and feedback independent of time and location as well as in a targeted and process-oriented manner.

Due to the rapid development of chatbot technologies as well as application areas, the classification of chatbots does not appear to be straightforward (Hussain et al. 2019). Chatbots are still an emerging technology, making the classification complex and multi-layered (Følstad et al. 2019). Chatbots can be classified according to different criteria and categories, including the interaction type (voice- or text-based), the knowledge area of the chatbot/its field of application (domain-specific or open domain), the goal or use case of chatbots (task-oriented or non-task-oriented), and the design approaches/techniques used in the development of chatbots (rule-based or conversational AI) (Følstad et al. 2019; Hussain et al. 2019; Ramesh et al. 2017; Sugisaki and Bleiker 2020).

In this paper, we classify chatbots according to their design approach or—more technically speaking—according to their system architecture (Hussain et al. 2019; Ramesh et al. 2017). Depending on their technical basis, chatbots can be divided into rule-based bots or conversational AI bots. Conversational AI bots can be further distinguished in terms of how they process dialog inputs and generate responses. Most current systems use retrieval-based methods,

while generative methods are still in research. Considering the rapid technological developments, one can assume that conversational AI chatbots will increasingly complement or overcome classic rule-based chatbots (Følstad et al. 2019).

Rule-based chatbots can only retrieve programmed content when communicating with a user, while conversational AI chatbots are based on machine learning, i.e., they use artificial intelligence that allows them to learn and follow patterns in order to understand the intent of the user, performing increasingly better the more data they are trained with (Brandão and Wolfram 2018; Hussain et al. 2019; Lömker et al. 2021; Ramesh et al. 2017). Also, conversational AI chatbots use text to interact with the user based on natural language (Jurafsky and Martin 2008; Kellner 2001). Through the use of conversational AI, the chatbot is also able to understand a user's intention even if they misspell words, as the chatbot learns to identify specific sentence patterns rather than looking for keywords. The use of text as an interaction method differs from chatbots that do not (only) include natural language as input but also use graphical user interfaces elements such as images, buttons, or lists to illustrate and select given options or choices (Klopfenstein et al. 2017; Shevat 2017). These types of chatbots are often rule-based chatbots.

In this paper, we follow definitions by Sugisaki and Bleiker (2020) and define writing-based chatbots as chatbots that allow users to freely express their utterances (via text) and initiate a conversation through the use of conversational AI. Click-based chatbots, on the other hand, are defined as chatbots that enable interaction with the user via predefined answers of the chatbot, which the user can click on. Both systems have advantages and disadvantages. Click-bots—which are always programmed as rule-based systems—are easier to construct and cheaper and faster to implement (Lömker et al. 2021). However, with click-bots, each scenario must be programmed in a time-consuming manner, and unforeseen requests from users cannot be answered by the chatbot (Hussain et al. 2019; Ramesh et al. 2017). In addition, the flow of conversation often seems somewhat rigid (Thanaki 2018). The advantages of conversational AI chatbots are that they save time and money in the long term through the use of machine learning since new expressions with the same intention do not have to be programmed over and over again (Hussain et al. 2019; Ramesh et al. 2017). They can also be programmed to “understand” different languages and different expressions with the same intention and can respond to users' emotions accordingly (Zhou et al. 2020). However, the disadvantages of conversational AI chatbots are that they rely on a large amount of training data and a long training period that allows them to respond individually to the user (Hussain et al. 2019; Ramesh et al. 2017). They sometimes learn things they are not supposed to learn, which can lead to them making of-

fensive or undesirable statements (Joshi 2020; Lee et al. 2019).

2.2 (AI-based) Coaching: Promotion Self Reflection (through Writing)

Coaching via chatbots can be classified as digital self-coaching (Kanatouri 2020). In coaching, chatbots, on the one hand, take over simple tasks, e.g., by suggesting internet resources relevant for the coachee or feedback and tips on specific topics. On the other hand, there are chatbots that enable coaching interactions based on coaching methods and models to encourage reflection from the coachee. While the machines cannot implement all the skills of a human coach in a conversation, they are helpful for certain coaching aspects, such as questions to reflect on strengths and approaches as well as guiding clients through decision-making processes (Graßmann and Schermuly 2021; Kanatouri 2020; Terblanche 2020). Følstad et al. (2019) classify chatbots according to the duration of relation in the interaction and to the locus of control. In this typology, coaching chatbots are long-term oriented and chatbot-driven. That means that coaching chatbots are characterized by a more chatbot-driven interaction that enables coachees to go through a structured coaching process.

Coaching literature provides insights into the opportunities and challenges of using coaching chatbots. Terblanche (2020) presents the Designing AI Coach framework (DAIC) in which he makes conceptual recommendations for designing organizational AI coaches. Although research on coaching chatbots is rare, there seems to be some consensus that coaching chatbots can “help clients to develop their self-coaching approach that is inexpensive and accessible” (Terblanche 2020, p. 156). Graßmann and Schermuly (2021) also note a (still) absence of research on AI-based coaching but can provide evidence from the literature of the effectiveness of AI-based therapy (e.g., the reduction of depression and anxiety; Fulmer et al. 2018). One of the few studies examining the acceptance of coaching chatbots concludes, among other things, that users are more willing to use a coaching chatbot if they derive a practical benefit from the chatbot interaction (Terblanche and Cilliers 2020). This means that both the technical design and the chatbot concept must be flawless and user-centric.

An important goal in coaching is to stimulate self-reflection, which can be through interventions such as systemic questioning techniques, reframing, perspective-shifting, or resource activation (Schlippe and Schweitzer 2016). That chatbots can stimulate self-reflection processes is shown by Fitzpatrick et al. (2017), who investigated a therapy chatbot. Woebot is supposed to help against depression and anxiety and uses techniques of cognitive behavioral therapy. It encourages the user to self-reflect and think positively. The

study shows that test subjects who interacted with the chatbot felt significantly better than the control group, which received an e-book with self-help strategies.

In coaching, methods such as creative or expressive writing (e.g., in the form of free writing or diary entries) from writing therapy can also be used to stimulate writing-based (self)reflection processes with regard to personality development (Reindl et al. 2012). Free writing can, therefore, be used as an intervention method of a coaching chatbot (here: free typing of thoughts by the user). One question in this context is to which extent free writing in chatbot coaching impacts the satisfaction, acceptance, or effectiveness of coaching. Especially against the background of different chatbot design approaches or system architectures (rule-based vs. conversational AI, see Chap. 2.1), this question needs to be investigated in particular in comparison with click-based coaching-chatbot interactions.

Insights regarding the interaction method of a chatbot (click-based or writing-based) are provided by a study by Hauser-Ulrich et al. (2020) that examines, among other things, the effectiveness, acceptance, and working alliance of the text-based health care chatbot SELMA. SELMA primarily offers predefined answer options in the interaction, e.g., as a selection of coping strategies. The results of the study show that subjects criticize the lack of free text input when interacting with SELMA. They wished to be able to enter free text as well as to be given more predefined answer options. The test subjects had the impression that the interaction was too static and inflexible.

