Human coaches and AI coaching agents: an exploratory quasi-experimental study of workplace client attitudes

Journal of Work-Applied Management

Human coaches and AI coaching agents: an exploratory quasi-experimental study of workplace client attitudes

Jonathan Passmore
Henley Business School, University of Reading, Greenlands Campus, Henley on Thames, UK
David Tee
Institute of Continuing Education, University of Cambridge, Cambridge, UK
Gina Palermo
EZRA Coaching, London, UK, and
Rebecca Rutschmann
Viva la Coaching Academy, Karlsruhe, Germany

Received 13 February 2025
Revised 19 March 2025
18 April 2025
5 June 2025
19 June 2025
Accepted 22 June 2025

Abstract

Purpose – The aim of this exploratory study was to examine workplace managers' attitudes to artificial intelligence (AI) coaching agents compared with human coaches, along with the factors influencing AI technology adoption in the workplace.

Design/methodology/approach – A small-scale quasi-experimental pilot study featured white collar workers ($n = 63$) and used a custom-designed questionnaire to compare client responses to engaging with a human coach or an AI coaching agent, Alpina, across six factors: developing new insights, working alliance, goal attainment, commitment, trust, confidentiality and shame.

Findings – Following a single coaching session, coaching clients reported higher scores when working with a human coach than clients working with an AI coaching agent on all factors.

Research limitations/implications – Caution is needed given the limited sample size, the use of a single session to evaluate and the use of a custom-designed measure.

Practical implications – Whilst AI coaches continue to improve technically and vary widely in functionality and sophistication, in this study, clients appear to rate human coaches more highly than the featured AI coach on a range of factors, such as insights and goal attainment. Further research is needed to validate these exploratory results and test how speech-to-speech or other factors may influence user ratings.

Originality/value – This is the first study exploring employee attitudes by comparing human and AI coaches across a range of factors.

Keywords AI coaching agent, Artificial intelligence, Coachbot, Coaching impact, Goal attainment

Paper type Research article

Introduction

AI is commonly defined as "*a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*" (Kaplain and Haenlein, 2019, p. 15). Developments in AI have profound implications for professional services, ranging from consulting to training and coaching (Passmore and Tee, 2024). The emergence of Generative AI marked a pivotal step in the

Journal of Work-Applied Management
Emerald Publishing Limited
e-ISSN: 2205-149X
p-ISSN: 2205-2062
DOI 10.1108/JWAM-02-2025-0032

development of AI, whose use has grown rapidly over the period 2022–2025. While traditional script-based AI operates based on predetermined rules, Generative AI can learn from data and autonomously generate unique content. The coming decade (2030s) is likely to see radical changes in the way coaching and professional services are provided, with the potential of AI solutions to increase accessibility, reduce costs and improve work quality, although the exact pattern is at present challenging to predict (Dell'Acqua *et al*., 2023).

The impact of AI on coaching is a topic of considerable debate, with divergent views on how an AI coaching agent will integrate into the coaching ecosystem. Three distinct perspectives can be identified. The first is that AI technologies will replace human coaches. A second school of thought is that AI tools cannot replace human coaches due to coaching being contingent upon unreplicable and intrinsic human qualities, meaning that AI is, in effect, unable to coach. Thirdly, AI will complement human coaching in a hybrid model, but design and deployment require careful management to protect individuals and organisations. We have labelled these three perspectives as "Zoomers", "Doomers" and "Bloomers".

Examining the first perspective, the "Zoomers", these proponents argue that AI is a learning technology which, over time, will improve and may ultimately replace human coaches. This perspective is grounded in the rapid advancements of AI technologies (Balasubramanian, 2023) and evidence from AI coaching (Terblanche *et al*., 2022a). A second school of thought comes from those more critical of AI. "Doomers" argue that AI conversation tools will never replace human coaches due to the unique qualities of being human, which are argued to be essential features of coaching (Bachkirova and Kemp, 2024). Human coaches bring interpersonal skills that facilitate deeper conversations, which AI is unable to replicate and, without which, no worthwhile coaching outcomes can be produced. Lastly, a more integrative view, "Bloomers", argues that AI has the potential to become a useful part of the coaching ecosystem, but care needs to be taken in managing ethical and practical aspects of its deployment (Passmore and Tee, 2024). However, to achieve a successful outcome, careful management is required in the design and deployment of AI technologies in coaching.

