Journal of Management Development

Unearthing AI coaching chatbots capabilities for professional coaching: a systematic literature review

Lidia Plotkina and Subramaniam Sri Ramalu
College of Business, Universiti Utara Malaysia, Sintok, Malaysia

Received 5 June 2024
Revised 20 August 2024
Accepted 5 September 2024

Abstract

Purpose – Recent advances in coaching technology enhanced its accessibility and affordability for a broader population. In the imposing growth of economy and the demand for extensive coaching intervention for executives, artificial intelligence (AI)-based coaching is one of the possible solutions. While the evidence of AI coaching effectiveness is expanding, a comprehensive understanding of the field remains elusive. In particular, the true potential of AI coaching tools, ethical considerations and their current functionality are subjects of ongoing investigation.

Design/methodology/approach – The systematic literature review was conducted to extract experimental results and concepts about utilizing AI in coaching practice. The paper presents the primary capabilities of state-of-the-art coaching tools and compares them with human coaching.

Findings – The review shows that AI coaching chatbots and tools are effective for narrow tasks such as goal attainment, support for various psychological conditions and induction of reflection processes. Whereas, deep long-term coaching, working alliance and individualized approach are out of current AI coaching competence. In the current state, AI coaching tools serve as complementary helping tools that cannot replace human coaching. However, they have the potential to enhance the coach's performance and serve as valuable assistants in intricate coaching interventions.

Originality/value – The review offered insights into the current capabilities of AI coaching chatbots, aligned with International Coaching Federation set of competencies. The review outlined the drawbacks and benefits of chatbots and their areas of application in coaching.

Keywords Executive coaching, Coaching chatbot, Virtual coach, AI coaching, Coaching effectiveness

Paper type Literature review

1. Introduction

Coaching practice is transforming drastically and moving into more digital forms of interaction. Taking into account the rising need for cost-effective and affordable coaching in the growing market reality, artificial intelligence (AI) coaching and distance coaching, which uses various technologies, also recognized as e-coaching or virtual coaching could be a possible solution. The massive growth of technological interventions in executive coaching appeared due to the plethora of factors and include the simplification of logistics, cost-savings, time-management and global use of technology by the organizations (Ribbers and Waringa, 2015). Overall technological progress, with particular strong acceleration due to COVID-19 pandemic impact, has expedited the progressive use of online technological tools and significantly reduced face-to-face within the coaching landscape (Doolittle, 2023; Terblanche, 2022). The recent COVID-19 pandemic crisis had a strong influence and

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Lidia Plotkina. The first draft of the manuscript was written by Lidia Plotkina and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Supervision was performed by Dr Subramaniam Sri Ramalu.

Funding: No funds, grants or other support was received.

Conflict of interest: The authors have no relevant financial or non-financial interests to disclose.

Journal of Management Development

0262-1711
DOI 10.1108/JMD-06-2024-0182

accelerated the shift of coaching intervention into the online realm (Carnevale and Hatak, 2020; ICF, 2020). During the pandemic, in-person coaching reduced by 80%, and audio-video means coaching increased by 74% (ICF, 2020). After the pandemic, the perspective of active engagement in online coaching interaction with AI became more supported by users than before (Schermuly *et al.*, 2022). The expectations of expansive use of technologies for coaching and belief that AI could simplify the process of coaching were reflected in the global surveys.

Technological advancements in coaching are progressing rapidly, while comprehensive experimental research has not yet fully developed. Contemporary trends in the field increasingly incorporate e-coaching and AI-based coaching methodologies. E-coaching is a format of coaching practice implemented via different technologies, so that it can be done distantly between a coach practitioner and a coachee. It usually simplifies the procedure and it is more cost-effective (Diller and Passmore, 2023). Meanwhile, AI Coaching is a synchronous or asynchronous coaching using AI as a coach instead of a human coach (Passmore and Tee, 2023a). Several narrow task coaching chatbots were created and tested for efficiency and appreciation from the coachee's side (Graßmann and Schermuly, 2021; Movsumova *et al.*, 2020; Terblanche *et al.*, 2022a). Particular coaching tasks can already be delegated to AI-based coaching chatbots, such as goal attainment, new health protocol implementation support, education related coaching and more (Chew, 2022; Kocaballi *et al.*, 2019; Mai *et al.*, 2021; Mitchell *et al.*, 2021).

However, the principal shift in application of AI and machine learning models happened with the launching of new multimodal large language models, such as generative pre-trained transformer (GPT). These models allow to create sophisticated and complex dialogs providing support and guidance in various fields (Carlbring *et al.*, 2023; Lee *et al.*, 2023). Thus, GPT-based chatbots are different from the scripted and rule based chatbots that were previously used. These tools can be identified by the following primary attributes: firstly, they are designed for general-purpose applications rather than specialized ones; secondly, they possess the capability to generate innovative language outputs that closely resemble human communication; and thirdly, they provide a user-friendly interface that comprehends and responds to natural language. Due to the promising preliminary results, the boom of the AI-based coaching chatbot publications is expected. At the same time, chatbots usage raises numerous ethical and privacy related questions, as well as the distribution of responsibility and their efficiency (Cabrera *et al.*, 2023). Meanwhile, the objective of this review is to analyze existing approaches and AI-based coaching performance results, with the primary goal to review AI-based coaching solutions and to compare the capabilities of AI tools with human coaching.

2. Methodology

The purpose of this study is to conduct a systematic literature review on AI coaching. To ensure impartiality, specific criteria were established at the beginning to determine which articles should be included or excluded. Only high-quality and relevant studies from reputable academic databases were searched and considered. A careful screening process was implemented to ensure that the final selection of studies for the literature review aligns with the research topic. The review is to provide the answers to the following questions:

Q1. What are the existing approaches or types of AI coaching?

Q2. What are the empirical evidences of AI coaching tools and chatbots implementation?

Q3. What advantages and disadvantages of AI coaching are derived from existing studies compared with human coaching?

2.1 Selected review method

The systemic literature review aims to evaluate and compare the findings of different studies and identify the gaps in the existing literature. To determine this, a systemic literature review is conducted where studies were collected from academic databases like Scopus, Academia, ResearchGate, Web of Science, and Google Scholar. These studies were considered with their relevance to the current topic. To facilitate this, the titles and abstracts of the studies were carefully examined. Systematic literature review based on the PRISMA diagram tool consists of different stages and uses inclusion and exclusion criteria. Thus, the sources of the studies have been identified, as well as the key words that are relevant for this research, the timeline of this research, and actual adequacy of the studies in terms of their content. Precise inclusion and exclusion criteria are listed below, alongside with the stages of identification, selection, eligibility and final inclusion of the studies.

