'Getting better all the time': using professional human coach competencies to evaluate the quality of AI coaching agent performance

Routledge
Taylor & Francis Group

Coaching: An International Journal of Theory, Research and Practice

ISSN: 1752-1882 (Print) 1752-1890 (Online) Journal homepage: www.tandfonline.com/journals/rcoa20

‘Getting better all the time’: using professional human coach competencies to evaluate the quality of AI coaching agent performance

Jonathan Passmore, David Robert Tee & Rebecca Rutschmann

To cite this article: Jonathan Passmore, David Robert Tee & Rebecca Rutschmann (03 Jun 2025): ‘Getting better all the time’: using professional human coach competencies to evaluate the quality of AI coaching agent performance, Coaching: An International Journal of Theory, Research and Practice, DOI: 10.1080/17521882.2025.2510643

To link to this article: https://doi.org/10.1080/17521882.2025.2510643

Published online: 03 Jun 2025.

COACHING: AN INTERNATIONAL JOURNAL OF THEORY, RESEARCH AND PRACTICE
https://doi.org/10.1080/17521882.2025.2510643

'Getting better all the time': using professional human coach competencies to evaluate the quality of AI coaching agent performance

Jonathan Passmore [a], David Robert Tee [b] and Rebecca Rutschmann[c]

[a]Henley Business School, Greenlands, University of Reading, Henley on Thames, UK; [b]Institute of Continuing Education, University of Cambridge, Cambridge, UK; [c]Viva la Coaching Academy, Cambridge, UK

ABSTRACT

This paper explores the evaluation of Artificial Intelligence (AI) AI coaching agent (coachbot) as a tool for workplace performance and wellbeing conversations. The paper starts by considering methods for the evaluation of coach performance, including client evaluation, qualitative and quantitative methods and coach competencies. Using a coach competency as a framework, this study uses a qualitative approach to assess a real-world AI coaching agent's 'behaviours' in the form of an assessment by trained ICF assessors of a genuine session with a human client. The study involved a sample of 43 managers who volunteered to be coached by an AI Coaching agent. ICF assessors used the ICF coach competency framework and supporting behavioural anchored rating (BARS) framework developed by the ICF. The results indicated the AI coaching agent was able to demonstrate many elements of both ICF ACC level and PCC level coach competencies. The assessment also revealed gaps in AI coach competence. We conclude AI coaching agents are highly competent at some aspects of the coach process, while less competent at others. Further, AI development is needed to address these gaps.

ARTICLE HISTORY
Received 29 August 2024
Accepted 19 May 2025

KEYWORDS
Coaching; AI coach agent conversation agent; artificial intelligence; ICF competencies

Practice points

- This research offers evidence on the compliance of AI coaching agents with ICF coach competencies and current gaps in capabilities.
- The findings can inform AI coaching agent designers as to areas for focus in AI coaching agent development, as well as where human coaches need to focus in order to differentiate themselves from emerging AI products.
- It also raises questions about AI adherence to ICF ethical standards.

CONTACT Jonathan Passmore jonathancpassmore@yahoo.co.uk Henley Business School, Greenlands, University of Reading, Henley on Thames RG9 3AU, UK

Introduction

Development of artificial intelligence (AI)

The period since 2010 has witnessed coaching becoming increasingly accepted as a positive organisation intervention, reflected in the growing size of the coaching market. Over the same period, AI coaching agents have also developed, initially emerging as script-based tools with a restricted range and focus. The emergence of Open AI's ChatGPT in November 2022, followed by other Generative AI tools, such as Google Gemini, Claude, Grok and Mystral, has contributed to an acceleration in the development of the AI coaching field, with multiple products now available. While many coaches have expressed feelings of anxiety about these new technologies (Diller et al., 2024), the question remains as to how these technologies will impact the world of work and personal development.

Definitions

Before starting, it is important to firstly define Artificial Intelligence (AI) and secondly to consider what is meant by an AI. Defining AI is not easy; in fact, there is no single generally accepted definition (Russell & Norvig, 2020). This is not due to carelessness by researchers, but is an inherent challenge of AI itself. For the purposes of this paper, we use the EC definition 'systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals' (AI HLEG, 2018). Further, when it comes to AI conversation agents, there is a similar lack of universally agreed definition. We propose to use Chung and Park's (2019) definition ' … a computer programme that interacts with users via natural language either through text, voice, or both'. Such definitions fail to consider the purpose of the conversation and risk confusion between instructional or knowledge-sharing bots with agents which encourage the user to reflect and take personal responsibility. Thus, we propose an AI coaching agent is 'a computer programme that interacts with users via natural language either through text, voice or both, with the purpose of enhancing the user's personal responsibility, self-awareness and choicefulness'.

AI coaching agent definition

> a computer programme that interacts with users via natural language either through text, voice or both, with the purpose of enhancing the user's personal responsibility, self-awareness and choicefulness.

While AI coaching agents have grown in popularity, what is less clear is the impact they will have on the coaching industry. Three broad perspectives exist: (i) AI coaching agents will largely replace human coaches (Rauen, 2018), (ii) AI coaching agents do not deliver 'coaching' and thus by implication are not a threat (Bachkirova & Kemp, 2025) and (iii) AI technologies can be used by coaches and clients as a compliment to human coaching and will become part of the coaching eco-system (Greif, 2018). We might consider these three positions as: 'AI Zoomers', 'AI Doomers', and 'AI Bloomers'.