Literature review

The past decade has provided significant evidence of the potential of human coaches to generate positive changes in clients' behaviours, cognition and emotions (Wang *et al*., 2022) and to achieve these outcomes within professional workplace environments (Athanasopoulou and Dopson, 2018). However, research on human coaching comparatively remains sparse compared with other disciplines. Reflecting the complexity of researching an aspect of development which features a personal, confidential relationship, where power in the relationship rests not with the therapist or service provider, but with the consumer or service purchaser and where often a third party, the organisation, is also cautious about its data and reputation.

A systematic literature review (SR) of AI coaching research suggests that the evidence on the impact is scant (Passmore *et al*., 2025b). The review identified 16 peer-reviewed journal papers ($n = 2{,}312$) which met its inclusion criteria and from which these four themes emerged: the potential of AI coaching agents, their current known impact, the risks and opportunities for the coaching industry and coaching ethics. On the theme of potential, the SR noted that AI coaches can already provide feedback, help with goal-tracking, serve as accountability partners and are seen as accessible, convenient and psychologically safe Terblanche *et al*., (2023, 2024). On the second theme, AI coaching impact, evaluation research has been limited in number and scope, with serious limitations. Studies have typically used student samples as opposed to workplace coaching clients, often focused on health outcomes or health care environments.

Reviewing this evidence, in a study by Ellis-Brush (2021), the participants were coached by an AI coaching agent trained in psychological techniques (specifically, cognitive behavioural coaching) and showed a growth in self-resilience. Hassoon *et al.* (2021) suggested that cancer

survivors coached by an AI coaching agent were able to increase their step count more than those assigned to a control group. Terblanche *et al*. (2022b) reviewed levels of goal attainment in a comparative study on the effectiveness of AI coaching and human coaching, with a quasi-experimental design featuring data from two separate longitudinal studies, different samples and different time periods. In one study, data were collected using human coaches by one group of researchers in 2018–19. In the second study, AI data were collected from a different sample during the UK COVID-19 Lockdown in 2020. While the AI coach participants achieved similar levels of goal attainment as those working with a human coach, the unique circumstances of the study make generalising the results problematic. Finally, in a study by Bamber (2025) comparing client ratings of a concealed human coach, pretending to be an AI coaching tool and a human coach (a Wizard of Oz study) and claimed to show similar levels of working alliance. Alongside these client-based evaluation studies, one study (Passmore *et al*., 2025c) used International Coaching Federation (ICF) assessors to evaluate AI bots using the ICF's coaching competency framework. The outcome suggests that the AI tool evaluated was capable of meeting the ICF Associate Certified Coach (ACC) standards.

Overall, these studies suggest AI's ability to effectively fulfil many competencies determined as necessary for effective coaching, but, given the limitations of previous studies, it is not yet possible to conclude if AI coaching agents offer comparable performance to human coaches, specifically in the context of real-world organisational coaching. Apart from a focus on goal attainment, few studies have sought to measure the wider features of AI coaching. Such measures might include the development of new client insights, perceptions about the confidentiality of the conversation (Passmore and von Bartheld, 2024), the working alliance (Graβmann *et al*., 2020), or trust (Schiemann *et al*., 2019).

Client and extratherapeutic factors are theorised to be the single greatest predictors of outcome variance (Asay and Lambert, 1999) in "helping-by-talking" interventions, with one client trait – their level of commitment–important both to the process and to the completion of subsequent actions at the close of the session (Passmore and Whybrow, 2019). High levels of commitment are associated with increased effort towards both the coaching process and goal attainment, which in turn are connected to a greater likelihood of the client achieving their desired outcomes. Tee *et al*.'s (2022) systemic literature review highlighted the importance of client commitment, which is reflected as a key theme across multiple studies (Audet and Couteret, 2012; Baron and Morin, 2009; Bouwer and van Egmond, 2012).