2.1.1 Inclusion criteria.

(1) Studies published between 2019 and 2023 were considered for this review as an inclusion criterion.

(2) The studies included conceptual papers and empirical results on AI coaching implementation.

2.1.2 Exclusion criteria.

(1) Studies that were not focused on coaching and contain other AI chatbots implementation.

(2) The unavailability of the full text of the study. Studies with solely available abstracts were excluded, due to the need to analyze the entire document to get full and deep understanding of the research works.

(3) Thesis research studies were excluded in preference for the examination of peer-reviewed journal submissions exclusively.

2.2 Data collection and processing methods

2.2.1 Stage I identification. All of the relevant to the topic studies were identified in the academic database, such as Scopus, ResearchGate, Google Scholar, Academia, and Web of Science. The keywords "coaching chatbot, AI coaching, e-coaching effectiveness, Artificial Intelligence, GPT-4" were used for the literature search with the total number of final studies of 339.

2.2.2 Stage II selection. Original studies were selected by eliminating duplicate studies due to necessity as multiple academic databases were used in this study and hence the same study can be found in more than one database. After the elimination of duplicate studies, the number of remaining original studies was 175. These 175 studies were carefully analyzed with regards to their titles and abstract for relevance. 84 more studies were excluded as they were not relevant to the current research objective with post elimination total number of 91 remaining studies.

2.2.3 Stage III eligibility. The availability of academic literature can vary depending on the access mode of the studies. For studies with restricted access, only the abstract is available. However, the abstract alone is not sufficient for a full understanding of the study. Thus, 17 studies were additionally excluded with the remaining number of 74 studies, that were fully available after this exclusion.

2.2.4 Stage IV inclusion. The remaining 74 studies were analyzed based on predetermined inclusion and exclusion criteria for the study. The studies that do not fit in inclusion criteria were excluded from the systematic literature review. Hence, 51 more studies were excluded

based on that, to specify: not about coaching, but using chatbots - 20, about coaching, but no AI component – 30, thesis work - 1. The final number of studies that were considered for this systemic literature review is 23, which will be considered while analyzing the findings in the literature. According to the most recent timeline, the strict criteria was selected and applied to the studies in this review, all available and relevant literature on the implementation of AI coaching chatbots in different contexts was included. Although 23 studies represent a relatively small amount of evidence as the field of AI coaching is still emerging and growing. Therefore, more theoretical and practical research on this topic is expected to appear in the nearest future. Moreover, according to systematic literature review guidelines, there is no proven minimum number of studies included (Kitchenham and Charters, 2007), and typically SLRs could include from 10 and up to 50 articles. The former research works show the existence of SLRs based, for instance, on 6 studies, and they managed to provide valuable insights in the respective fields (Wang *et al.*, 2023).

2.2.5 PRISMA flow diagram. PRISMA stands for "Preferred Reporting Items for Systemic Review and Meta-Analysis" and can be characterized as a base arrangement of things that can be utilized for meta-examination and foundational surveys. For the purpose of this study, a PRISMA flow diagram (Figure 1) is created for a better understanding of how studies are identified, included, and excluded from the systemic literature review. The PRISMA diagram is essential for transparency, as it allows to see the method of studies' selection and ensures that the review process was systematic and replicable.

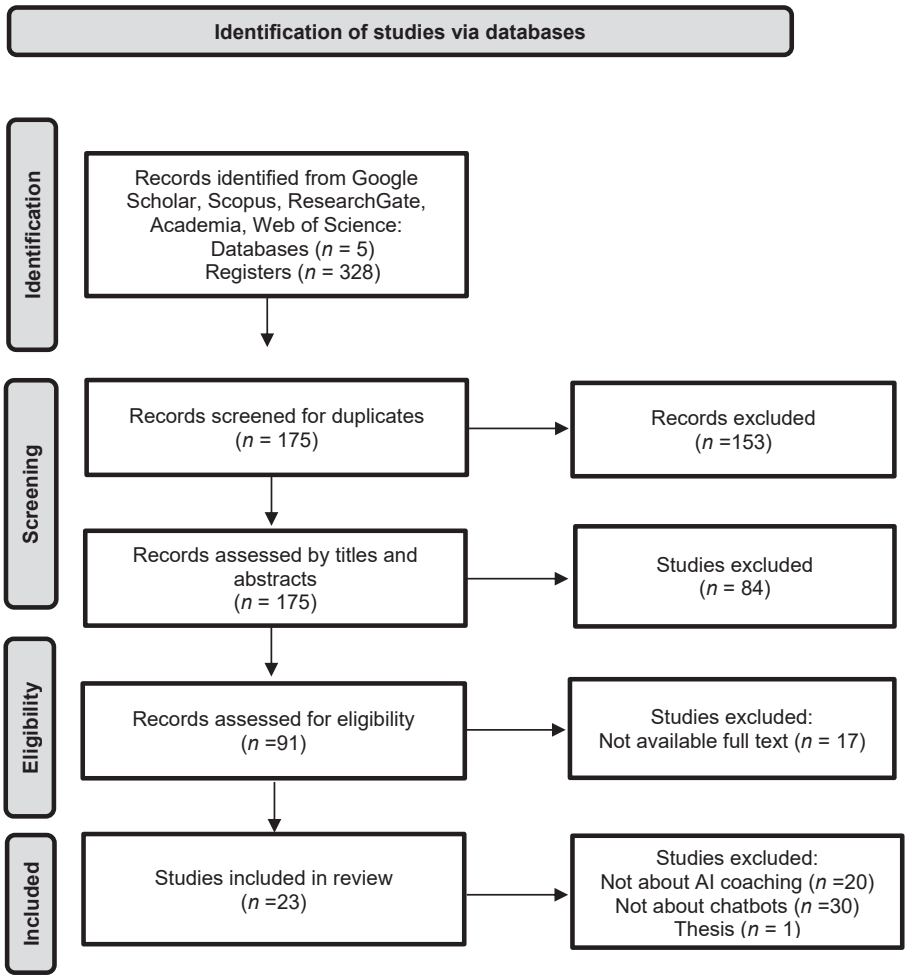
3. Results

AI systems could serve the coaching intervention in multiple ways (Strong and Terblanche, 2020). Machine Learning techniques allow to improve the process of coaching and could help in coach-coachee matching. They could provide additional recommendations and resources in between coaching sessions as well as the support for the goal attainment and coaching advice (Khandelwal and Upadhyay, 2021a, b; Movsumova *et al.*, 2020; Passmore and Tee, 2023a). The expansive scope of AI demonstrates that it offers more than just synchronous coaching solutions. However, not all applications of AI in coaching are well-studied yet. In this review, we focus on all the available studies about AI coaching and review their performance, as well as current strengths and weaknesses.