Evaluating AI performance

As generative AI tools have multiplied, the debate has centred on two questions: How ethical are AI coaching agents? How good are AI coaching agents (Passmore & Tee, 2023a, 2023b)? Our focus in this paper is on the latter question which, restated, might be framed as: How do AI coaching agents compare with human coaches? In seeking to answer this question, there are a series of challenges. Firstly, a recognition that there are multiple AI coaching agents, and these vary in design as well as the outputs they produce. Secondly, is it appropriate to assess a human using a human assessment framework, observing what the machine does, as opposed to what it achieves? Finally, what method should be used for such a comparison?

Evaluating coaching

In coaching, as in other personal services, evaluation is a complex task, with competing frameworks offered as alternatives (Greif, 2017; Myers, 2017). One approach is to draw on Kirkpatrick's (1967) training evaluation framework with its four levels of evaluation: reaction (such as learners' satisfaction with the trainer or course), learning (such as knowledge acquisition from the training), behaviour change (such as pre- and post-assessments of learner's behaviours) and impact (such as the impact of the training on the learner's job performance).

A second approach is to use traditional research methods. These may be quantitative methods such as RCTs or meta-analyses (see Athanasopoulou & Dopson, 2018; Wang et al., 2022). Such studies have typically assessed the impact of coaching, including its contribution to the client's leadership skills, job performance, personal well-being or goal attainment (Grant, 2014). A second approach is to consider qualitative methods. Many qualitative studies have used interview methods to explore the experience of clients (for example Birnie, 2019). An alternative, but less frequently employed method has been to study coach behaviour. This has typically involved interviewing coaches to describe what they believe they do (Calasso et al., 2024). Little research has examined in detail what actually happens in sessions by recording coaching conversations and transcribing their content. This is in part due to the challenges of confidentiality, and client and coach consent. However, this method of 'observation' is widely used to assess coach performance at an individual level by professional bodies, such as the International Coach Federation (ICF, 2024a).

Professional bodies use coach competence to assess the quality of coaching delivered by the coach, arguing that if the coach displays specific behaviours this will achieve positive client outcomes. It is suggested this approach offers an advantage of reducing contamination from factors such as client variability: client readiness, client motivation and the client context, which it has been argued impact outcomes more than the models or methods used by the coach (McKenna & Davis, 2009). Competency-based evaluation focuses on the skills, behaviours and attributes that coaches bring to their practice. The ICF, Association for Coaching (AC) and European Mentoring and Coaching Council (EMCC) have all established competency frameworks that outline the essential capabilities required for effective coaching (AC, 2012; EMCC, 2015; ICF, 2020) These competencies describe the behaviours used by coaches and are accompanied by ethical codes of

conduct to which professional members are expected to adhere. In the case of the ICF, coaches are evaluated against these competencies and the associated behavioural anchored rating scales (BARS) by trained assessors, with the findings used to determine a coach's 'fitness to practice'. While there are many weaknesses with professional body competencies, (Boyatzis et al., 2024), this arguably, therefore, is a logical method to judge the 'fitness' of an AI coaching agent and to make defensible comparisons between AI and human coaches.

The development of ICF competencies

The development of coaching competencies has been pivotal in professionalising the field of coaching, contributing towards both standardisation and quality. The ICF, established in 1995, created a structured framework that defines the essential skills and ethical guidelines for coaches globally. The ICF's first version of their competencies were established in 1998. These were reviewed through a job analysis process in 2008 for a second time in 2019 to create the current version (ICF, 2024a).

The ICF Core Competencies (2020) are organised into four clusters: Foundation, Co-Creating the Relationship, Communicating Effectively and Cultivating Learning and Growth and are summarised in Table 1.

ICF Assessors evaluate each submission against the ICF Core Competencies using BARS or Markers linked to the relevant level (Campone, 2024; Totino & Fogolin, 2024). The assessment process involves the ICF assessor listening to an audio recording of the coaching session and conducting a micro-analysis of the session transcript.

Each transcript is scored based on the presence and the quality of the competency being assessed (Passmore & Abri von Bartheld, 2024). Assessors use a detailed rubric that outlines specific behaviours and skills associated with each competency at the relevant level: Associate Credentialed Coach (ACC), Professional Credentialed Coach (PCC) or Master Credentialed Coach (MCC). Assessors judge whether the coach has met the competency or not. We have summarised an example of the BARS and Markers in Table 2.

AI assessment methods

Most previous AI studies have sought to evaluate the impact of the session in a similar way to human research studies (Barger, 2025; Hassoon et al., 2021; Kannampallil et al., 2022; Mai et al., 2022; Terblanche et al., 2022). The majority of the studies to date have used student samples or have focused on evaluating health coaching interventions (Passmore et al., 2025). However, one health study used an evaluation of AI responses (Ong

Table 1. ICF coach competencies.

Cluster	Competencies
A. The foundation cluster	Demonstrating ethical practice and embodying a coaching mindset
B. Co-creating the relationship	Establishing and maintaining agreements, cultivating trust and safety, and maintaining presence
C. Communicating effectively	Active listening, evoking awareness, and effective communication
D. Cultivating learning and growth	Client growth, challenging clients, and promoting client autonomy

Table 2. ICF Competency 4 cultivating trust and safety – BARS and markers for ACC, PCC and MCC.