Hauser-Ulrich et al. (2020) interpret this result in terms of the media-equation-theory (Reeves and Nass 1998) and computers-as-social-actors paradigm (CASA) (Nass et al. 1994), according to which humans tend to interact with computers (or chatbots) as with other humans. Accordingly, the desire for a writing-based coaching-chatbot interaction suggests a desire to interact in chatbot coaching in the same way as in human-to-human coaching. The authors of the study recommend a combination of predefined answer options and free-text input for future coachbot designs to allow users more autonomy in the interaction.

2.3 Working Alliance in (Human-Machine-) Coaching

In coaching, the development of a sustainable working alliance between coach and coachee is considered one of the key effective factors and has been demonstrated vividly in research (de Haan et al. 2016; Graßmann et al. 2019; Künzli 2019; Lindart 2016; Lippmann 2013; Wasylshyn 2003). The concept goes back to Bordin (1979), according to whom the working alliance consists of three aspects: agreeing on the goals to be achieved, determining the tasks that need to be completed to achieve the goals, and estab-

lishing a bond, which includes building trust, appreciation and respect. To operationalize Bordin’s working alliance, Horvath and Greenberg (1989) developed the Working Alliance Inventory (WAI). In Germany, the short form WAI-SR has proven useful (Wilmers et al. 2008).

For a working alliance to develop, both client and coach should be actively involved in the coaching process so that the client can self-reflect and reveal insight into their own thoughts, feelings, and behavior (Grant 2014). If this is not possible, coaches may not be able to obtain sufficient information about the clients’ situation. This lack of information can hinder the proper handling of clients’ concerns (Graßmann et al. 2019; Graßmann and Schermuly 2016). Lindart (2016) also emphasizes the importance of promoting client autonomy.

Research shows that relationship building can also emerge in human-machine interaction. There are various studies on the importance of rapport building with virtual agents (Gratch et al. 2007; Huang et al. 2011; Zhao et al. 2014). The concept of rapport in human-machine interaction is similar to the factor bonding in the working alliance (Sharpley 1997). Subjects perceived, among other things, lower resistance to self-disclosure and higher ease of use when they believed they were interacting with a virtual agent than when they believed they were interacting with a human (Gratch et al. 2014). Bickmore and Picard (2005) found that conversational agents who can build rapport are more respected and trusted by users than purely task-oriented conversational agents. For AI-based therapy, research also shows that clients can form a bond with their virtual agent (Bickmore et al. 2010).

Terblanche (2020) compiles key aspects from the literature on human-machine interaction to support a strong coach-coachee relationship. It can be assumed that these effects also apply to AI-based coaching and that coachbots are able to establish a working alliance with clients (Graßmann and Schermuly 2021; Hauser-Ulrich et al. 2020; Mai et al. 2021). With regard to the three aspects of the working alliance (task, bond, goal), it can be assumed that an AI-based coach might be more successful on the bond- and task- level and show challenges on the goal-level. For the task-level, Graßmann and Schermuly (2021) show that AI-based coaching might be well able to guide coachees through a coaching process in a structured way by using a variety of methods (psychoeducational stimuli, psychometrics, creativity and goal development methods, visualizations, scaling). Terblanche and Cilliers (2020) show first empirical evidence for this. Hauser-Ulrich et al. (2020) transferred the concept of working alliance to the interaction patterns between a health care chatbot and users (Kiluk et al. 2014). Their results show that the working alliance between chatbot and client is comparable to the working alliance in internet-based human-to-human therapies. The

subjects enjoyed using the chatbot and found it helpful and easy to use. For the bond-level, Mai et al. 2021 show first empirical hints. They investigated to what extent the self-disclosure of a coaching chatbot influences the self-disclosure of the test subjects and which effects this shows on the bond-level of the working alliance and the rapport building. The results show that students disclose themselves to the chatbot and establish a relationship with it. The fact that the coaching is conducted by a machine makes it easy for students to open up. There is a great willingness to use it. However, AI-based coaching seems to have the greatest difficulties in problem identification by clients and in providing individualized feedback. This concerns the goal-level. Accordingly, AI-based coaching processes seem to have the greatest chance of success when “the client is already aware of the core problem and simply needs to be prompted with questions to reflect upon it” (Graßmann and Schermuly 2021, p. 9). Studies show that self-coaching works well when coachees are already aware of what behaviors they want to improve in the future (Sue-Chan and Latham 2004).

2.4 Research Question and Hypothesis

The aim of this study is to investigate the working alliance in a coaching process between a coaching chatbot and a human coachee. Various aspects can influence the working alliance when designing a coaching chatbot. Few have been studied so far (see Chap. 2.2 and 2.3). The question arises how aspects of the working alliance can be promoted in a coaching process between chatbot and human. The related work in Chap. 2.2 suggests that the interaction method of a coaching chatbot could be an influencing factor. Does it make a difference in establishing a working alliance whether clients can write freely with the coaching chatbot or click on predefined answers? To what extent does the type of interaction method influence aspects conducive to working alliance, such as promoting client autonomy and active engagement as well as facilitating self-reflection and self-disclosure (see Chap. 2.3)?

In this paper, we investigate which coachbot interaction method enables a stronger working alliance between coach and coachee: a click-based or a writing-based coaching chatbot. We hypothesize that there are differences in the strength of the working alliance—especially in the outlined use case of chatbot coaching—depending on whether a click-based or writing-based coaching chatbot is used. In doing so, two chatbot design approaches are compared with each other (see Chap. 2.1): A click-based chatbot, where the chatbot provides predefined answer choices and the coachee can only click on one answer, and a conversational AI chatbot, which asks the same questions but allows the coachee to type their answers freely.

3 Research Design

3.1 Chatbot Concept and Interaction Script

The chatbot in this study is designed as a coaching chatbot for students. The topic of exam anxiety was chosen as the use case. Coaching can help with exam preparation, one of the areas where students perceive a high need for support (Traus et al. 2020). In the coaching process designed here, students are encouraged—guided by questions—to reflect on their exam anxiety and their own experiences. The goal is to expand the students’ options for action and to develop their own strategies for dealing with their exam anxiety. The overarching goal of chatbot coaching is to stimulate students’ self-reflection and solution-oriented thinking so that students can more successfully manage their studies. In contrast, the explicit non-goal of coaching is to provide students with as much knowledge as possible on the topic of exam anxiety or to give advice or instructions. The coaching approach presented here is based on systemic coaching, which is oriented towards a solution- and resource-oriented perspective (Berninger-Schäfer 2018; Schlippe and Schweitzer 2016).

Based on a study (Mai et al. 2021), an existing interaction script of a coaching chatbot was further developed. The coaching process in this study is oriented on coaching process steps from literature (Berninger-Schäfer 2018) and can be divided into three phases: onboarding, situation description & clarification, and goodbye. A clarification of the concern was deliberately omitted since, in the case of the topic of exam anxiety, the concern had already been established in advance. Moreover, research in the field of coaching chatbots suggests that coaching can develop its strengths better if the client is already aware of their core problem in advance (see Chap. 2.3).