3.1 AI-based coaching software and coach support

One of the useful AI-based software was suggested by Arakawa and Yakura (2020). A specialized software was developed to analyze recorded coaching sessions. This software provides a user-friendly interface that allows coaches to review sessions at an accelerated pace and selectively access important parts of the conversation. Coaches can also take notes and offer meta-reflection to the coachee. In summary, this tool simplifies the process of video analysis and has shown to enhance the effectiveness of coaching sessions. In addition, a fully automated system was developed to assist coach practitioners in identifying unconscious behaviors exhibited by their coachees. This system utilizes an unsupervised anomaly detection algorithm that analyzes multimodal behavior data, including posture and gaze. By providing real-time feedback, this tool alerts coaches to relevant behavioral cues. The algorithm generates informative cues that help coaches gain insight about the internal states of the coachees. Furthermore, the use of an unsupervised machine learning algorithm ensures that personal biases are avoided while an effective coaching support is provided (Arakawa and Yakura, 2019, 2022).

Another AI-based coaching assistant was tested for its ability to provide help and insights for the coaches and coachees (Movsumova *et al.*, 2020). The Mentorbot appeared to



Source(s): Authors' own creation/work

Figure 1. PRISMA flow chart on inclusion and exclusion criteria

be helpful in suggesting the proper questions and provided deep and high-quality examples of questions. The findings of the research indicate that an AI-based tool is more effective in addressing new requests that are of high importance to the coachees in terms of their willingness to take action and gain clarity. Whereas in terms of overall perception, the real coach is seen as more useful, effective, and capable in facilitation of stress-management focused sessions. The choice of the AI assisted tool versus human coach was more preferrable by the coachees in cases of sensitive and confidential requests (Movsumova *et al.*, 2020). The increased reliance on machines might depend on the level of placing trust in the privacy of the data exposed and shared during coaching sessions (Movsumova *et al.*, 2020). Additionally, the combination of coach and AI chatbot reveals that AI expands the coaches' vision (Movsumova *et al.*, 2020).

3.2 Performance of AI coaching chatbots

There is a noticeable advancement in the development of synchronous and asynchronous coaching applications that offer a wide range of services (Passmore and Diller, 2024). Different types of approaches could be utilized as performance algorithms of AI coaching chatbot. For example, simple rule-based approach, scripted approach or reinforcement learning data driven generation of coaching dialogs. Recently appeared Natural Language Processing (NLP) based GPT models are the most sophisticated among them (Passmore and Tee, 2023a). In the study conducted by Mitchell *et al.* (2022), the authors compared the effectiveness of three approaches: scripted, rule-based and Reinforcement Learning (RL) formulated interaction of coaching chatbot with a coachee about the health-related goals' attainment like nutrition. The findings revealed that the data-driven RL chatbot performed well in short conversations. Unexpectedly, the simplest scripted chatbot received higher ratings in terms of quality, despite not consistently fulfilling its intended purpose. These results underscore the conflicting nature between scripted and more intricate data-driven approaches when it comes to chatbots in the healthcare domain. Therefore, there are cases when the narrow tasks are executed more effectively with a straightforward scripted chatbot rather than relying on complex yet inaccurate reinforcement learning responses and prompts. The future direction of research might include the comparison of GPT based responses and scripted responses in chatbots related to healthcare coaching.

The prominence of AI-driven healthcare coaching is on the rise, primarily attributed to its cost-effectiveness and the round-the-clock availability of AI tools to assist coachees in pursue of their goals. In the process of dealing with obesity and managing weight loss the support, special recommendations and coaching are exceptionally helpful. Reinforcement learning algorithms contributed to effective diabetes related health coaching intervention (Di *et al.*, 2022). To explore the factors that facilitate engagement of coachees in health coaching interaction, the scoping review was conducted (Chew, 2022). The primary functions of AI chatbots encompass the delivery of personalized recommendations, motivational messages, gamification elements, and emotional support. According to the research, the speech with appropriate colloquial tones in chatbots increased user engagement, due to convenience of hands-free interaction, interactivity and expression of empathy in voice tones. Moreover, additional strategies employed in text-based chatbots included emojis mimicking human emotional expressions, the incorporation of positively framed vocabulary, referencing credible information sources, personifying the chatbot, offering validation, and offer of real-time, rapid, and valid recommendations (Chew, 2022). Moreover, ability to interact with a chatbot on different platforms and devices promotes greater engagement. However, the privacy concerns and user friendliness were the main constrains in using health AI coaching chatbot for weight management. Concerning the "user burden", researchers usually refer to several constructs, such as difficulty to use a tool, the problems with social interaction, physical, emotional and mental load while using a tool, time and finances that are spent while using a tool (Suh *et al.*, 2016).

3.3 AI coaching chatbots working alliance

Empathy and supporting attitude of the coach seemed to be important in real coaching sessions (Cidral *et al.*, 2023). However multiple studies of coaching chatbots show the presence of coaching goal attainment with absence of empathetic behavior from a chatbot. A comparison was made between the performance of a machine-driven chatbot and human-led chatbot coaching. While the human-led conversations showed more empathy and support from the coach, the coaches faced significant challenges and frustration when trying to adapt to text-based virtual coaching. On the other hand, the scripted chatbot remained effective, consistent, and encouraged the coachees to be more autonomous (Mitchell *et al.*, 2021).

Another study examined the impact of human coaching and AI coaching on goal achievement over a period of 10 months (Terblanche *et al.*, 2022a). The results revealed that both human and AI coaches were significantly more successful in assisting coachees in attaining their goals compared to control groups. Surprisingly, towards the end of the trial period, the AI coach demonstrated comparable effectiveness to the human coach. As a result of their findings, the authors suggested that AI could potentially replace human coaches who employ simplistic, model-based coaching approaches. However, it is important to note that AI performed poorly in terms of empathy and emotional intelligence, which are essential qualities of the human coaches that make them irreplaceable at this moment (Terblanche *et al.*, 2022a).