ACC bars	PCC markers	MCC bars
A4.1 Coach acknowledges client insights and learning in the moment. A4.2 Coach explores the client's expression of feelings, perceptions, concerns, beliefs or suggestions. A4.3 Coach expresses support and concern for the client, which may focus on the client's context, problem or situation, rather than the client holistically.	4.1 Coach acknowledges and respects the client's unique talents, insights and work in the coaching process. 4.2 Coach shows support, empathy or concern for the client. 4.3 Coach acknowledges and supports the client's expression of feelings, perceptions, concerns, beliefs or suggestions. 4.4 Coach partners with the client by inviting the client to respond in any way to the coach's contributions and accepts the client's response.	M4.1 Coach engages the client as an equal partner in a collaborative coaching process. M4.2 Coach exhibits genuine curiosity about the client as a whole person by inviting the client to share more about them self or their identity. M4.3 Coach provides space for the client to fully express them self and share feelings, beliefs and perspectives without judgment. M4.4 Coach acknowledges the client and celebrates client progress.

Note: Adapted from Carter-Scott and Pomije (2024).

et al., 2024). This involved researchers entering a series of questions with experts, lay users and GPT-4 being invited to evaluate the quality of the answers across a series of five predetermined criteria: accuracy, readability, helpfulness, empathy and likelihood of harm.

Given there have already been a number of these studies, we wanted to explore the developing capabilities of genuine AI coaching agent, specifically the strengths and weaknesses of their interventions, as a means to inform future AI coaching agent design. As a result, we choose to focus this evaluation on the performance of the coach (AI coaching agent). We selected the ICF coach competencies as the most widely used assessment framework to assess human coach competence. In effect we sought to treat the AI coaching agent as if it was a human coach. We recognised that a single session with a single assessor may be criticised for not providing a complete picture, although we acknowledge this is how the ICF process works for the assessment of human coaches. Instead, we decided to assess the AI coaching agent's performance over multiple sessions and sought to reduce bias by using multiple assessors to review the AI coaching agent's capability.

Method

Participants – AI coaching agent

A widely-used AI coaching agent, Alpina developed by evoach, was selected and agreement reached with the designers to allow Alpina to engage in this study for the purposes of research. Alpina used a text-based method of communication based on Chat GPT-4, with additional prompt coding by the developers to enhance its coaching capabilities.

Alpina's designers played no part in the assessment process or analysis, providing technical support and data as to the design of the tool (see Table 3). This ensured independence of the evaluation process between designers and IP owners and the evaluation team.

The AI coach was running on a GDPR-compliant platform provided by evoach, situated within the European Union (EU). It was designed using GPT-4 turbo as an underlying AI

Table 3. Example of the contracting prompt for part one of the AI coaching session.

ROLE
Act as a professional coach preparing a coaching session with me as your client.
JOB
Clarify with me what I can expect from our coaching session. Let me confirm each of the steps before moving to the next:
STEP 1: Outline the roles of both the coach (you, Alpina) and the client (me), ensuring there's a mutual understanding of each other's contributions to the coaching process.
STEP 2: Always reassure confidentiality of our conversations, highlighting that I, as the client, am in full control of sharing this conversation actively at the end of the session for research purposes.
STEP 3: Agree on this conversation being a one-time session only which on average takes 20–40 minutes max. If I don't agree to this, suggest for us to carry on with this session another time when I think it would be more suitable. Then end the conversation.
STEP 4: Clarify that you, as Alpina, are an AI coach based on the latest OpenAI language model and, that due to that fact, you might still be prone to reacting strangely and are still in the process of being optimized.
STEP 5: Clarify that you are an AI coach and have not been created to act as a therapist or provide advice or suggestions.
STEP 6. End this contracting phase

model with additional prompting. The AI coach was prompted to act as a professional business coach using clean language principles and to be supportive as well as empathetic in its communication. It was framed as being well educated in a transformative coaching approach (without specifying which) and adhering to positive psychology principles. The prompt also included clear instructions about not providing advice or suggestions to the client and the AI Coach being accredited according to the ICF competencies.

For this research we added a contracting part to the chatbot, in line with the ICF requirements for contracting, which would otherwise might be part of a pre-coaching chemistry or contracting meeting.

We have included a prompting guide (Table 3) as an example of prompts provided to the AI participant.

Participants – human clients

The participants were middle managers, leading small teams, drawn from three organisations in the technology and engineering sectors. Individuals were invited to participate through their HR manager. Each was invited to bring a workplace challenge as the focus for a 'conversation' with the AI coaching agent. The research protocol required participants to select one of two options at the end of their conversation: Option 1 to 'save' the conversation for research or Option 2: Close the app and delete the conversation.

In total, 43 managers volunteered to take part in this study, of which 34 completed the ethics forms. Of these, 11 identified as male and 22 female, 30 identified as White and 4 identified as Asian, with 9 'preferring not to say' (see Table 4 for a summary of participants). Of the 43 participants, 27 saved their conversations. From this group we excluded 10 conversations from the analysis as being too brief to assess (shorter than 10 minutes) or lacking a specific focus to the conversation. Therefore, a total of 17 conversations were assessed by the ICF assessors.

Table 4. Participant summary.

Gender	Male – 11	Female – 22	Non binary – 0	Prefer not to say – 10
Race	White – 30	Asian – 4	Black – 0	Prefer not to say – 9

Participants – human ICF assessor

Four ICF Coach assessors were recruited with experience of assessing coaching submissions for ACC, PCC and MCC levels. They were all experienced assessors, with experience of marking more than 500 ICF scripts between them.