Onboarding is an integral part of a chatbot concept and includes welcoming and getting to know the user as well as the chatbot. In onboarding, the chatbot introduces itself and explains what its tasks are. This sets the expectations for the chatbot’s capabilities and defines its boundaries (Lovejoy 2019; Terblanche 2020). In addition, the user’s name is asked for in order to address them more personally in the course of the conversation.

Onboarding is followed by the situation description & clarification. It starts with a scale question in which the coachee is asked to rate their exam anxiety on a scale of 1 to 10. This question serves as an introduction to the topic. It makes the subjective feeling of exam anxiety tangible and serves to check whether exam anxiety is present and the user thus falls into the target group of the coaching. Specific systemic coaching questions are then used to describe and clarify the situation. So-called problem-oriented questions are used to illustrate to clients how problems are

Fig. 1 Activity Diagram of the Interaction Script

actively produced and maintained, from which, in reverse, it is also made clear what should be avoided if one wants to get rid of the problem (Schlippe and Schweitzer 2016) (e.g., “What would you have to do to make your exam anxiety worse?”). In contrast, questions such as “Have you talked to other people about your test anxiety?” target the social context of the problem Wolf (2017). For clarification, solution-oriented questions, such as so-called miracle questions, are used: “How would you know your exam anxiety is gone?” (Schlippe and Schweitzer 2016). Afterward, there is an opportunity for the user to receive mini tips against exam anxiety, such as information on self-organization or breathing exercises. After the situation description & clarification, the interaction ends with a goodbye.

To make the interaction with the chatbot seem more natural, the chatbot responds to the user’s answer with an utterance at some points. Since it is not clear in advance what the coachee’s response will be, empathetic expressions from the chatbot such as “I think it’s great that you’re talking to me about this” were integrated.

The structure of the interaction script can be shown in an activity diagram (see Fig. 1). The activity diagram shows

all possible communication paths. It can be interpreted as a behavior diagram and describes a basic flow of the use case. The activity diagram is written in the Unified Modeling Language 2.0 (Object Management Group 2005).

3.2 Program Architecture Landbot/Rasa

The two system architectures Rasa and Landbot were selected for the study. The differences between the two platforms relate to the development process and the interaction method. Both systems have advantages and disadvantages (see Chap. 2.1). Landbot is rule-based and designed for intuitive chatbot design, which does not require programming knowledge. Moreover, the user interacts with the chatbot mainly through buttons (Landbot 2020). In contrast, Rasa is a conversational AI and provides YAML and Python-based programming of the chatbot. Rasa, as a conversational AI framework, is used to recognize the user’s intention in order to enable the writing-based interaction method. It offers the advantage over a rule-based framework of reducing technical difficulties in the interaction between the user and the bot.

Fig. 2 Program Structure of the Coaching Chatbot in Landbot

3.2.1 Landbot Mode of Operation: Rule-based and Click-based

A rule-based chatbot developed on the Landbot platform is based on a modular structure. It is advisable to design an interaction script in advance (see Chap. 3.1). Based on this, the different conversation paths that the user can take in the course of the conversation can be modeled. This is done with the help of modules and interaction paths, as shown in Fig. 2. The white boxes represent the chatbot's statements and the red boxes correspond to the defined response options. The number and definitions of the choices can be customized as we go along. Landbot also provides additional interaction options, such as graphical interaction or text-based input of keywords (Landbot 2020). However, this was deliberately omitted in this chatbot development so that comparability of the systems can be ensured.

3.2.2 Rasa Mode of Operation: Conversational AI and Writing-based

In Rasa, conversation is enabled via conversational AI. One main function for this is *Natural Language Understanding* (NLU). Rasa converts human speech into structured files. It thus converts messages from the user into so-called *intents* so that the chatbot can understand them. Subordinate machine learning libraries such as *Tensorflow* and *spaCy* are used for this purpose. Accordingly, an *intent* corresponds to the possible responses of a recurring conversation segment (such as the greeting, the goodbye, the agreement).

To ensure high accuracy in predicting a conversation flow, a large number of possible user responses in the *intents* is conducive. In addition, keywords in the *intents* can be defined as *entities*. *Entities* are used, for example, to recognize the user's name in the course of the conversation. The cornerstone of the chatbot is *Natural Language Processing* (NLP). NLP is a category of machine learning that analyzes the free form of text and then converts it into structured files. NLU is a subset of NLP. Thus, while NLP only converts the texts into files, NLU interprets the texts and also takes into account the chat histories. The important keywords and the meaning of user input can thus be determined by Rasa (Barton and Müller 2021).

One advantage of using Rasa is that the security of user data is a top priority. Rasa protects user data by not having to send it to a third-party service for message processing. The data can be stored on-site, namely on a private server—in the case of this study, on a server set up specifically for this study.

3.3 Experimental Design

The two developed and programmed coaching chatbots were embedded in an experimental design that included a combination of experiment and survey. Both Rasa and Landbot are based on the same interaction script.

Fig. 3 Comparison of the Click-based Chatbot (a) and the Writing-based Chatbot (b)

3.3.1 Test Conditions and Pretest

The chatbot developed on the Landbot platform is based on an interaction with the user via suggested answer options and a click interaction. In contrast, the chatbot in the Rasa environment is designed so that the user formulates and enters their answers independently. Since in this way, a conversation progression in Rasa can take an infinite number of directions, while the progression in the Landbot is strictly predefined, this results in different starting conditions for the research question. To counteract this phenomenon, an activity diagram was designed that forms the basis for both chatbot systems (see Chap. 3.1). This ensures that despite any input from the user in the writing-based chatbot, the course of the conversation is the same throughout and the systems differ only by the distinguishing feature of clicking vs. writing.

To ensure this, initial pretests evaluated the most common responses in Rasa so that they could form the basis for the choices in Landbot and be implemented in the system (for an example, see Fig. 3). In addition, the congruent structure of the chatbots made it possible to compare the course of the conversation in the evaluation process and to examine it for anomalies.

3.3.2 Survey and Test Procedure

To answer the research question identified in this paper, a questionnaire consisting of three parts was developed (see Table 1) for participants in the study to fill out after their interaction with the chatbot. The first and second parts of the questionnaire aimed to measure the working alliance in terms of bonding and agreement on coaching tasks (see Chap. 2.3) and were based on the German short form of the Working Alliance Inventory WAI-SR (Wilmers et al. 2008). Questions from the WAI-SR bonding and agreement on tasks subscales were selected and adapted for the chatbot coaching context. The words “therapy” and “therapist”

were replaced with “coaching” and “coachbot” to refer to the chatbot coaching. The goal subscale was left out since the basic coaching concern exam anxiety as well as the coaching process were already defined in advance in the chatbot concept and interaction script (see Chap. 3.1).

The third part of the questionnaire referred to the interaction method and the technical implementation. For a functioning chatbot interaction and for a working alliance to develop, the system architecture must be stable. Due to the different interaction methods of the two chatbot systems, some questions were worded differently (see Table 1, questions 7 and 8). The aim of this section was to evaluate the technical implementation from the users’ point of view. Since both chatbot systems are still in an early stage of development, early identification of bugs is beneficial for further development of the bot. Furthermore, the perception of the technical framework gave an indication of feasibility and adaptability. In addition, the questionnaire contained a free text field in which subjects could enter comments.