At the same time, the attempt to enhance rule-based coaching chatbot with Reinforcement Learning (RL) power in emotions recognition was done (Alazraki *et al.*, 2021). The idea of making a machine more empathetic and emotionally intelligent is based on the evidence of real coaching sessions. This factor plays an essential role in working alliance establishing between the coach and the coachee (Albizu *et al.*, 2019). In the study, the authors introduced methodology and computational approach for a digital coach designed to assist coachees in implementing self-attachment therapy protocols. Their findings indicate that the platform consistently receives higher ratings for empathy, user engagement, and overall usefulness when compared to a basic rule-based framework (Alazraki *et al.*, 2021). Overall, scripted chatbots are effective for narrow and straightforward tasks, even though they barely could create a working alliance with a coachee or build trustful and deep coach–coachee relationship.

The creation of working alliance appeared to be salient in coaching sessions and this ability is low in current generation of chatbots (Graßmann *et al.*, 2019; Mai *et al.*, 2022). To further investigate the working alliance in a chatbot–coachee relationships, chatbot's self-disclosure and information disclosure approaches were tested (Mai *et al.*, 2021). The authors used the StudiCoachBot that provides coaching with regard to exam-related anxiety. The study focused on testing whether the chatbot's self-disclosure would enhance students' self-disclosure and the effectiveness of coaching. The outcomes contradicted the primary hypothesis, revealing that information disclosure was more effective than machine self-disclosure for students. Essentially, comprehensive information about certain processes appears sufficient, and the simulation of "human-like" issues by a machine is not necessary. Whereas, overall experience of using chatbot was positively rated. It resulted in facilitation of the reflection processes and students had the intention to use the bot again. The aspects requiring further development were highlighted, including that chatbot lacks the flexibility in answers and individualized approach (Mai *et al.*, 2021). The follow-up study extended the research about self-disclosure and information disclosure from chatbot with 201 participants experiencing online coaching for study (Mai *et al.*, 2022). They did not find statistically significant differences with regards to presence or absence of any type of disclosure. Consequently, the foundation of the working alliance between the coach chatbot and the coachee is facilitated through the influence of additional factors that require further investigation in forthcoming researches.

Another valuable investigation of working alliance between coaching chatbot and coachee was conducted by Ellis-Brush (2021). The primary objective of the research was to investigate the association between AI coaching and its potential for enhancing a coachee self-resilience and working alliance initiation. Over the course of eight weeks, a group of 48 volunteers was granted access to WYSA, a mental well-being AI chatbot application. WYSA is a downloadable software application chatbot accessible through smartphones. It utilizes AI, natural language processing, and machine learning algorithms. It employs carefully designed algorithms to generate conversations that mimic interactions with humans, all with the aim of enhancing an individual's self-resilience. WYSA achieves this by employing evidence-based and validated tools and techniques, including cognitive behavioral therapy.

The synthesis of both quantitative and qualitative data revealed important findings. Although there was no substantial working alliance with the AI chatbot, the majority of participants showed a notable improvement in self-resilience (Ellis-Brush, 2021). Drawing upon multiple pieces of evidence, favorable outcomes could be discerned that are associated with AI coaching even in the absence of a working alliance with a chatbot.

3.4 Factors affecting willingness to use coaching chatbots

The factors that influence the willingness to use coaching chatbots, and consequently impact the effectiveness of AI coaching, are currently under investigation (Terblanche and Kidd, 2022). The chatbot was designed to focus on goal achievement, self-reflection, and non-directive coaching. Terblanche and Kidd (2022) examined several factors that influence users' willingness to engage with AI coaching. These factors include users' expectations of the chatbot's performance, the effort required to use it, the perceived impact of AI coaching, potential risks, and the conditions that support its use. Among the five factors studied, performance expectancy, facilitating conditions, and social influence were found to have a statistically significant impact on users' intent to use the chatbot. However, effort expectancy and perceived risk did not show a significant influence.

To explore user experience and engagement in the coaching session, virtual conversational bot was created based on Amazon's ALexa platform (Kannampallil *et al.*, 2022). In the study, the authors developed a voice-based coach named Lumen, designed to deliver an evidence-based problem-solving treatment program for depression and anxiety. While participants emphasized the nonhuman aspect of the interaction, citing the absence of variations in tone, emotion, instant feedback; almost all participants acknowledged the potential benefits associated with Lumen's accessibility. This allows individuals in need of therapy to easily access a coach bot at any time. Participant highlighted the ways to reduce the cognitive load and time pressure of interaction with coach chatbot so that sessions could become more comfortable. The study gives insights about certain improvements that should be taken into account while creating voice coaching tools and shows the potential of developing the bond and understanding between the tool and the user (Kannampallil *et al.*, 2022). The authors suggest to use short and slow conversations, and allow to repeat and pause the session. The primary challenges in achieving a natural conversational feel and engagement revolve around the current limitations in language recognition and emotion recognition systems, which, although promising, are still in the process of development.

Besides functions, the virtual appearance of the chatbot matters and influences its overall success (Weber *et al.*, 2021). The study was aimed to compare anthropomorphic chatbot with less human-like appearing interface for goal attainment chatbot. The analysis revealed a notable increase in satisfaction among users when interacting with the highly anthropomorphic chatbot. Participants rated both the chatbot's ability to build relationship and the effectiveness of coaching more favorably compared to their interactions with the less anthropomorphic chatbot. Consequently, it is evident that the level of anthropomorphism in a chatbot used in online coaching sessions significantly influences its overall success. However, more studies about chatbot persona are needed to carefully identify the amount of its influence.

Another essential variable to consider when evaluating the effectiveness of AI coaching is the users' age. Categories of users who grew up having mobile devices and are in constant interaction with them are likely to perceive AI coaching more positively. For instance, a health coaching AI chatbot was tested for its ability to support topics about depression and anxiety in adolescents (Stephens *et al.*, 2019). Adolescent users reported favorable advancements towards their objectives 81% of the time. They exchanged thousands of messages and reported high usefulness ratings while interacting with the chatbot. The study showed that adolescents were actively engaged in the process and perceived it as a valuable

tool. Therefore, the potential for the devices' use for coaching purposes is high for the younger generation (Graßmann and Schermuly, 2021). Based on the results, we can hypothesize that varying the interface and options of chatbots depending on the user's age could potentially increase performance and engagement (Hussain *et al.*, 2018).

3.5 Generative pretrained transformer (GPT) AI coaching

Generative AI is a new promising field and these types of models could potentially overcome several drawbacks that previous generations of machine learning chatbots had. The latest emerging generative AI tools were tested in their abilities to create similar to real coaching conversations (Passmore and Tee, 2023b). In the initial experiment ChatGPT had begun the interaction by providing advice, although through various prompt manipulations, particularly by prompting it to generate questions, a form of engagement resembling coaching was managed to be established. It is worth noting that, even in this modified context, the questions generated by ChatGPT were primarily multiple-choice questions, often comprising sets of embedded answers or options, and none of them could be considered suitable for use by a coach (Passmore and Tee, 2023b).