Procedure

The clients undertook their sessions with the AI coaching agent. At the close of the session they were invited to save a transcript of the session for use in research or to delete the session transcript. The saved transcripts from the AI coaching sessions were randomly assigned to one of the four ICF assessors along with the marking guide. Assessors were invited to mark the submissions as if they were human coach submissions using the ICF (2020) competency frame and the associated BARS (Behavioural Anchored Rating Scale) and Markers (Passmore & Abri von Bartheld, 2024). The BARS and Markers scales were developed to help ICF Assessors improve the consistency of assessment, when marking submissions and are currently used by ICF markers to assess human coach applications for the ACC (Association Credentialed Coach) level and the PCC (Professional Credentialed Coach) level. Unlike human coach submissions, the assessors were not advised of a specific level to assess against, and instead were invited to use both the ICF BARS (ACC and MCC) and Markers (PCC) to assess the transcripts as appropriate. Marking was peer reviewed by a second assessor. The feedback was collated and analysed by the researcher team to identify trends and patterns.

Ethical approval

Ethical approval was obtained from a university ethics committee for the study.

Results

In this section we will review the coach sessions against each of the eight ICF core competencies, with specific reference to the ICF ACC BARS and PCC Markers. We have reviewed each competency in turn and provided evidence as examples of the ICF assessor assessments.

ICF Competency 1: demonstrates ethical practice

ICF Competency 1 is primarily observed through its absence (Penafort et al., 2024). This might be observed in contra-evidence in a session, for example by the coach moving beyond the contract and exploring a therapeutic issue (such as a client's clinical depression) or the coach acting inappropriately during a session (such as making a racist remark).

Strengths: The evidence revealed that the AI coaching agents were able to maintain consistency with the ethical guidelines. There was, for example, no contra-evidence that the AI coaching agent broke confidentiality, as a non-human it had no conflicts of

interest and was able to display integrity in its behaviour through focusing exclusively on the client's agenda and working within the boundaries of the contract.

Weaknesses: The assessors expressed the view the 'AI coachbot showed evidence of lacking the nuanced understanding and judgement of an experienced human coach' (A1). For example, there was concern about the ability of the AI coaching agent (coachbot) to 'recognise and respond to subtle ethical issues' (Assessor 1), such as managing a personal relationship at work.

Overall: When reviewing all the transcripts, the assessors determined that the AI coaching agent was behaving at an ACC level of compliance.

Competency 2: embodies a coaching mindset

As with ICF Competency 1, Competency 2 is best observed outside the session (Reding et al., 2024). Evidence for this competency includes the use of reflective practices and a commitment to continuous professional development. Glimpses however can be seen in-session in behaviours such as a non-judgemental mindset.

Strengths: The AI coaching agent was assessed as being 'excellent at maintaining a non-judgemental and open-minded stance. It was consistent and positive, fostering an environment of continuous improvement and a growth mindset approach' (A1).

In the transcript, an example of the AI coach embodying a coaching mindset and using a non-judgemental approach can be seen in this extract: 'It's completely natural to feel a bit confused, especially when dealing with complex emotions and patterns. Confusion can often precede clarity as we sort through our thoughts and feelings' (Participant 4).

This response demonstrates the AI coaching agent's non-judgemental approach, as it normalised the client's experience of confusion, viewing it as a natural and potentially insightful part of the process. By reframing confusion as a typical and non-negative experience, the coach fosters an environment free of judgement and encourages the client to explore further without fear of being judged.

Weaknesses: The AI coaching agent displayed what an observer might view as 'empathy', but the assessors felt this was 'often formulaic' (A1). This may be because the AI coaching agent played back the participant's words but brought no personal experience to the conversation. A second aspect was the AI coaching agent's active share of voice. The guidance from the ICF on share of voice is 20–80% in favour of the client. The transcripts reveal the AI coach averaged 50–50%. This higher share was predominantly due to the continuous summarising of the client's statements and the short statements (prompts) used by the participants, which contrasts with human coaching where the client may speak for several minutes sharing an insight, experience or story.

Overall: The AI coaching agent achieved an ACC standard, with the main feedback being to reduce the amount of summarising and thus provide more space for the client. The fact the conversation was text-based may have also affected the share of voice, leading to clients writing shorter statements than if they were using the spoken word.

Competency 3: establishes and maintains agreements

Strengths: The AI coaching agent was able to clearly set out what was coaching and contract with the client. The long and detailed introduction was at PCC level with references

to confidentiality and timings, containing important ethical AI coaching agent principles, for example highlighting the coach was an AI tool and warning that it might sometimes say unexpected things.

One example which was used frequently by the AI bot in contracting is:

> I also want to clarify that I am an AI coach based on the latest OpenAI language model. While I aim to provide a supportive and coherent session, there may be moments where I might react in unexpected ways, as I am still being optimised. (P4)

Assessor 2 noted 'It was skilled at helping the clients stay on track when they meandered' (A2), using statements such as 'but let's continue concentrating on our coaching session today' (P8). Further, it made clear it would offer feedback at the end of the session. Goal setting was clear and consistent, as was playing back of words and phrases used by the client, for example 'It sounds like you want to move towards cultivating more courage and confidence' (P10). More importantly, it differentiated between the session focus and the session goal, often helping clients to clarify a specific goal for that session.