The target group for chatbot coaching were students who suffer from exam anxiety. The term exam anxiety in this case was defined very broadly to reach more test subjects and included nervousness before or during exams. The strength of exam anxiety was not asked in the questionnaire but was integrated into the chatbot coaching (see Chap. 3.1). The questionnaire was provided to the test subjects in combination with the respective chatbot web link via an integrated random number generator so that an even distribution of the test subjects could be ensured. The test procedure included one chatbot coaching conversation. Following the chatbot coaching, the subjects participated in the online questionnaire survey.

3.3.3 Data Evaluation

In order to assess the participants’ perceived bonding, agreement on tasks, and preferred method of interaction & technical realization, the questionnaire results were an-

Table 1 Questionnaire

Question Category	Item	Question to Rasa User (writing-based)	Question to Landbot User (click-based)
<i>Working Alliance: Bonding</i>	1 Bonding	I think my coachbot likes me	
	2 Bonding	My coachbot and I respect each other	
	3 Bonding	I feel that my coachbot appreciates me	
<i>Working Alliance: Agreement on tasks</i>	4 Task-Agreement	The coachbot has made me aware of how I can change	
	5 Task-Agreement	Coaching opens up new ways of looking at my problem	
	6 Task-Agreement	I believe that the way we are working on my problem is right	
<i>Interaction Method and Technical Realization</i>	7 Interaction Method	I like the fact that I can formulate free answers	I like the fact that I get predefined answers
	8 Interaction Method	I believe that predefined answers would have supported me better in the coaching process	The suggested answers apply to me
	9 Technical Difficulties	I felt hindered by technical difficulties during the conversation	
	10 Technical Framework	I find the technical framework of a chatbot appropriate for coaching support	
<i>Additional Comments</i>			

alyzed using statistical methods. For a first insight into the data set, the socio-demographic data of the subjects were analyzed. Subsequently, the answers to each question were analyzed according to their frequency. For this purpose, the categorical attributes from “rarely” to “always” were converted into numerical values on a Likert scale from one to five. By forming the arithmetic mean, a statement can be made about the distribution of the data in the data set.

To determine the statistical significance of the results, despite the small sample sizes of 10 and 11 participants for each group, the p -value of each item was calculated. For item 8, the results of the writing-based interaction method were inverted, as the question for this item negated the question for the previous item and thus the used scale. Additionally, as the questions for item 8 ask different aspects for the writing-based and click-based chatbots and are therefore not comparable one-to-one, the p -value was omitted for this item. Additionally, the Cronbach’s alpha value was calculated to determine scale reliability for the categories bonding, task-agreement and interaction method. As items 7 and 8 for the click-based chatbot enquire different things, the Cronbach’s alpha value was omitted for this item.

In addition, the chat histories were scanned and examined to see which conversation paths were more frequented than others. Furthermore, the statements in the comment section of the questionnaire were evaluated.

4 Results

4.1 Sample Description

21 students participated in the experiment, 10 of whom interacted with the writing-based Rasa coaching chatbot, and 11 of whom interacted with the click-based Landbot coachbot. 19 trial participants were male, and 2 were female. The age of the subjects ranged from 20 to 39 years, with 18 subjects between 20 and 29 years old, which is a typical age range for students in Germany (VUMA 2021). Half of the subjects were in the bachelor’s program in Mechanical Engineering, while the other half were in the master’s program in Mechanical Engineering. One subject was studying Computer Science.

The sample can be further described according to the strength of the students’ exam anxiety. On a scale from 1=very low to 10=very high, the average exam anxiety of the subjects was $M=5.6$. The minimum of test anxiety was 2, and the maximum was 9. In the click-based group, the subjects had higher exam anxiety ($M=6.5$) than in the writing-based group ($M=4.7$). On average, subjects interacted with the writing-based chatbot for 7.63 min and with the click-based chatbot for 6.89 min. Participants, therefore, spent an average of ten percent more time on the writing-based platform.

4.2 Evaluation of the Questionnaire

The statistical analysis of the questionnaire results is shown below (see Table 2). Overall, the results indicate good scale reliability for the items of each category (Cronbach’s alpha between 0.66 and 0.89). The majority of the p -values range from 0.11 to 0.98 and therefore do not show statis-

Table 2 Statistical Analysis of Questionnaire Results

Item	Writing-based mean (and standard derivation) (<i>N</i> =10)	Click-based mean (and standard derivation) (<i>N</i> =11)
1 Bonding	3.10 (1.73)	3.00 (1.34)
2 Bonding	3.20 (1.03)	2.36 (0.92)
3 Bonding	3.20 (1.48)	3.09 (1.38)
Total	3.17 (1.39)	2.81 (1.24)
<i>Cronbach’s alpha</i>	0.66	0.84
4 Task-Agreement	2.60 (1.43)	2.55 (1.21)
5 Task-Agreement	2.40 (1.58)	3.00 (1.18)
6 Task-Agreement	2.70 (1.16)	3.55 (1.13)
Total	2.57 (1.36)	3.03 (1.21)
<i>Cronbach’s alpha</i>	0.79	0.73
7 Interaction Method	3.70 (1.57)	2.45 (1.51)
8 Interaction Method	3.90 (1.37) (inverted)	2.91 (1.14)
<i>Cronbach’s alpha</i>	0.89	–
9 Technical Difficulties	2.80 (1.55)	1.55 (0.82)
10 Technical Framework	2.80 (1.48)	2.82 (1.40)

tical significance. Only items 9 ($p=0.04$), 2 ($p=0.07$), and 7 ($p=0.08$) show statistical significance or a trend. Further tests with a larger sample size are therefore necessary.

4.2.1 Working Alliance: Bonding and Agreement on Tasks

The evaluation of the questions in the bonding section shows that the bonding with an average value of $M=3.17$ ($SD=1.39$) is higher for the writing-based coaching chatbot (Rasa) than for the click-based chatbot (Landbot), which can be rated with $M=2.81$ ($SD=1.24$) (see Table 2). In particular, for item 2, “My coachbot and I respect each other”, the writing-based chatbot performs significantly better than the click-based chatbot.

In the questionnaire section on agreement on tasks, the click-based coaching chatbot achieves higher values. The mean here is $M=3.03$ ($SD=1.21$), which is higher than for the writing-based chatbot ($M=2.57$) ($SD=1.36$) (see Table 2). Interestingly, the results for item 4, “The coachbot has made me aware of how I can change” hardly differ, indicating that both chatbot types stimulate change processes. However, the responses for items 5 and 6 suggest that the predefined responses in the click-based chatbot (Landbot) are more likely to trigger change processes.

Our hypothesis that there are differences in the strength of the working alliance depending on the type of interaction method can thus be confirmed—separately for the subscales bonding and task. However, it is tied which interaction method facilitates a stronger working alliance.