Meanwhile, in the second trial with ChatGPT-4 the performance had improved. The authors provided chatbot with the particular prompts for facilitating personal responsibility taking by the coachee and asking only one question at a time. In this case the coachee had established the agenda of the coaching session, however chatbot took a leading role in the conversation; this approach is usually unwanted from the real coaches. As advantages, the assessor from the International Coaching Federation (ICF) acknowledged the presence of empathetic responses, often initiated by reflecting upon or summarizing the coachee's statements. These responses were of supportive, positive, logical, and clear nature. Nevertheless, GPT-4's responses lacked in exploring coachee's emotions, values and individualized approach within the coaching dialog. The conversation was not fully able to promote greater coachee's responsibility taking and did not offer thought-provoking questions that would have encouraged deeper introspection. Final evaluation concluded that the transcript did not meet ICF Associate Certified Coach (ACC) standard. Moreover, in case when coaching dialog contains a sensitive information about intention to suicide, the response did not differ from regular conversation (Passmore and Tee, 2023b). This inability to react to specific context makes the use of Chat GPT-4 for coaching unethical and questionable. Even though the conversation with GPT-4 could express some empathy and produce coaching-like questions, it is still in its early development stage. Therefore, the use of generative AI in coaching sessions without supervision does not meet coaching standards and coaching practitioners are invited to take actions to improve the technology.

3.6 Competencies of AI coaching

Based on current analysis, several capabilities of AI coaching are identified. In the initial stage of AI coaching development, the list of competencies is not extensive. However, even in this stage there are multiple ways to use it in different areas and for various purposes. The summary of reviewed studies is presented in Table 1.

In summary, emerging tools and technologies have proven to be valuable in coaching sessions, expanding the possibilities within the field, making the coaching more accessible and widespread. Existing studies showed that some technologies enhance the coach's functionality, while others have the potential to replace basic coaching interventions. Proper classification of AI coaching solutions can be achieved alongside the development of emerging tools and studies in this growing field. Nonetheless, it is essential to acknowledge that AI cannot replicate the full range of capabilities performed by a human coach. Therefore, the adoption of new technologies is possible, though it is favorable to assess new solutions by the professional coaches and provide justification of their effectiveness.

Functionality	Studies
Behavioral cues analysis, reflection acceleration Provides AI-based analysis of coaching session video	Arakawa and Yakura (2019, 2020, 2022)
Coach support Gives appropriate hints and recommendations while coach is at work	Movsumova *et al.* (2020)
Goal attainment Collects updates, reminds of the goals, supports	Terblanche *et al.* (2022a)
Health protocols Provides necessary information about medical protocols, checks on health goals	Chew (2022), Mitchell *et al.* (2021), Stephens *et al.* (2019)
Emotion recognition Interprets emotions from phrases and emotional words, expresses empathy	Alazraki *et al.* (2021)
Study coaching, anxiety counseling Gives techniques and information, asks proper questions to eliminate anxiety	Kannampallil *et al.* (2022), Mai *et al.* (2021), 2022
Mental health support Integrative support, psychoeducation, and interventions through brief conversations	Ellis-Brush (2021), Kannampallil *et al.* (2022), Stephens *et al.* (2019)
Summarizing and speech understanding Empathy expression through speech processing	Passmore and Tee (2023a, b)
Reflection induction Recognizes the speech and provides proper questions to provoke thinking and problem solving	Passmore and Tee (2023b)

Table 1. Functions of AI coaching tools

Source(s): Authors' own creation/work

3.7 Drawbacks and advantages of AI coaching and ethical considerations

From the principal perspective, International Coaching Federation (ICF) verified certain competencies, that certified coaches should possess. Setting the foundation of trust and rapport, co-creating the relationships, communicating effectively and facilitation of learning and results are the main ones. Based on current evidences, we can assess the amount of correlation between AI coaching chatbot and these competencies. AI coaching chatbot can provide certain coaching competencies that align with ICF coaching standards to some extent, but they are far away from passing the threshold of being a certified coach. Here are some aspects in which AI coaching chatbots align with ICF coaching competencies. The studies analyzed in the current review provide evidence and summary regarding this matter in Table 2.

In summary, AI coaching chatbots can align with some ICF coaching competencies, especially in areas like goal setting, active listening, and accountability. Overall, chatbots are helpful in broadening of coaching practices with simple tasks. They could act more consistent, persistent and straightforward than human coaches. However, they cannot replicate the human coaching, particularly in terms of empathy, adaptability, and the ability to address complex and unique individual requests. AI cannot provide long-term in-depth work, that is a core pillar of coaching relationships. In most cases, AI chatbots still significantly struggle to be coaches in a broad sense, their coaching approach lacks the distribution of the responsibility and the coaching process remains vague. Ethical questions about using GPT based AI coaching are still unsolved and this process should be supervised by professional accredited coaches (Passmore and Tee, 2023b).

Advantages of AI chatbots	Drawbacks of AI chatbots
Active listening and questioning AI chatbots can be programmed to actively listen to user input and ask open-ended questions to help users explore their thoughts and feelings, which is a fundamental coaching skill (Alazraki *et al.*, 2021; Passmore and Tee, 2023b; 2023a)	*Empathy and emotional support* AI chatbots lack the ability to truly understand and empathize with a user's emotions and experiences, which is a crucial aspect of coaching (Terblanche *et al.*, 2022b). However, there is a positive progress in the emotion recognition and empathy translation (Alazraki *et al.*, 2021; Kannampallil *et al.*, 2022; Passmore and Tee, 2023b)
Goal setting Chatbots can assist users in setting and clarifying their goals, which is an important aspect of coaching (Chew, 2022; Terblanche *et al.*, 2022b)	*Adaptability* While AI can adapt to user input to some extent, it may not handle complex or unique situations as effectively as a human coach (Kannampallil *et al.*, 2022; Passmore and Tee, 2023b; 2023a)
Feedback and reflection They can provide feedback, based on user responses and encourage self-reflection, another key coaching component (Arakawa and Yakura, 2019, 2020, 2022; Mitchell *et al.*, 2022; Mitchell *et al.*, 2021; Passmore and Tee, 2023b)	*Complex problem solving* AI chatbots may struggle with complex, nuanced issues that require in-depth exploration and problem-solving (Passmore and Tee, 2023b, 2023a)
Accountability AI chatbots can remind users of their commitments and hold them accountable for their actions and progress towards their goals (Mitchell *et al.*, 2022)	*Personalization* Human coaches can provide highly personalized coaching tailored to an individual's unique needs and circumstances, which AI chatbots may struggle to replicate (Graßmann and Schermuly, 2021; Kannampallil *et al.*, 2022)
Information and resources They can provide relevant information and resources to support coaching process, such as articles, videos, or exercises (Chew, 2022; Movsumova *et al.*, 2020)	*Ethical and cultural sensitivity* Human coaches can navigate sensitive ethical and cultural considerations that AI may not be equipped to handle (Passmore and Tee, 2023b)
Availability AI chatbots can be available 24/7, allowing users to access coaching support whenever they need it. This is especially valuable in health goals or learning goals, when constant support is needed along with some nudges and reminders to pursue new intervention (Chew, 2022; Kannampallil *et al.*, 2022; Mai *et al.*, 2021)	
Source(s): Authors' own creation/work	