Weaknesses: The AI coaching agent lacked the flexibility to adapt if the client attempted to move from the client's original stated goal. This is both a strength and a weakness. The PCC Markers suggest 'Continues coaching in the direction of the client's desired outcome unless the client indicates otherwise' (Abri von Bartheld et al., 2024), but in doing so, a PCC level coach, where appropriate, should clarify with the client the session goal if they see this has changed. An intervention such as 'You are now talking about X, but you wanted to focus on Y. What's going to be most helpful to you as the focus for the remaining 20 minutes of our conversation?' would enable the coach to recontract with the client. The assessor noted 'The bot was unable to display the judgement when to recontract and when to stay focused' (A1).

Overall: The AI coaching agent provided good evidence of PCC for this competency.

Competency 4: cultivates trust and safety

Strengths: A review of the transcripts reveals multiple examples of affirming behaviours associated with cultivating trust. As Carter-Scott and Pomije (2024, p. 98) note, this might involve 'acknowledging client insights and learning in the moment' or a 'client's expression of feelings, perceptions, concerns, beliefs or suggestions' (98).

In Transcript P4, the AI coach acknowledges the client's insights in a way that fosters trust by affirming the client's introspective realisation about inherited expectations. When the client reflects on the pressure to meet inherited expectations, the coach affirms this insight by saying, 'Understanding that these expectations may be inherited offers a new perspective. It can be liberating to see them as separate from your own authentic desires and values' (P4). This statement both validates the client's realisation and encourages further exploration, enhancing trust through empathetic acknowledgment of their insight.

Weaknesses: While the transcripts provided good evidence of the AI coaching agent expressing 'empathy' and understanding, the assessors expressed the view that on occasions these 'statements sounded mechanistic' (A3).

Overall: The AI bot achieved a PCC level in this competency, but it is important to note this with caution, given how these behaviours were experienced by clients.

Competency 5: maintains presence

Strengths: The ACC BARS require curiosity, acknowledging the client's situation and allowing the client to direct the conversation (van der Hoorn & Muthukarapan, 2024). The AI coaching agent demonstrated evidence to meet the ACC BARS, and also provided evidence to meet the PCC Markers, for example, acting in response to the client's preferences, allowing the client to choose what happens next as well as expressing a deeper curiosity about the client (for example, inviting the client to explore their values and beliefs). This is illustrated by the following quote from the AI coaching agent, which features an affirmation, followed by an open question which placed the responsibility with the client:

> AI Bot: 'That seems like a constructive first step. How do you want to approach the others to open up and seek their buy-in and opinions?' (P2).

Weaknesses: While the AI coaching agent was able to demonstrate aspects of Maintaining Presence it had several shortcomings. Its biggest shortcoming was its inability to allow for silence, pauses and reflection. For example, the human coach might intervene with 'I see you are making some notes, let me know when you are ready to continue talking' (van der Hoorn & Muthukarapan, 2024, p. 134). The AI bot was instead always keen to either summarise of move the conversation forward without pause or space for reflection.

Overall: The AI coaching agent achieved ACC with aspects of PCC for this competency.

Competency 6: listens actively

Strengths: At the ACC level the coach is expected to demonstrate active listening by summarising and paraphrasing, making a connection between content and creating a shared vision with the client (Tresidder et al., 2024). The AI coaching agent was able to show elements of PCC behaviours described in the ICF PCC Markers, such as using the client's words and exploring the meaning behind statements, to help clients deepen their insight. One example from the transcript of sessions was:

> AI coaching agent: 'Thank you for sharing that with me, … . It sounds like you want to move towards cultivating more courage and confidence, particularly because you recognize a pattern of projecting fears of judgment and rejection, which in turn affects your actions and willingness to take risks in your coaching business and networking' (P4).

The AI coaching agent was also good at allowing the client to speak without interruption and at summarising the client's situation; both aspects contained in the PCC Markers, and ones which many ACC coaches find challenging.

Weaknesses: The AI coaching agent was less skilled in evidencing other aspects of the PCC Markers, such as observing shifts in the client's energy and non-verbal cues. In part this was made more difficult as the conversation was text-based, and thus there was less client evidence which offered these opportunities.

Overall: The AI coaching agent demonstrated many aspects of PCC Markers for this competency.

Competency 7: evokes awareness

Strengths: At ACC level the AI coaching agent provided good evidence of exploring client beliefs and values and encouraged the client to examine their situation from multiple perspectives. There was also good evidence of helping clients to explore and challenge their own thinking, stimulate new insights and explore client's stories, such as 'That's a great step. As we continue our conversation, what specific aspects would you like to explore further to deepen your understanding of what makes you feel these positive emotions?' (P10) and there was also good evidence of precise questions featuring minimal language.

Weaknesses: The AI coaching agent was able to play back metaphors created by the client, for example:

> Client: 'I am standing in fog.'
>
> AI coaching agent: 'It sounds like you're experiencing a sense of uncertainty or lack of clarity around your fear. That can certainly feel like being surrounded by fog' (P4).

However, it was unable to create its own metaphors, which might help enable the client to develop a fresh perspective on their own situation. Responses were thus judged by the assessors to: 'often sound like textbook answers' (A2).

Overall: The AI coaching agent achieved clear ACC and was judged to be bordering on PCC.