The working alliance, or coach-client relationship, has been suggested as another key predictor of coaching outcomes. Graβmann *et al*., (2020) study results suggested that the working alliance was positively related to desirable coaching outcomes (range: $r = 0.32$ to 0.64), while negatively related to unintended negative effects of coaching ($r = -0.29$). However, not all writers have shared this perspective. DeRubeis *et al*. (2005) suggest that the relationship is mainly determined by the client's ability to form such an alliance, with those same clients being the ones who make the most progress in therapy and, by extension, in coaching. Similarly, de Haan *et al*. (2020) evidenced that the working alliance was not strongly related to coaching's effectiveness. In considering AI coaching, Terblanche and Cilliers (2020) argued that, rather than focussing on working alliances, AI coaching studies should investigate technology acceptance. We assert that, for AI coaching clients, higher levels of technological competence and confidence might contribute towards enhanced outcomes.

A third factor is the clients' perception of confidentiality, fostering a safe environment for clients to openly discuss personal and professional challenges, which is mentioned repeatedly in coaching practice texts (such as Brennan and Wildflower, 2018; Passmore and Sinclair, 2023; Smith and Arnold, 2023) as well as in professional standards (ICF, 2019). Confidentiality is an important ingredient enabling the development of a trusting relationship and thus promoting a genuine dialogue, facilitating deeper insights and meaningful progress (Ebrahimi, 2024). The issue of trust is associated with confidentiality. Trust is argued to be a critical factor in influencing coaching effectiveness (Graβmann *et al*., 2020). Boszoremyi-Nagy (2013) has highlighted the importance of trust within clinical

relationships as a key building block of the relationship. Clients are more likely to disclose more truthful information when they trust the coach to honour their commitments, from respecting confidentiality to working competently and supporting them towards their goals. Related to trust is shame-embarrassment and our willingness to disclose possibly embarrassing information to others. Ellis-Brush (2021) has suggested that a technology-based conversation can be perceived as a safe space because there is no judgement from other people. This aligns with the online disinhibition effect (Suler, 2004), which suggests that people feel less inhibited in cyberspace than in the real world, leading them to behave differently (Suler, 2016). Coaching evaluation remains a widely debated issue within the field (see, for example, Peterson, 1996; Seay and Muscarella, 2024). While the number of evaluation studies remains small, most have drawn on either psychological tools, such as growth mindset and psychological safety (Passmore *et al.*, 2025a), which were not specifically designed for coaching or organisational use. Given organisational measures are space, researchers must develop and experiment with new and different measures.

Based upon the above theory and research, ten hypotheses relating to the comparison of human and AI coaching (Table 1) were proposed. This is a comparatively large number of hypotheses, but given the lack of research and the breadth of factors within the custom coaching questionnaire (Seiler, 2021), we felt this number was justified. The hypotheses were based on six themes. Participants would evaluate the AI coaches more highly on new insights and goal attainment, but poorly on commitment, working alliance and trust compared to human coaches. Finally, there would be no difference in ratings for AI and human coaches on confidentiality.

Method

Design

The study employed a quasi-experimental, between-participant design. There was one categorical independent variable featuring two non-randomly assigned conditions: human coaching and AI coaching, drawn from different organisations. There were six continuous dependent variables: new insights, positive working alliances, goal attainment, commitment, shame and confidentiality.

Participants–human clients

A purposive sampling strategy of volunteer participants, who were aged 18 or over, in professional, knowledge-based (white collar) work and wanting to engage in coaching was adopted. Human coaches were drawn from a range of European Union (EU) and recruited

Table 1. Research hypotheses

#	Hypotheses
H_1	Clients will evaluate the AI coaching positively on new insights
H_2	Clients will evaluate the AI coaching positively on goal attainment
H_3	Clients will evaluate the AI coaching negatively on commitment
H_4	Clients will evaluate the AI coaching negatively on working alliance
H_5	There will be no significant difference in AI coaching and human coaching on new insights
H_6	There will be no significant difference in AI coaching and human coaching on goal attainment
H_7	There will be no significant difference in AI coaching and human coaching on confidentiality
H_8	Clients will evaluate human coaching more highly than AI coaching in generating a positive working alliance
H_9	Clients will evaluate human coaching more highly than AI coaching in generating commitment
H_{10}	Clients will evaluate human coaching more highly than AI coaching in trust

Source(s): Table created by author

directly by human coaches who volunteered to participate in the study. Those in the AI coaching condition were members of a single EU-based organisation and were recruited by the head of learning and development.