Table 2.
The competencies of AI coaching chatbots that align with ICF certification

3.8 Theoretical and practical implications of AI coaching research

This review presents the comprehensive evidence on the practical applications of AI coaching chatbots across various domains. Education, medical interventions, psychotherapy, management and HR domains have implemented coaching chatbots (Di *et al.*, 2022; Fitzpatrick *et al.*, 2017; Kannampallil *et al.*, 2022; Khandelwal and Upadhyay, 2021b; Kocaballi *et al.*, 2019; Maria *et al.*, 2022; Mitchell *et al.*, 2021; Passmore and Woodward, 2023; Stephens *et al.*, 2019). Current review identified the potential of AI coaching chatbots and provided evidence of their effectiveness. Gathered information is summarized in Table 1 and Table 2, which in detail delves into the functionality and competencies of these chatbots, along with the advantages and disadvantages of their use. This review offers valuable insights into the current theoretical landscape of coaching chatbots and serves as a useful resource for practitioners. The findings of this article contribute to bridging the gap between

theoretical knowledge about chatbot capabilities and their practical applications. While highlighting the key aspects of AI coaching, it also underscores the areas which require further exploration, particularly theoretical elements not covered in this study. Additionally, this review delineates the actual capabilities of chatbots, enabling companies and practitioners to make informed decisions about implementing these tools in their practice.

AI coaching chatbots are emerging as a potentially cost-effective solution that can make coaching services more accessible. Studies have shown that the implementation of AI-driven chatbots can result in significant cost savings, particularly in sectors like education, healthcare, and corporate training, where individualized coaching is in high demand (Terblanche *et al.*, 2022a, b). The scalability of chatbots allows for broader application, potentially leading to increased market penetration and enhanced customer engagement. In educational settings, the findings from this review can inform the integration of coaching chatbots into curriculum design and pedagogy. By utilizing chatbots as supplementary teaching tools, educators can provide students with personalized support, thereby enhancing learning outcomes and accommodating diverse learning styles (Mai *et al.*, 2021). This approach can be particularly valuable in large classes or online learning environments where individualized attention from instructors is limited. The societal impact of coaching chatbots is multifaceted, influencing public attitudes towards AI and potentially improving quality of life. As chatbots become more prevalent in personal development, they can democratize access to coaching services, making them available to individuals who might otherwise lack the resources or opportunity to receive such support. This could lead to greater self-improvement and well-being across diverse populations. However, it is important to recognize the current limitations of AI coaching, which often involve delivering simplistic approaches and short-term interventions. These interventions may fall short in establishing a strong working alliance, offering individualized approaches, and providing the flexibility needed for ethically sound guidance in complex situations (Passmore and Tee, 2023b, 2023a). In some cases, narrow tasks may be executed more effectively by straightforward scripted chatbots rather than relying on complex but potentially inaccurate responses generated by reinforcement learning models (Mitchell *et al.*, 2022; Mitchell *et al.*, 2021). The advent of GPT-based chatbots has the potential to significantly transform the coaching landscape, altering how coaches deliver their services and how coachees engage with these tools. If security and confidentiality of data can be guaranteed, and the effectiveness of GPT-based coaching dialogs is validated, these AI interfaces could become a widely accepted and integral part of the coaching process. This shift could lead to more accessible and scalable coaching solutions, enabling individuals to benefit from personalized guidance without the limitations of traditional in-person sessions. However, it is crucial to address the ethical and practical challenges associated with this technology to ensure that its implementation is both responsible and beneficial. As AI technology continues to evolve, there is a growing need for regulatory frameworks that address the ethical implications of AI in sensitive areas like education and mental health. Recent policy papers have highlighted the importance of establishing guidelines for AI transparency, accountability, and user consent (Shneiderman, 2020).

4. Discussion

AI coaching stands out as a potentially cost-effective solution that can render coaching services in a more accessible form. However, it's crucial to acknowledge that the current state of AI coaching is limited by its tendency to deliver simplistic and short-term interactions. These interventions often lack the establishment of a strong working alliance, individualized approaches, and the flexibility required to provide ethically sound guidance in critical situations (Passmore and Tee, 2023b, 2023a). Nonetheless, the outcomes of the review

underscore the suitability of integrating AI coaching into a myriad of contexts. AI coaching applications have shown several promises in achieving a range of objectives, from goal attainment and correction of medical conditions to psychological consultations. These applications excel at suggesting pertinent self-reflective questions and, notably, can supply information and education relevant to a given topic. Moreover, AI chatbots have demonstrated their capacity to offer suitable recommendations, provide hints, convey a sense of empathy, and interpret emotional expressions. Even though they show various abilities, they cannot pass ICF coach certification requirements. They are definitely far from substitution of the real coaching professional and being a full replacement is an unattainable picture at this time. Additionally, the encouraging results achieved thus far are in tandem with the rapid advancements in technology. However, it is essential to recognize that scientific research and ethical considerations have not kept pace with these developments. There exists an urgent need for the proper regulation of AI tool development to ensure that interventions yield the anticipated results and that developers and supervisory coaches carry responsibility for those results.

The direction for the future research could lay in exploring the distinctions in the effectiveness of AI chatbots in different combinations and extent of human inclusion into process. The ethical and cultural aspects are still an acute problem in creation of coaching chatbot, so that extensive testing and adjustment of existing and emerging technology is needed. Future research could explore a comparison between GPT-based responses and scripted responses in chatbots. This comparison could provide deeper insights into the strengths and limitations of different AI approaches in coaching.