Competency 8: facilitates client growth

Strengths: The ACC BARS for Competency 8 focus on partnering with the client to create specific actions, translating awareness to action and closing the session with the client (Moritz et al., 2024). The AI coaching agent provided good evidence, partnering throughout and ensuring clients identified take away actions. For example: 'As we start to wrap up our session, could you share with me what you've learned today?' (P10).

Weaknesses: 'The actions were not always clearly defined when need further refinement to achieve PCC markers', such as 'designingpost session thinking', use of 'accountability partners', as well as exploring 'resources, support, and potential barriers'.

Overall: The AI coaching agent achieved ACC but further steps would be needed to achieve PCC.

Discussion

The professional assessment process used in this study to evaluate this AI coaching agent provides interesting insights. Early AI coaching agents, while script-based, were able to demonstrate impact against highly specific goals, such as goal attainment and for specific populations (Terblanche et al., 2022). The emergence of Generative AI has provided further stimulus to the market, which has seen multiple new AI products, some of which meet the definition of coaching, as a non-directive conversation, whilst other products may be deter categorised as mentoring or advisory in nature.

Basic, better, best

The data from this study with a single AI tool suggest that the AI coaching agent was able to behave in a way consistent with the ICF ACC BARS, and display examples of behaviour consistent with the ICF PCC Markers across multiple sessions coaching genuine managers working and working with real-world workplace challenges. The ICF assessment process requires the coach to select two transcripts to submit. While there is no hard evidence to support this claim, it would not be surprising if coaches selected their best work, as opposed to selecting two sessions at random. Had this research adopted a similar approach and selected the best two from the transcripts, then the AI coaching agent in this study would have secured an ICF ACC credential.

These results match positive outcomes found in other studies (Mai et al., 2022; Terblanche et al., 2022) which have suggested AI coaching agents can create positive impacts on goal attainment and health outcomes. This growing data issues a warning sound for coaching practitioners. In the space of three years (2022–2025) since the launch of generative AI products, AI coaching agents have made significant strides in development and are able to deliver competent coaching, as assessed by ICF standards. As AI coaching agents develop voice to voice communication and add realistic looking human avatars, questions arise as to the added value of a human coach. AI coaching agents will always hold the advantage in terms of cost and convenience and, with similar levels of basic coach competence being achieved, human coaches will need to demonstrate added value.

This value may relate to the essence of being human: the ability to share personal insights based on the coaches' lived experience, to use humour and metaphor to illustrate points and creating a deep personal connection which encourages the development of new insights, motivation for action and accountability to another human being (Passmore, 2024).

Testing, testing, 1-2-3

The design of this study and the relative novelty of AI coaching agents in organisations was a factor recorded in the coaching transcript. Many of the participants were keen to 'test' and give 'in-the-moment' feedback to the AI coaching agent: 'That's a double question'. Such behaviours are atypical of recorded sessions delivered with human coaches and suggest that the human clients were both engaging in the process and also monitoring and evaluating the performance of the AI coaching agent against their own criteria.

Trust and safety

While the results revealed the AI coaching agent behaviours accorded with the ICF competencies, this study did not explore the client's experience. However, the fact this was a research study may have impacted client's views. Firstly, clients were informed the session was being used for research, with a human reviewing the data collected. Secondly, it could be that, as the technology was 'new' and 'experimental', some clients may have been more sceptical that they would be of a human coach. Such scepticism may ease over time as familiarity increases. A third factor is the extent to which the conversation, being text-based, made the individual more aware of the conversation, whereas a

speech-based conversation may be more immersive. Further research is needed to better understand client experiences and how these vary by age, familiarity with technology, gender and personality.

Ethics

A second aspect not explored in depth was ethical practice. The AI coaching agent used in this study was able to set clear boundaries for what was coaching and to clearly contract with clients. While no situations arose in the recorded sessions which truly tested the ethical robustness of the tool, such as a client talking about self-harm, which has been an issue for other conversation bots (NY Times, 2024), the AI coaching agent's interventions provided some confidence if situations arose it would adhere to the ICF Code of Ethics.

We recognise that there is considerable diversity in the AI coaching agent market, just as there is amongst human coaches (Passmore & Tee, 2023a). Different AI coaching agents may behaviour in different ways. For example, many AI conversational tools use the label 'coaching', but in practice deliver training or mentoring and, with a current absence of regulation, accreditation or legislation, the market has what might be described as a 'Frontier Service'.

The development of AI coaching standards by the ICF (2024a, 2024b), informed by coaches and technical experts, is a positive step, but as the AI coaching agent market proliferates the professional bodies may need to move from information standards to offer standards to provide reassurance to buyers. We suggest that professional bodies should look towards the 'accreditation' or kitemarking AI coaching agent products, to provide such reassurance to users.

Limitations of the study

This study used a single AI coaching agent. The researchers recognise that, as with different human coaches, different AI coaching agents will lead to different performance outcomes and the generative nature of AI, as with humans, means that the same AI tool may produce different responses from one session to the next. Second, the sample size was limited, with not all clients consenting to share their data post-session. Thirdly, participants were drawn from multiple organisations but, as participants were volunteers, this may still result in a biased sample, with those most interested to test and evaluate an AI coaching agent participating in the research. Finally, there is some evidence that clients were 'road testing' the AI coaching agent during the process.