The study involved $n = 63$ participants, of which $n = 28$ with a mean age of 41 were assigned to the AI coaching condition and $n = 35$ with a mean age of 47 were assigned to the human coaching condition. Data about the participants' race and gender are in Table 2.

Participants – human coaches

A total of $n = 28$ human coaches were recruited through the researchers' professional networks to participate in the study. Human coaches were ICF- or EMCC-accredited and undertook coaching as a professional activity with clients. Each human coach was restricted to a maximum of three clients. They were required to deliver the coaching sessions online, with each session required to last between 30 and 45 min. There was no guidance provided as to the model or approach that human coaches should adopt.

Materials and measures

For the AI coaching condition, this study used "Alpina", a commercially available AI coaching agent developed by the company "evoach", with permission for its use obtained from the designers. Alpina was running on a GDPR-compliant platform and used a text-based method of communication using Open AI's GPT-4 turbo Large Language Model, with additional prompt coding by the developers to enhance Alpina's coaching capabilities. Alpina was prompted to act as a professional business coach using clean language principles and being 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 the principles of positive psychology. The AI prompts also included clear instructions about not providing advice or suggestions and how to apply the ICF's coaching competencies. In line with ICF requirements, a contracting process was built into the AI coaching condition. Table 3 contains an example of prompts provided to the AI coaching condition participants. All participant data from the AI coaching sessions were supplied to the researchers in anonymised form.

Coaching client attitude measure

A customised measure was used, based upon the EXCBOS coaching evaluation questionnaire (Seiler, 2021), an exploratory instrument for evaluating human coaching. To encourage completion rates, an abbreviated version was created featuring 11 items, each scored using a percentile range and selected from the original 35-item questionnaire. Given Seiler's original focus, small changes in language were needed to ensure retained items made sense for this study's AI coaching condition, such as adding the term "AI" to the item. The featured themes were: (1) new insights, (2) goal attainment, (3) commitment, (4) working alliance, (5) confidentiality and (6) shame. An example item is: "*I felt I could trust my coach with confidential information*". A positive relationship was operationalised as a mean score of 50%

Table 2. Summary of participant demographics

AI coaching condition				
Gender (28)	Male – 10	Female – 16	Non binary –0	Prefer not to say - 2
Race (28)	White - 20	Asian – 3	Black – 0	Prefer not to say - 5
Human coaching condition				
Gender (35)	Male – 7	Female – 27	Non binary –0	Prefer not to say - 1
Race (35)	White - 30	Asian – 2	Black – 2	Prefer not to say - 1

Source(s): Table created by author

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 min max. If I don't agree to this, suggest that we 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 optimised
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

Source(s): Table created by author

or higher for the theme. H_{10} used the mean score from the shame items to determine the participant's level of trust. A list of the questions is provided in Appendix 1. In addition, biodata items relating to age, gender and race were also collected.

Procedure

For the AI coaching condition, written consent was obtained from the host organisation. An e-mailed anonymous link then invited potential participants to find out more about the research and provide their individual consent to participate. For the human coaching condition, informed consent was obtained from individual participants through a written agreement that described the research.

To ensure comparability, clients in both conditions were offered a single coaching session, after which they were invited to complete the coaching client attitude measure. Data were collected using SurveyMonkey, accessible via the participant's smartphone or computer. All responses were anonymised. No data were collected from the human coaching condition participants detailing which human coach provided which session.

Ethical approval was secured from a university ethics committee.

Results

Dataset properties and descriptive analysis

Prior to analysis, skewness and kurtosis were assessed for all variables. The Shapiro–Wilk test indicated that none of the variables were normally distributed (all $p > 0.05$). In addition, two variables (shame and confidentiality) displayed suboptimal skewness and kurtosis. No data transformations were applied due to the small sample size. The means and standard deviations for all variables are displayed in Table 4.

Hypotheses testing

Given the sample sizes of both the AI and human coach conditions, a series of independent-samples Mann–Whitney *U* tests were conducted.