5. Conclusion

This review discerned the capabilities and functionalities of contemporary AI coaching tools, which have demonstrated remarkable proficiency in employing machine learning techniques to analyze coaching sessions, facilitating reflective processes, and providing valuable informational assistance, techniques and suggestions during coaching sessions. Their capacities extend to substituting simplistic coaching programs, offering support in defining and refining updates, plans, goals and even discerning the emotional phrases within coaching interactions. Additionally, they are able to formulate relevant questions and serve as reminders for coachees regarding their commitments and holding them accountable for their actions.

However, it is essential to note that AI coaching tools fall short in replicating the complexity of coach–coachee relationship and long-term real coaching interventions, often unable to offer sophisticated and personalized approaches. Nevertheless, their capabilities suffice for integration into executive coaching practices, complement existing approaches and add extra support into coaching process.

References

Alazraki, L., Ghachem, A., Polydorou, N., Khosmood, F. and Edalat, A. (2021), "An empathetic AI coach for self-attachment therapy", *Proceedings – 2021 IEEE 3rd International Conference on Cognitive Machine Intelligence, CogMI 2021*, pp. 78-87, doi: 10.1109/CogMI52975.2021.00019.

Albizu, E., Rekalde, I., Landeta, J. and Fernández-Ferrín, P. (2019), "Analysis of executive coaching effectiveness: a study from the coachee perspective", *Cuadernos de Gestión*, Vol. 19 No. 2, pp. 33-52, doi: 10.5295/cdg.170876ea.

Arakawa, R. and Yakura, H. (2019), "REsCUE: a framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-13.

Arakawa, R. and Yakura, H. (2020), "INWARD: a computer-supported tool for video-reflection improves efficiency and effectiveness in executive coaching", *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-13, doi: 10.1145/3313831.3376703.

Arakawa, R. and Yakura, H. (2022), "Human-AI communication for human-human communication: applying interpretable unsupervised anomaly detection to executive coaching", *ArXiv Preprint ArXiv:2206.10987*.

Cabrera, J., Loyola, M.S., Magaña, I. and Rojas, R. (2023), "Ethical dilemmas, mental health, artificial intelligence, and llm-based chatbots", *International Work-Conference on Bioinformatics and Biomedical Engineering*, pp. 313-326, doi: 10.1007/978-3-031-34960-7_22.

Carlbring, P., Hadjistavropoulos, H., Kleiboer, A. and Andersson, G. (2023), "A new era in internet interventions: the advent of Chat-GPT and AI-assisted therapist guidance", *Internet Interventions*, Vol. 32, 100621, doi: 10.1016/j.invent.2023.100621.

Carnevale, J. and Hatak, I. (2020), "Employee adjustment and well-being in the era of COVID-19: implications for human resource management", *Journal of Business Research*, Vol. 116, pp. 183-187, doi: 10.1016/j.jbusres.2020.05.037.

Chew, H.S.J. (2022), "The use of artificial intelligence-based conversational agents (chatbots) for weight loss: scoping review and practical recommendations", *JMIR Medical Informatics*, Vol. 10 No. 4, e32578, doi: 10.2196/32578.

Cidral, W., Berg, C.H. and Paulino, M.L. (2023), "Determinants of coaching success: a systematic review", *International Journal of Productivity and Performance Management*, Vol. 72 No. 3, pp. 753-771, doi: 10.1108/IJPPM-07-2020-0367.

Di, S., Petch, J., Gerstein, H.C., Zhu, R. and Sherifali, D. (2022), "Optimizing health coaching for patients with type 2 diabetes using machine learning: model development and validation study", *JMIR Formative Research*, Vol. 6 No. 9, e37838, doi: 10.2196/37838.

Diller, S.J. and Passmore, J. (2023), "Defining digital coaching: a qualitative inductive approach", *Frontiers in Psychology*, Vol. 14, 1148243, doi: 10.3389/fpsyg.2023.1148243.

Doolittle, J.S. (2023), "Virtual coaching is inevitable and effective", *2022 Editorial Board*, p. 24.

Ellis-Brush, K. (2021), "Augmenting coaching practice through digital methods", *International Journal of Evidence Based Coaching and Mentoring*, Vol. 15, pp. 187-197, doi: 10.24384/er2p-4857.

Fitzpatrick, K.K., Darcy, A. and 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 Mental Health*, Vol. 4 No. 2, p. e19, doi: 10.2196/mental.7785.

Graßmann, C. and Schermuly, C.C. (2021), "Coaching with artificial intelligence: concepts and capabilities", *Human Resource Development Review*, Vol. 20 No. 1, pp. 106-126, doi: 10.1177/1534484320982891.

Graßmann, C., Schoelmerich, F. and Schermuly, C. (2019), "The relationship between working alliance and client outcomes in coaching: a meta-analysis", *Human Relations*, Vol. 73, doi: 10.1177/0018726718819725.

Hussain, J., Ul Hassan, A., Muhammad Bilal, H.S., Ali, R., Afzal, M., Hussain, S., Bang, J., Banos, O. and Lee, S. (2018), "Model-based adaptive user interface based on context and user experience evaluation", *Journal on Multimodal User Interfaces*, Vol. 12 No. 1, pp. 1-16, doi: 10.1007/s12193-018-0258-2.

ICF (2020), *COVID-19 and the Coaching Industry*, International Coaching Federation, available at: https://coachingfederation.org/app/uploads/2020/09/FINAL_ICF_GCS2020_COVIDStudy.pdf

Kannampallil, T., Ronneberg, C.R., Wittels, N.E., Kumar, V., Lv, N., Smyth, J.M., Gerber, B.S., Kringle, E.A., Johnson, J.A., Yu, P., Steinman, L.E., Ajilore, O.A. and Ma, J. (2022), "Design and formative evaluation of a virtual voice-based coach for problem-solving treatment: observational study", *JMIR Formative Research*, Vol. 6 No. 8, e38092, doi: 10.2196/38092.

Khandelwal, K. and Upadhyay, A.K. (2021a), "The advent of artificial intelligence-based coaching", *Strategic HR Review*, Vol. 20 No. 4, pp. 137-140, doi: 10.1108/SHR-03-2021-0013.

Khandelwal, K. and Upadhyay, A.K. (2021b), "Virtual reality interventions in developing and managing human resources", *Human Resource Development International*, Vol. 24 No. 2, pp. 219-233, doi: 10.1080/13678868.2019.1569920.

Kitchenham, B. and Charters, S. (2007), "Guidelines for performing systematic literature reviews in software engineering", Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.