4.2.2 Interaction Method and Technical Implementation

In the third questionnaire section, items 7 and 8 asked about the interaction method (clicking vs. writing) of the chat-

bots. Different questions were asked depending on the chatbot that was interacted with. For the item “I like the fact that I can formulate free answers” (writing-based coachbot), 50% of the test subjects answer with “always”, another 20% agree with the statement with “very often” or “often”. For the click-based chatbot, 73% answer that they “rarely” or “sometimes” like the fact that they get predefined answers (see Fig. 4). This shows that the ability to freely formulate answers is preferred in the experiment. On the one hand, Landbot users state that they rather perceive predefined answers as a restriction in the coaching process; on the other hand, Rasa users confirm this tendency, as they do not want predefined answers.

This statement of the writing-based chatbot users is supported by item 8 “I believe that predefined answers would have supported me better in the coaching process.”—the control question for item 7. Here, the subjects answered with a mean of only $M=2.1$ (inverted $M=3.90$, $SD=1.37$). The users of the click-based chatbot answered the question “The suggested answers apply to me” in item 8 with a mean of $M=2.91$ ($SD=1.14$) (see Table 2). This result is in line with the results of the second part of the questionnaire on the agreement on tasks, which indicate that the selected answers from the pretest were appropriate.

Items 9 and 10 asked about the technical feasibility of the respective chatbot systems. Both chatbot systems achieve similar mean scores ($M=2.80$, $SD=1.48$ for Rasa vs. $M=2.82$, $SD=1.40$ for Landbot) on the question “I find the technical framework of a chatbot suitable for coaching support” (Item 10, see Table 2). This indicates that the chatbot coaching is generally well accepted by subjects. However, technical complications occurred with significantly different frequency (Item 9): The click-based chatbot users rarely felt hindered by technical difficulties (64%), while

Fig. 4 Preferred Interaction Method (Writing-based vs. Click-based); $N=21$, Data in Percent

Fig. 5 Technical Difficulties (Writing-based vs. Click-based); $N=21$; Data in Percent

Rasa often, very often, or always led to misinterpretation of statements or general communication problems among users (60%) (see Fig. 5).

4.3 Evaluation of the Chat Histories and Comment Area

The fact that users of the coachbot Rasa encountered technical difficulties was confirmed both by the analysis of the chat histories as well as in the comments section. Rasa users stated that the chatbot's responses were repeatedly steered into incorrect conversational situations (*intents*). Fig. 6 shows a chat excerpt describing such a situation. Here you can see that the chatbot does not know the user's statement (*intent*) "preparation", so it does not understand it and therefore incorrectly responds with a goodbye. This is due to the short learning time in Rasa. Due to the shortness of the project, not enough training data could be collected to optimize the chatbot.

A further comparison of the chat histories in Rasa and Landbot reveals that the subjects repeatedly took the same course of conversation. Even if they rated their exam anxiety as below three, for example, all subjects wanted tips against exam anxiety. This shows that despite seemingly low exam anxiety, the tips are accepted by the users. Some participants also expressed that they would like to receive more personal tips.

In addition, it can be seen in the comment section that subjects expected more answer options from the click-based chatbot. Some students stated that they were not able to express their individual thoughts and others were less engaged in the self-reflection process. This was evident, for

example, when the chatbot asked, "What is the reason for your exam anxiety?" Users of the click-based chatbot had only three choices, whereas ten users of the writing-based chatbot gave eight different reasons for their exam anxiety. Consequently, users in the click-based chatbot were given a choice of causes for their exam anxiety, but they did not question whether the suggested answers were actually mainly responsible for their exam anxiety. One participant noted that users are presented with "thoughts" and were more likely to grapple with which answer choice most closely matched their feelings.

Fig. 6 Wrong Intent Recognition Leading to an Early Exit from Conversation (with English Translations)

5 Discussion: Clicking vs. Writing?

5.1 Working Alliance and Interaction Method

Although the results of not all items are statistically significant, they allow for trends. The results show that the working alliance (bonding and task subscales) in both study conditions (click-based and writing-based) can be classified as medium to high overall. These results can be placed in the literature on AI-based coaching and the working alliance in coaching (see Chap. 2.2 and 2.3): Studies show the effectiveness of AI-based therapy (Fulmer et al. 2018) and that it can stimulate self-reflection processes (Fitzpatrick et al. 2017) as well as forming a bond with clients (Bickmore et al. 2010). Hauser-Ulrich et al. (2020) also show that a working alliance can be established in chatbot interaction with a health care chatbot. Studies on interaction with conversational agents in human-machine interaction further show the establishment of bonding or rapport (Bickmore and Picard 2005; Gratch et al. 2007, 2014; Huang et al. 2011; Zhao et al. 2014).

Various aspects of the chatbot dialog concept may have contributed to the establishment of a working alliance (see Chap. 3.1): During onboarding, the coachbot clarified its tasks and limitations in both experimental conditions. In addition, it reacted with feedback to the answers of the test subjects at some points. Terblanche (2020) sees these as important factors in building a relationship. Graßmann and Schermuly (2021) further highlight the importance of a structured coaching process. Both chatbots in the present study guided the test subjects through such a structured process. Limiting factors for establishing a working alliance may have been the technical implementation difficulties with the writing-based chatbot (see Chap. 5.2).

The results further indicate that a coaching chatbot with free-writing options can create stronger bonding in the working alliance (see Chap. 4.2.1). This result is supported by the statement of the writing-based chatbot users that they found it good to be able to formulate freely and that predefined answer options would not have supported them in the coaching process (see Chap. 4.2.2). In addition, statements in the comment area allow the conclusion that predefined answer options limit the self-reflection process (see Chap. 4.3). The coachees are less concerned with what they would answer (e.g., what the reasons for their exam anxiety are or what strategies they have already developed themselves), but try to select the most suitable options from those given. These results are in line with Hauser-Ulrich et al. (2020), who show in their study that the test subjects would have liked the possibility of more free text input in order to have a more flexible interaction with the chatbot. Literature also supports this by saying that in order to ensure proper handling of clients’ concerns in coaching,

it is necessary that clients disclose (sensitive) information (Graßmann et al. 2019; Graßmann and Schermuly 2016). Moreover, the client’s active participation, as well as their autonomy, is a crucial contributor to the success of coaching (Grant 2014; Lindart 2016). Thus, one can argue that given answer options direct the attention in the coaching process in a certain direction, whereby the self-reflection process remains on a rather superficial level. In line with the media-equation-theory (Reeves and Nass 1998) and computers-as-social-actors paradigm (CASA) (Nass et al. 1994), one can further argue that the desire for a writing-based coaching-chatbot interaction suggests a desire to interact in chatbot coaching in the same way as in human-to-human coaching (Hauser-Ulrich et al. 2020).

With respect to the agreement on tasks, the click-based coaching chatbot was rated better by the test participants (see Chap. 4.2.1). This is interesting insofar as that the coaching interventions of both chatbots (click-based and writing-based) were identical (see Chap. 3.1.). A look into the individual items of this part of the questionnaire leads to the conclusion that especially the predefined answer options of the click-based chatbot were helpful for the coaching process. This is confirmed by the statements of the click-based chatbot users that the predefined answer options applied to the coachees and were obviously selected appropriately in the pretest (see Chap. 4.2.2). The test subjects seem to have felt picked up by the predefined answers. This is further supported by the analysis of the chat histories and the statements in the comment area, which show that the test participants of both chatbot systems would like to receive tips on how to deal with exam anxiety—i.e., more guidance from the coaching chatbot (see Chap. 4.3). This is in line with a study by Terblanche and Cilliers (2020), who show that coaching chatbots are well accepted when they offer practical benefits, and a conceptual paper by Graßmann and Schermuly (2021), who show that coaching chatbots have their strength in the use of a variety of methods. Thus, in a low-threshold self-coaching offer—such as chatbot coaching—given answers can provide an initial orientation in dealing with the issue (in this case: dealing with exam anxiety).