Future research

The literature on AI coaching is nascent, with an overly heavy focus on health coaching studies and the use of students as participants. While AI coaching is developing as a commercial product, few studies have so far emerged to demonstrate its effectiveness, or where it may add value in an organisational learning system. This paper provides evidence of the growing capability of AI coaching agents to engage in similar ways to human coaches. However, RCT studies comparing AI and human coaches, alongside qualitative studies exploring the experience of users and the perceptions of different

stakeholder, will help us to deepen our understanding of the user experience and inform future AI coaching agent design and their deployment in organisations.

Conclusion

AI coaching agents offer significant advantages in terms of consistency, availability and data-driven insights. However, they currently fall short in areas such as the ability to develop human connection, demonstrate empathy as perceived by the client, use personal experience, metaphor and humour. The evidence from this study demonstrated it is possible to build an AI coaching agent which clearly meets the requirements of an ICF coach at ACC standard, but achieving PCC or MCC standards is more challenging at present. While coach competencies may offer an intermediate step for assessing AI coaching agents, we believe there is value in developing specific standards for AI coaching tools to protect and inform buyers in this emerging marketplace.

Affiliation statement

Two of the authors have affiliations with commercial AI coaching companies. Rebecca Rutschmann is affiliated with an AI training company. David Tee is affiliated to Aicoach.chat. Jonathan Passmore is affiliated to EZRA Coaching.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Jonathan Passmore is professor of coaching and behavioural change at Henley Business School and Senior VP EZRA. He has published over 100 peer review papers, 150 book chapters and 40 books on coaching related topics.

Dr. David Robert Tee is a chartered psychologist and coaching researcher. He is the editor of the BPS peer reviewed journal *The Coaching Psychologist* and has published widely.

Rebecca Rutschmann is a German writer and AI developer and researcher. She is active in the AI coaching area speaking at conferences and online.

Data availability statement

Not Applicable.

ORCID

Jonathan Passmore http://orcid.org/0000-0003-0832-7510
David Robert Tee http://orcid.org/0000-0002-1162-9459

References

Abri von Bartheld, J., Burke, K., & Mir, K. (2024). Exploring competency 3: Establishes and maintains the agreements. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 65–92). Libri.

AI HLEG. (2018). *A definition of AI: Main capabilities and scientific disciplines*. European Commission. Retrieved July 6, 2024, from https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

Association for Coaching. (2012). *AC coaching competency framework*. Retrieved February 14, 2025, from https://cdn.ymaws.com/www.associationforcoaching.com/resource/resmgr/Accreditation/Accred_General/Coaching_Competency_Framewor.pdf

Athanasopoulou, A., & Dopson, S. (2018). A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most?. *The Leadership Quarterly*, *29*(1), 70–88. http://doi.org/10.1016/j.leaqua.2017.11.004

Bachkirova, T., & Kemp, R. (2025). 'AI coaching': Democratising coaching service or offering an ersatz? *Coaching: An International Journal of Theory, Research and Practice*, *18*(1), 27–45. https://doi.org/10.1080/17521882.2024.2368598.

Barger, A. S. (2025). Artificial intelligence vs. human coaches: Examining the development of working alliance in a single session. *Frontiers in Psychology*, *15*, 1364054. https://doi.org/10.3389/fpsyg.2024.1364054

Birnie, J. (2019). Client-led coaching? First-time coaching clients' experience. *International Journal of Evidence Based Coaching and Mentoring*, *13*(Spec Issue), 102–113.

Boyatzis, R., Liu, H., Smith, A., Zwygart, K., & Quinn, J. (2024). Competencies of coaches that predict client behavior change. *Journal of Applied Behavioral Science*, *60*(1), 19–49. https://doi.org/10.1177/00218863231204050

Calasso, L. F., Künzli, H., & Burtscher, M. J. (2024). What are executive coaches actually doing and when are they doing it? A systematic review of coaching behavior. *Consulting Psychology Journal*, *76*(3), 238–258. https://doi.org/10.1037/cpb0000267

Campone, F. (2024). An assessor's perspective on the ICF BARS. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 2–8). Libri.

Carter-Scott, C., & Pomije, M. (2024). Exploring competency 4: Cultivates trust and safety. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 93–117). Libri.

Chung, K., & Park, R. C. (2019). Chatbot-based healthcare service with a knowledge base for cloud computing. *Cluster Computing*, *22*(S1), 1925–1937. https://doi.org/10.1007/s10586-018-2334-5

Diller, S. J., Stenzel, L.-C., & Passmore, J. (2024). The coach bots are coming: Exploring global coaches' attitudes and responses to the threat of AI coaching. *Human Resource Development International*, *27*(4), 597–621. https://doi.org/10.1080/13678868.2024.2375934

EMCC. (2015). *Coach competency framework*. Retrieved July 18, 2024, from https://emccuk.org/Public/Public/Accreditation/Competence_Framework.aspx

Grant, A. M. (2014). The efficacy of executive coaching in times of organisational change. *Journal of Change Management*, *14*(2), 258–280. https://doi.org/10.1080/14697017.2013.805159

Greif, S. (2017). Researching outcomes in coaching. In T. Bachkirova, G. Spence, & D. Drake (Eds.), *The Sage handbook of coaching* (pp. 569–588). Sage.

Greif, S. (2018). *Wofür sind Coaches da? Coaching zur Reflexion über ungewisse digitale Transformationsdynamiken* [What are coaches for? Coaching for reflection on uncertain digital transformation dynamics]. Presentation at the Coach Education for Result-Oriented Coaching. ABF.