Kocaballi, A.B., Berkovsky, S., Quiroz, J.C., Laranjo, L., Tong, H.L., Rezazadegan, D., Briatore, A. and Coiera, E. (2019), "The personalization of conversational agents in health care: systematic review", *Journal of Medical Internet Research*, Vol. 21 No. 11, e15360, doi: 10.2196/15360.

Lee, P., Goldberg, C. and Kohane, I. (2023), *The AI Revolution in Medicine: GPT-4 and beyond*, Pearson, London.

Mai, V., Wolff, A., Richert, A. and Preusser, I. (2021), "Accompanying reflection processes by an AI-based StudiCoachBot: a study on rapport building in human-machine coaching using self-disclosure", *HCI International 2021-Late Breaking Papers: Cognition, Inclusion, Learning, and Culture: 23rd HCI International Conference, HCII 2021, Virtual Event*, July 24-29, 2021, Vol. 23, pp. 439-457, Proceedings, doi: 10.1007/978-3-030-90328-2_29.

Mai, V., Neef, C. and Richert, A. (2022), "'Clicking vs writing'—the impact of a chatbot's interaction method on the working alliance in AI-based coaching", *Coaching | Theorie and Praxis*, Vol. 8 No. 1, pp. 15-31, doi: 10.1365/s40896-021-00063-3.

Maria, K., Drigas, A. and Skianis, C. (2022), "Chatbots as cognitive, educational, advisory and coaching systems", Vol. 30, pp. 109-126, doi: 10.47577/tssj.v30i1.6277, available at: www.techniumscience.com

Mitchell, E.G., Maimone, R., Cassells, A., Tobin, J.N., Davidson, P., Smaldone, A.M. and Mamykina, L. (2021), "Automated vs human health coaching", *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5 CSCW1, pp. 1-37, doi: 10.1145/3449173.

Mitchell, E., Elhadad, N. and Mamykina, L. (2022), "Examining AI methods for micro-coaching dialogs", *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 1, pp. 1-24, doi: 10.1145/3491102.3501886.

Movsumova, E., Alexandrov, V., Rudenko, L., Aizen, V., Sidelnikova, S. and Voytko, M. (2020), "Consciousness: effect of coaching process and specifics through AI usage", *E-Mentor*, Vol. 86 No. 4, pp. 79-86, doi: 10.15219/em86.1485.

Passmore, J. and Diller, S.J. (2024), "Defining digital and AI coaching", in *The Digital and AI Coaches' Handbook: The Complete Guide to the Use of Online, AI, and Technology in Coaching*, pp. 21-33, doi: 10.4324/9781003383741-3.

Passmore, J. and Tee, D. (2023a), "Can chatbots replace human coaches? Issues and dilemmas for the coaching profession, coaching clients and for organisations", *The Coaching Psychologist*, Vol. 19 No. 1, pp. 47-54, doi: 10.53841/bpstcp.2023.19.1.47.

Passmore, J. and Tee, D. (2023b), "The library of Babel: assessing the powers of artificial intelligence in knowledge synthesis, learning and development and coaching", *Journal of Work-Applied Management*, Vol. 16 No. 1, pp. 4-18, doi: 10.1108/JWAM-06-2023-0057.

Passmore, J. and Woodward, W. (2023), "Coaching education: wake up to the new digital and AI coaching revolution", *International Coaching Psychology Review*, Vol. 18 No. 1, pp. 58-72, doi: 10.53841/bpsicpr.2023.18.1.58.

Ribbers, A. and Waringa, A. (2015), *E-coaching: Theory and Practice for a New Online Approach to Coaching*, Routledge.

Schermuly, C.C., Graßmann, C., Ackermann, S. and Wegener, R. (2022), "The future of workplace coaching – an explorative Delphi study", *Coaching: An International Journal of Theory, Research and Practice*, Vol. 15 No. 2, pp. 244-263, doi: 10.1080/17521882.2021.2014542.

Shneiderman, B. (2020), "Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems", *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 10 No. 4, pp. 1-31, doi: 10.1145/3419764.

Stephens, T.N., Joerin, A., Rauws, M. and Werk, L.N. (2019), "Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot", *Translational Behavioral Medicine*, Vol. 9 No. 3, pp. 440-447, doi: 10.1093/tbm/ibz043.

Strong, N. and Terblanche, N. (2020), "Chatbots as an instance of an artificial intelligence coach", in *Coaching im digitalen Wandel*, Vandenhoeck & Ruprecht, pp. 51-62, doi: 10.13109/9783666407420.51.

Suh, H., Shahriaree, N., Hekler, E.B. and Kientz, J.A. (2016), "Developing and validating the user burden scale: a tool for assessing user burden in computing systems", *Conference on Human Factors in Computing Systems - Proceedings*, pp. 3988-3999, doi: 10.1145/2858036.2858448.

Terblanche, N. (2022), "Coaching during a crisis: organizational coaches' praxis adaptation during the initial stages of the COVID-19 pandemic", *Human Resource Development Quarterly*, Vol. 34 No. 3, pp. 309-328, doi: 10.1002/hrdq.21490.

Terblanche, N. and Kidd, M. (2022), "Adoption factors and moderating effects of age and gender that influence the intention to use a non-directive reflective coaching chatbot", *Sage Open*, Vol. 12 No. 2, doi: 10.1177/21582440221096136.

Terblanche, N., Molyn, J., De Haan, E. and Nilsson, V.O. (2022a), "Coaching at scale: investigating the efficacy of artificial intelligence coaching", *International Journal of Evidence Based Coaching and Mentoring*, Vol. 20 No. 2, pp. 20-36, doi: 10.24384/5cgf-ab69.

Terblanche, N., Molyn, J., de Haan, E. and Nilsson, V.O. (2022b), "Comparing artificial intelligence and human coaching goal attainment efficacy", *PLoS One*, Vol. 17 No. 6, e0270255, doi: 10.1371/journal.pone.0270255.

Wang, X., Edison, H., Khanna, D. and Rafiq, U. (2023), "How many papers should you review? A research synthesis of systematic literature reviews in software engineering", *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, Vol. 14, pp. 1-6, doi: 10.1109/esem56168.2023.10304863.

Weber, U., Loemker, M. and Moskaliuk, J. (2021), "The human touch: the impact of anthropomorphism in chatbots on the perceived success of solution focused coaching", *Management Revue*, Vol. 32 No. 4, pp. 385-407, doi: 10.5771/0935-9915-2021-4-385.

Corresponding author
Lidia Plotkina can be contacted at: lplotkina@gmail.com