5.2 Technical Implementation

To classify the presented results, the technical realization of the two chatbot system architectures (Rasa vs. Landbot) is fundamental. On the one hand, the results clearly show that both system architectures are perceived as appropriate for a self-coaching process (see Chap. 4.2.2). On the other hand, these systems have to work technically flawlessly to be able to initiate self-coaching processes effectively. Here, the conversational AI Rasa (writing-based system architecture) had major technical implementation problems that

hindered the coaching process and the self-reflection of the test persons (see Chap. 4.2.2). Both the evaluation of the chat histories and the statements in the comment area confirm this (see Chap. 4.3).

Since the productivity of a chatbot system is key to its acceptance (Brandtzaeg and Følstad 2017; Terblanche and Cilliers 2020), the present results on the working alliance must be seen in this light: The technical problems in programming the writing-based chatbot in Rasa are due to the insufficient training time. Using a conversational AI for chatbot coaching is much more complex than using a rule-based system. Training effort and time are significantly higher and longer (Hussain et al. 2019; Lömkker et al. 2021; Ramesh et al. 2017). However, once such a system is appropriately trained, it offers many advantages for the outlined use case compared to a rule-based system. With an increasing amount of training data, the system starts to understand the user's free text input, learns to respond individually to user input and to select different conversation paths, which means that the self-coaching process can be triggered in a more individualized and process-oriented manner.

Thus, it can be surmised that a more stable system architecture will, at a minimum, provide clearer results in terms of bonding in the working alliance. Our hypothesis that there are differences in the strength of the working alliance depending on the type of interaction method can be confirmed—separately for the subscales bonding and task. However, it is tied which interaction method facilitates a stronger working alliance. The research question of this study, “Which coachbot interaction method enables a stronger working alliance between coach and coachee: a click-based or a write-based coaching chatbot?” cannot be answered with either-or but merely with both-as-well.

The idea of “both-as-well” in coachbot design corresponds in this sense to a systemic understanding of coaching, which has the goal of opening up spaces of possibility and expanding possibilities for action (Berninger-Schäfer 2018; Schlippe and Schweitzer 2016). Insofar, the results of this study can be placed in the context of effectiveness factors in chatbot coaching. A study by Mai et al. (2021) investigated the concept of self-disclosure as an effectiveness factor in AI-based coaching (see Chap. 2.3). The results show that self-disclosure, or the similar concept of information disclosure, has a positive effect on students' perceived bonding and thus on the working alliance in coaching. With the choice of interaction method, the present study now investigated another contributing effectiveness factor in the establishment of a working alliance between a coaching chatbot and a human coachee.

5.3 Design and Practical Implications

The findings of this study are highly interesting and relevant for the (further) development of digital (AI-based) self-coaching tools. Instead of an either-or solution in coachbot development (clicking vs. writing), the results of this study suggest developing a “mixed” coachbot that allows interaction via clicking as well as via free writing. Different approaches are conceivable here: As a low-threshold introduction to a coaching conversation, more clickable question answers can play a role at the beginning of the process to give the coachee confidence in chatbot coaching. As the process progresses, the coachbot can move to more open-ended coaching questions that allow free writing options. It is also conceivable to provide answer options or let the coachee formulate freely, depending on the coaching topic or phase.

This results in the following design implications: A key aspect is both the technical and the further conceptual development of the coaching chatbot in conversational AI. The quality of the conversational flow and the intent accuracy increases as the training time progresses. The interaction script can also be further developed through more intensive evaluation of the chat histories available through this study.

From a coach's perspective, it makes sense to use different system architectures when developing a coachbot. The click-based chatbot was explicitly chosen as a no-code or low-code variant to enable people without programming knowledge to work with a chatbot. The results show that reflection processes can also be triggered with a click-based coachbot and interested coaches, consultants and trainers can thus venture a low-threshold introduction to the use of a coaching chatbot.

5.4 Limitations and Future Research

The present study has limitations: The sample of 21 test subjects is very small, and the majority of the *p*-values indicate the need for further testing with a larger sample size. Only three items show a statistical significance or trend. Nonetheless, we assume that the results are initial trends that need to be verified in studies with a larger sample. In addition, the coaching process with the chatbot involves only one interaction. For follow-up studies, the dialog concept of the coachbot should be extended so that multiple (and longer) coaching sessions between chatbot and coachee can be investigated. Furthermore, the sample consisted of students. It remains to be investigated whether the effects on the coaching topic of exam anxiety are also evident in other target groups such as pupils or professionals. Moreover, the influence of the chatbot interaction method should also be investigated for other coaching topics (such as coaching on goal achievement). Last but not

least, the technical functionality of the writing-based chatbot was not mature in some places, which could have an impact on the results.

There are several approaches for follow-up studies that involve expanding the research design: On the one hand, the questionnaire should be expanded to include the goal level of the WAI-SR in order to be able to comprehensively survey the effectiveness of the working alliance. On the other hand, a follow-up study should also apply qualitative research methods so that, for example, guided interviews are conducted on the perceived effectiveness in order to gain deeper insights. The chat histories should also be subjected to a structured analysis in order to draw conclusions about typical conversation paths and to be able to incorporate these into the chatbot design. Finally, the selection of predefined answers in the click-based chatbot system architecture should not only be determined via pretests but should also be developed against a theoretical background, e.g., via empirical studies.

6 Conclusion

This study investigated the influence of the interaction method (clicking vs. writing) in chatbot coaching on the working alliance. Overall, a moderate to good working alliance was established in both chatbot coaching interaction methods (click-based and writing-based). The bonding is higher when using a writing-based system architecture than when using a click-based system. However, the results also show that click-based systems are more helpful as a low-threshold entry point to a coaching process, as they better guide coachees through the process by providing predefined answers. Through further technical and conceptual development, however, these effects could also become visible with conversational AI.

In total, the results of the study indicate that self-reflection processes through digital self-coaching via chatbots are generally well accepted by students. In a next step, the coaching chatbot will therefore be further developed in terms of both its technical functionality and its conception in conversational AI. In studies with a larger sample size and multiple coaching sessions, the influence of a writing-based, click-based, and “mixed” interaction will be further investigated.

Consideration will also be given to how coaching chatbots will be applicable in specific contexts. The target group of students seems to be open to the use of AI-based coaching. It is conceivable, for example, to provide benefits in the context of study counseling/student advisory service and/or for the preparation and follow-up of seminars for students on relevant topics such as exam anxiety, learning and work strategies, project management, and communication.

Acknowledgements This journal contribution is based on a project carried out as part of the module “Research Seminar” in the master’s program Mechanical Engineering at TH Köln/University of Applied Sciences. We kindly thank the master’s students Fabian Gerz, Markus Hakschteter and Christina Schneider for carrying out this project.