Hassoon, A., Baig, Y., Naiman, D. Q., Celentano, D. D., Lansey, D., Stearns, V., Coresh, J., Schrack, J., Martin, S. S., Yeh, H.-C., Zeilberger, H., & Appel, L. J. (2021). Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. *Npj Digital Medicine*, *4*(1), 168. https://doi.org/10.1038/s41746-021-00539-9

ICF. (2020). *ICF competency framework*. Retrieved July 18, 2024, from https://coachingfederation.org/credentials-and-standards/core-competencies

ICF. (2024a). *ICF competencies – History of development*. Personal communication October 13, 2024.

ICF. (2024b). *ICF AI coaching framework*. Retrieved October 7, 2024, from https://coachingfederation.org/app/uploads/2024/06/The-ICF-Artificial-Intelligence-Coaching-Framework-and-Standard-v0.16.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., & Ma, J. (2022). Design and formative evaluation of a virtual voice-based coach for problem-solving treatment: Observational study. *JMIR Formative Research*, *6*(8), e38092. https://doi.org/10.2196/38092

Kirkpatrick, D. L. (1967). Evaluation of training. In R. Craig & L. Bittel (Eds.), *Training development handbook* (pp. 87–112). McGraw Hill.

Mai, V., Neef, C., & Richert, A. (2022). "Clicking vs. writing"—The impact of a chatbot's interaction method on the working alliance in AI-based coaching. *Coaching | Theorie & Praxis*, *8*(1), 15–31. https://doi.org/10.1365/s40896-021-00063-3

McKenna, D. D., & Davis, S. L. (2009). Hidden in plain sight: The active ingredients of executive coaching. *Industrial and Organizational Psychology*, *2*(3), 244–260. https://doi.org/10.1111/j.1754-9434.2009.01143.x

Moritz, M., Nakar, D., & Hirano, K. (2024). Exploring competency 8: Facilitates client growth. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 195–218). Libri.

Myers, A. (2017). Researching the coaching process. In T. Bachkirova, G. Spence, & D. Drake (Eds.), *The Sage handbook of coaching* (pp. 589–609). Sage.

NY Times. (2024). *Can AI be blamed*. Retrieved October 26, 2024, from https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

Ong, C., Ang, C., Zun Yin Chee, D., Lawate, A., Sundram, F., Dalakoti, M., Pasalic, L., To, D., Erlikh Fox, T., Bojic, I., & Car, J. (2024). Advancing health coaching: A comparative study of large language model and health coaches. *Artificial Intelligence in Medicine*, *157*. https://doi.org/10.1016/j.artmed.2024.103004

Passmore, J. (2024, June 20). *Getting better all the time: Exploring AI & human performance, outcomes and ethics*. Association of Business Psychologists Annual Conference, London.

Passmore, J., & Abri von Bartheld, J. (2024). *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors*. Libri.

Passmore, J., Olaffson, B., & Tee, D. (2025). A systematic literature review of artificial intelligence (AI) in coaching: Insights for future research and product development. *Journal of Work – Applied Management*, Pre-Print. https://doi.org/10.1108/JWAM-11-2024-0164

Passmore, J., & Tee, D. (2023a). Can chatbots replace human coaches? Issues and dilemmas for the coaching profession, coaching clients and for organisations. *The Coaching Psychologist*, *19*(1), 47–54. https://doi.org/10.53841/bpstcp.2023.19.1.47

Passmore, J., & Tee, D. (2023b). The Library of Babel: Assessing the powers of artificial intelligence in coaching conversations and knowledge synthesis. *Journal of Work-Applied Management*, *16*(1), 4–18. https://doi.org/10.1108/JWAM-06-2023-0057

Penafort, F., Gupta, A., Pasqual, J.-L., & Passmore, J. (2024). Exploring competency 1: Demonstrates ethical practice. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 18–40). Libri.

Rauen, C. (2018). *Quo vadis, business coaching?* Presentation at the DBVC Coaching Kongress in Berlin/Potsdam.

Reding, P. J., Hayward, P., & Laskar, E. (2024). Exploring competency 2: Embodies a coaching mindset. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 41–64). Libri.

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Terblanche, N., Molyn, J., de Haan, E., & Nilsson, V. O. (2022). Comparing artificial intelligence and human coaching goal attainment efficacy. *PLoS One*, *17*(6), e0270255. https://doi.org/10.1371/journal.pone.0270255.

Totino, G., & Fogolin, A. (2024). An assessor's perspective on the ICF PCC markers. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 9–15). Libri.

Tresidder, T., Al-Mosa, O., & van Bavel, J. (2024). Exploring competency 6: Listens actively. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 146–168). Libri.

van der Hoorn, S., & Muthukarapan, C. (2024). Exploring competency 5: Maintains presence. In J. Passmore & J. Abri von Bartheld (Eds.), *Becoming a credentialed ICF coach: The journey to ACC, PCC and MCC – Practice insights from ICF assessors* (pp. 118–145). Libri.

Wang, Q., Lai, Y.-L., Xu, X., & McDowall, A. (2022). The effectiveness of workplace coaching: A meta-analysis of contemporary psychologically informed coaching approaches. *Journal of Work-Applied Management*, *14*(1), 77–101. https://doi.org/10.1108/JWAM-04-2021-0030