Participants evaluated their human coaches significantly higher on new insights (Mdn = 82.5) and goal attainment (Mdn = 90) compared to participants who had AI coaches (Mdn = 42.5 and Mdn = 27.75, respectively), and these differences were significant

Table 4. Means and standard deviations of the variables

Variable	Human coach			AI coach		
	N	Mean	Standard deviation	*N*	Mean	Standard deviation
New Insights	35	83.33	12.94	28	40.29	32.05
Positive working alliance	35	88.23	11.59	28	42.61	32.76
Goal attainment	35	86.80	13.73	28	38.29	32.41
Commitment	35	81.56	15.64	28	47.43	35.82
Shame	35	91.60	8.06	28	87.45	20.10
Confidentiality	35	92.26	10.82	28	75.36	34.55

Source(s): Table created by author

($U = 116.5$, $p < 0.001$ and $U = 84$, $p < 0.001$, respectively). Therefore, the null hypotheses for H_1 and H_2 cannot be rejected.

The results indicated support for H_3 and H_4, indicating that participants assessed their human coaches significantly higher on commitment (Mdn = 82.5) and on generating a positive working alliance (Mdn = 89.5) compared to participants with AI coaches (Mdn = 50 and Mdn = 35.25, respectively). Both differences were statistically significant ($U = 222$, $p < 0.001$ and $U = 507$, $p < 0.001$, respectively).

No significant differences were found between human (Mdn = 94) and AI coaches (Mdn = 98) on participant evaluations of trust ($U = 963$, $p = 0.343$). Similarly, there was no significant difference between human (Mdn = 97) and AI coaches (Mdn = 97.5) on participant assessments of confidentiality ($U = 446.5$, $p = 0.528$). Therefore, the null hypotheses for $H_5$and H_6 cannot be rejected.

Human coaching was considered more confidential (H_7) and scored as developing a stronger working alliance (H_8), developing commitment (H_9) and on experiencing no shame (H_{10}) (i.e. feeling judged by the human or AI coach), although the difference between the human and the AI coaching agent for H_{10} was not significant: Human coach (Mdn = 94), AI coach (Mdn = 98), $U = 963$, $p = 0.343$).

Discussion

The results from this study provide a complex picture, neither supporting the view that AI's coaching created an existential threat to the human coach (Diller *et al*., 2024), nor the view that AI coaching agents are not coaches (Bachkirova and Kemp, 2024). Instead, a more nuanced perspective is offered, reflecting both the strengths of AI as an emerging technology, which can engage in one-to-one conversations and its weaknesses at the time of the study in 2024, in high trust conversations such as coaching. In particular, it must be considered that AI's advancements move at a faster rate than the academic research dissemination process, so it would be wrong to draw conclusions as to all AI coaching's capabilities in 2025 and beyond, based upon findings generated in 2024.

Nonetheless, based on the results from this small-scale quasi-experimental study with professional, white collar workers, it appears that, as a population, professionals have a general preference for human-to-human relationships in coaching based on a one-off session. Participants rated the human coaches more positively in their ability to create a positive working alliance than the AI coach. Similarly, the human coaches were also rated more highly in terms of client commitment and goal attainment.

Evidence from Graβmann *et al*. (2020) suggests the importance of the working relationship to effective outcomes such as goal attainment, whereas findings from a large-scale longitudinal Randomised Control Trial (RCT) by de Haan *et al*. (2020) suggest that the working alliance is not strongly related to coaching effectiveness over the full coaching

relationship, only at the start. This might therefore have a distorting effect on the evaluation of a single session coaching intervention, as used in this study, with the recognition that this design has low ecological validity. There is evidence that client motivation and commitment are important factors in the coaching process (Rayiu and Baban, 2012).

Other aspects not explored in this study, such as humour and presence, are currently difficult for AI coaching agents to replicate, as is the contextual sensitivity displayed by some human coaches who have lived experience and direct organisation working experience. This may allow them to understand wider organisational dynamics, cultural factors, or the "lived experience" of their clients.

In this sense, AI at the time of this study was behind the subset of human coaches who bring these elements into their work and, through this, enable clients to feel heard, understood and respected for the unique human being they are (Passmore and von Bartheld, 2024). For those human coaches unable to bring these qualities to their work, an AI coach's ability to engage at a transactional level makes their coaching offers comparable.