Conflict of interest V. Mai, C. Neef and A. Richert declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albrecht, A. (2016). *Student Coaching—Selbstentwicklung mit professioneller Unterstützung*. MBS insights. <https://www.munich-business-school.de/insights/2016/student-coaching/>. Accessed 25 Sept 2021.
- Barton, T., & Müller, C. (2021). *Künstliche Intelligenz in der Anwendung*. Springer.
- Berninger-Schäfer, E. (2018). *Online-Coaching*. Springer.
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2), 293–327. <https://doi.org/10.1145/1067860.1067867>.
- Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24, 648–666.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252–260. <https://doi.org/10.1037/h0085885>.
- Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. *Internet Science. Lecture Notes in Computer Science*, 10673, 377–392. https://doi.org/10.1007/978-3-319-70284-1_30.
- Brandão, T. K., & Wolfram, G. (2018). *Digital Connection: Die bessere Customer Journey mit smarten Technologien – Strategie und Praxisbeispiele*. Springer.
- De Gennaro, M., Krumhuber, E. G., & Lucas, G. (2020). Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Frontiers in psychology*, 10, 3061. <https://doi.org/10.3389/fpsyg.2019.03061>.
- Deepika, K., Tilekya, V., Mamatha, J., & Subetha, T. (2020). *Jollity chatbot—a contextual AI assistant*. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). (pp. 1196–1200). <https://doi.org/10.1109/ICSSIT48917.2020.9214076>.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health*, 4(2), e19. <https://doi.org/10.2196/mental.7785>.
- Følstad, A., & Brandtzaeg, P. B. (2017). Chatbots and the new world of HCI. *interactions*, 24(4), 38–42.
- Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different chatbots for different purposes: towards a typology of chatbots to

- understand interaction design. *Internet Science. Lecture Notes in Computer Science*, 11551, 145–156. https://doi.org/10.1007/978-3-030-17705-8_13.
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, 5(4), e64. <https://doi.org/10.2196/mental.9782>.
- Grant, A. M. (2014). Autonomy support, relationship satisfaction and goal focus in the coach-coachee relationship: which best predicts coaching success? *Coaching: An International Journal of Theory, Research and Practice*, 7(1), 18–38.
- Graßmann, C., & Schermuly, C. (2016). Side effects of business coaching and their predictors from the coachees' perspective. *Journal of Personnel Psychology*, 15(4), 152–163.
- Graßmann, C., & Schermuly, C. C. (2021). Coaching with artificial intelligence: concepts and capabilities. *Human Resource Development Review*, 20(1), 106–126.
- Graßmann, C., Schölmerich, F., & Schermuly, C. C. (2019). The relationship between working alliance and client outcomes in coaching: a meta-analysis. *Human Relations*, 73(1), 35–58.
- Gratch, J., Lucas, G. M., King, A. A., & Morency, L. P. (2014). It's only a computer: The impact of human-agent interaction in clinical interviews. In *Proceedings of the 2014 international conference on autonomous agents and multi-agent systems* (pp. 85–92). <https://doi.org/10.5555/2615731.2615748>.
- Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. *Lecture Notes in Computer Sciences*, 4722, 125–138. https://doi.org/10.1007/978-3-540-74997-4_12.
- de Haan, E., Grant, A. M., Burger, Y., & Eriksson, P.-O. (2016). A large-scale study of executive and workplace coaching: the relative contributions of relationship, personality match, and self-efficacy. *Consulting Psychology Journal: Practice and Research*, 68(3), 189–207. <https://doi.org/10.1037/cpb0000058>.
- Hauser-Ulrich, S., Künzli, H., Meier-Peterhans, D., & Kowatsch, T. (2020). A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. *JMIR mHealth and uHealth*, 8(4), e15806. <https://doi.org/10.2196/15806>.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *J Counsel Psychol.*, 36(2), 223–233. <https://doi.org/10.1037/0022-0167.36.2.223>.
- Huang, L., Morency, L. P., & Gratch, J. (2011). Virtual Rapport 2.0. *Lecture Notes in Computer Science*, 6895, 68–79. https://doi.org/10.1007/978-3-642-23974-8_8.
- Hussain, S., Ameri Sianaki, O., & Ababneh, N. (2019). A survey on conversational agents/chatbots classification and design techniques. *Web, Artificial Intelligence and Network Applications. Advances in Intelligent Systems and Computing.*, 927, 946–956. https://doi.org/10.1007/978-3-030-15035-8_93.
- Joshi, N. (2020). Choosing Between Rule-Based Bots And AI Bots. Forbes. <https://www.forbes.com/sites/cognitiveworld/2020/02/23/choosing-between-rule-based-bots-and-ai-bots/>. Accessed 25 Sept 2021.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.
- Kanatouri, S. (2020). *The digital coach*. Routledge.
- Kellner, A. (2001). Dialogsysteme. In K.-U. Carstensen, C. Ebert, C. Endriss, S. Jekat, R. Klabunde & H. Langer (Eds.), *Computerlinguistik und Sprachtechnologie: Eine Einführung* (pp. 115–138). Spektrum.
- Kiluk, B. D., Serafini, K., Frankforter, T., Nich, C., & Carroll, K. M. (2014). Only connect: the working alliance in computer-based cognitive behavioral therapy. *Behav. Res. Ther.*, 63, 139–146. <https://doi.org/10.1016/j.brat.2014.10.003>.
- Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The rise of bots: a survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 conference on designing interactive systems* (pp. 555–565). <https://doi.org/10.1145/3064663.3064672>.
- Künzli, H. (2019). Spielstand 1:0 – Die Wirksamkeit von Coaching. In A. Ryba & G. Roth (Eds.), *Coaching und Beratung in der Praxis: Ein neurowissenschaftlich fundiertes Integrationsmodell* (pp. 102–124). Klett-Cotta.
- Landbot (2020). Knowledge center. <https://landbot.io/knowledge-center>. Accessed 25 Sept 2021.
- Lee, N., Madotto, A., & Fung, P. (2019). Exploring social bias in chatbots using stereotype knowledge. In *Proceedings of the 2019 workshop on widening NLP* (pp. 177–180).
- Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020). “I hear you, I feel you”: encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–12). <https://doi.org/10.1145/3313831.3376175>.
- Lindart, M. (2016). *Was Coaching wirksam macht: Wirkfaktoren von Coachingprozessen im Fokus*. Springer. <https://doi.org/10.1007/978-3-658-11761-0>.
- Lippmann, E. (2013). *Coaching: Angewandte Psychologie für die Beratungspraxis*. Springer.
- Lömker, M., Moskaliuk, J., & Weber, U. (2021). *Chatbots im Coaching. essentials*. Springer. https://doi.org/10.1007/978-3-658-32830-6_22.
- Lovejoy, J. (2019). The UX of AI. <https://design.google/library/ux-ai/>. Accessed 15 Dec 2021.
- Mai, V. (2020). Projektcoaching und Leadership-Coaching als integrative Elemente in der Ingenieurausbildung. Eine Wirksamkeitsstudie zur Entwicklung von Metakompetenzen. In S. Heuchemer, R. Hochmuth, N. Schaper, & B. Szczyrba (Eds.), *Forschung und Innovation in der Hochschulbildung*, 8, Research Paper. <https://cos.bibl.th-koeln.de/frontdoor/index/index/docId/926>. Accessed December 15, 2021.
- Mai, V., & Richert, A. (2020). AI Coaching: Effectiveness factors of the working alliance in the coaching process between coachbot and human coachee – an explorative study. In I. Gómez Chova & A. López Martínez (Eds.), *EDULEARN20 Proceedings* (pp. 1239–1248). <https://library.iated.org/view/MAI2020AIC>. Accessed September 25, 2021.
- Mai, V., & Richert, A. (2021). StudiCoachBot an der TH Köln - Reflexionsprozesse KI-basiert begleiten. *fnma Magazin – Forum neue Medien in der Lehre Austria* (01/2021), pp. 21–23. <https://www.fnma.at/publikationen/magazin>. Accessed September 25, 2021.
- Mai V., Wolff A., Richert A., Preusser I. (2021). Accompanying Reflection Processes by an AI-Based StudiCoachBot: A Study on Rapport Building in Human-Machine Coaching Using Self Disclosure. In *HCI International 2021 – Late Breaking Papers: Cognition, Inclusion, Learning, and Culture. Lecture Notes in Computer Science*. 439–457. https://doi.org/10.1007/978-3-030-90328-2_29.
- Nass, C., Steuer, J., & Tauber, E. (1994). Computers are social actors. In *SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). <https://doi.org/10.1145/191666.191703>.
- Object Management Group (2005). Unified modelling language. <https://www.omg.org/spec/UML/2.0/About-UML/>. Accessed 15 Dec 2021.
- Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017). A survey of design techniques for conversational agents. *Communications in Computer and Information Science*, 750, 336–350.
- Reeves, B., & Nass, C. (1998). *The media equation: how people treat computers, television, and new media like real people and places*. CSLI Publ.
- Reindl, R., Hergenreider, M., & Hünninger, J. (2012). Schriftlichkeit in virtuellen Beratungssettings. In H. Geißler & M. Metz (Eds.),

- E-Coaching und Online-Beratung* (pp. 339–357). VS. https://doi.org/10.1007/978-3-531-19155-3_20.
- Reis, (2006). Gruppencoaching für Lehramtsstudierende – Ein leistungsfähiger Beitrag zur Studienreform. In J. Wildt, B. Szczyrba & B. Wildt (Eds.), *Consulting. Coaching. Supervision. Eine Einführung in Formate und Verfahren hochschuldidaktischer Beratung* (pp. 158–173). Bertelsmann.
- Schlippe, A., & Schweitzer, J. (2016). *Lehrbuch der systemischen Therapie und Beratung I. Das Grundlagenwissen* (3rd edn.). Vandenhoeck & Ruprecht.
- Schumacher, E.-M. (2006). LernCoaching. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer & J. Wildt (Eds.), *Neues Handbuch Hochschullehre (Griffmarke A 3.5)*. DUZ Medienhaus.
- Shah, H., Warwick, K., Vallverdú, J., & Wu, D. (2016). Can machines talk? Comparison of ELIZA with modern dialogue systems. *Comput. Hum. Behav.*, 58, 278–295.
- Sharpley, C.F. (1997). The influence of silence upon client-perceived rapport. *Counselling Psychology Quarterly*, 10(3), 237–246. <https://doi.org/10.1080/09515079708254176>.
- Shevat, A. (2017). *Designing bots: creating conversational experiences*. O’Reilly Media.
- Singh, S., & Thakur, H. K. (2020). Survey of various AI chatbots based on technology used. In *8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 1074–1079). <https://doi.org/10.1109/ICRITO48877.2020.9197943>.
- Sue-Chan, C., & Latham, G.P. (2004). The relative effectiveness of external, peer, and self-coaches. *Applied Psychology*, 53, 260–278.
- Sugisaki, K., & Bleiker, A. (2020). Usability guidelines and evaluation criteria for conversational user interfaces: a heuristic and linguistic approach. In *Proceedings of the Conference on Mensch und Computer (MuC ’20)* (pp. 309–319). <https://doi.org/10.1145/3404983.34055055>.
- Terblanche, N. (2020). A design framework to create artificial intelligence coaches. *International Journal of Evidence Based Coaching & Mentoring*, 18(2), 152–165. <https://doi.org/10.1109/ICRITO48877.2020.919794310.24384/b7gs-3h05>.
- Terblanche, N., & Cilliers, D. (2020). Factors that influence users’ adoption of being coached by an artificial intelligence coach. *Philosophy of Coaching: An International Journal*, 5(1), 61–70.
- Thanaki, J. (2018). *Machine learning solutions: expert techniques to tackle complex machine learning problems using python*. Packt Publishing Ltd.
- Traus, A., Höffken, K., Thomas, S., Mangold, K., & Schröer, W. (2020). *Stu.di.Co. – Studieren digital in Zeiten von Corona*. Hildesheim: Universitätsverlag. <https://doi.org/10.18442/150>.
- Vu, M. (2021). VuMA Touchpoints monitor. <https://touchpoints.vuma.de/#/zielgruppen/>. Accessed 25 Sept 2021.
- Wasylyshyn, D.K.M. (2003). Executive coaching: an outcome study. *Consulting Psychology Journal: Practice and Research*, 55(2), 94–106. <https://doi.org/10.1037/1061-4087.55>.
- Wiemer, M. (2012). Begleitung anspruchsvoller Bildungswege: Coaching für Studierende. *Organisationsberatung, Supervision, Coaching*, 19(1), 49–57. <https://doi.org/10.1007/s11613-012-0271-3>.
- Wilmers, F., Munder, T., Leonhart, R., Herzog, T., Plassmann, R., Barth, J., & Linster, H. (2008). Die deutschsprachige Version des Working Alliance Inventory – Short revised (WAI-SR) – Ein schulübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz. *Klinische Diagnostik & Evaluation*, 1(3), 343–358.
- Wolf, C. (2017). Konzeption und Untersuchung eines technologiegestützten Selbstcoachings als Intervention zur Förderung von ergebnisorientierter Selbstreflexion bei Studierenden am Karlsruher Institut für Technologie (KIT). <https://d-nb.info/1138708682/34> Dissertation, Karlsruher Institut für Technologie. Accessed December 15, 2021.
- You, Y., & Gui, X. (2020). Self-diagnosis through AI-enabled chatbot-based symptom checkers: user experiences and design considerations. In *AMIA Annual Symposium Proceedings* (pp. 1354–1363).
- Zhao, R., Papangelis, A., & Cassell, J. (2014). Towards a dyadic computational model of rapport management for human-virtual agent interaction. *Intelligent Virtual Agents*, 8637, 514–527. https://doi.org/10.1007/978-3-319-09767-1_62.
- Zhou, L., Jianfeng, G., Li, D., & Shum, H.-Y. (2020). The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *Computational Linguistics*, 46(1), 53–93. https://doi.org/10.1162/coli_a_00368.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.