We hypothesised that the AI coaching agent would be rated more highly by participants for both confidentiality and shame (embarrassment). These hypotheses were based on Suler's (2004) Theory of Disinhibition, which suggests that inhibitions are reduced as distance grows, for example, from face-to-face to online and, by extension, from online human-to-human to human-to-AI coaching agent. Based on this theory, we expected AI coaches to score more highly on the two sub-scales of shame (embarrassment) and confidentiality. However, there was no statistically significant difference between participant ratings of AI and human coaches.

The fact that participants rated the human and AI coaches at similar levels may be related to the participants' awareness that they were taking part in a research study and that human researchers would be accessing the details of their sessions. In a pure AI coaching session, participant ratings of confidentiality might be increased by making the AI coach more clearly articulate the confidential nature of the conversation, affirming that no human will access the data. Future research is needed to test this hypothesis.

We maintain the view that, for some coaching topics where embarrassment or shame is a feature, the AI coach may offer a more appropriate space for the client to disclose and explore this content. This may only occur where the topic constitutes one that would engender a strong emotional response of embarrassment. Such topics are more likely to occur in, for example, counselling on sexual practices, as opposed to coaching about task prioritisation or career development at work. These effects may therefore be less relevant in work-based coaching. Furthermore, these effects may only be seen when clients feel in control of their data; that is, researchers and others do not have access to the contents of the session.

Aspects which we did not measure in this study include coaching client evaluations of the cost comparison between AI and human coaching or convenience (needing to schedule diaries in advance with a human coach or having access whenever required, 24–7, to an AI coach). AI coaching agents offer significant advantages. They are always on, never get tired, or take holidays and are available at a fraction of the cost of a human coach. AI can also be configured to engage in any language and offer both text and speech options for engagement, making them more accessible for different user preferences or requirements.

Our literature-informed hypotheses suggested that the AI coach might perform strongly on some of the six factors we measured, either matching or outperforming the human coach. However, the results indicated lower average ratings across all factors.

Data from this study showed a large standard deviation (SD) for the client responses to AI coaching, suggesting a wider range of feelings from different participants (Figures 1 and 2). This was most marked in goal attainment and new insights, where some participants rated AI coaching at a similar level to human coaching, whilst others rated it significantly worse. This contrasts with the participant response for other factors such as Shame (Figure 3).

These results suggest that, for some clients, the experience of AI coaching met their expectations in terms of insights and goal attainment and was comparable to the ratings given

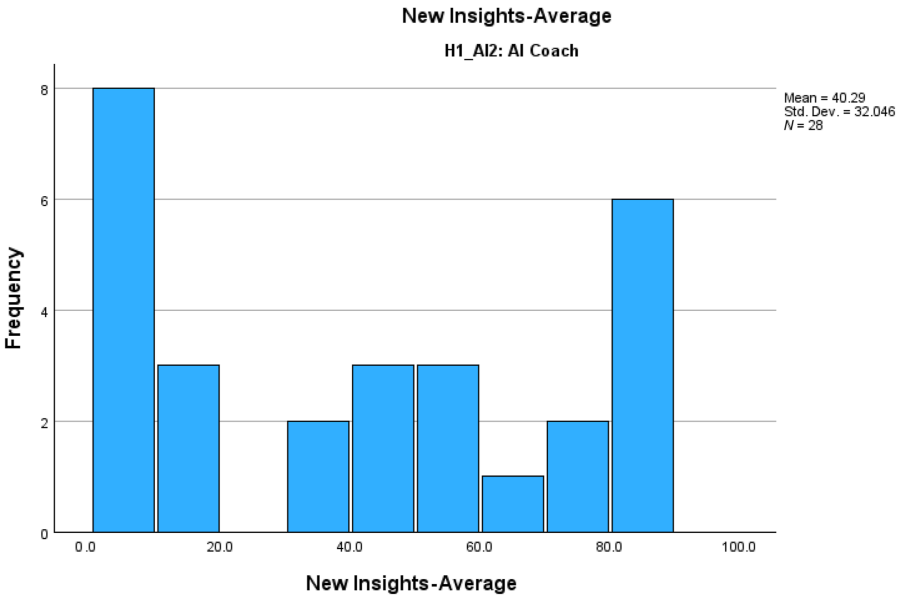

Figure 1. Histogram AI Coachbot – New Insights. Source: figure created by author

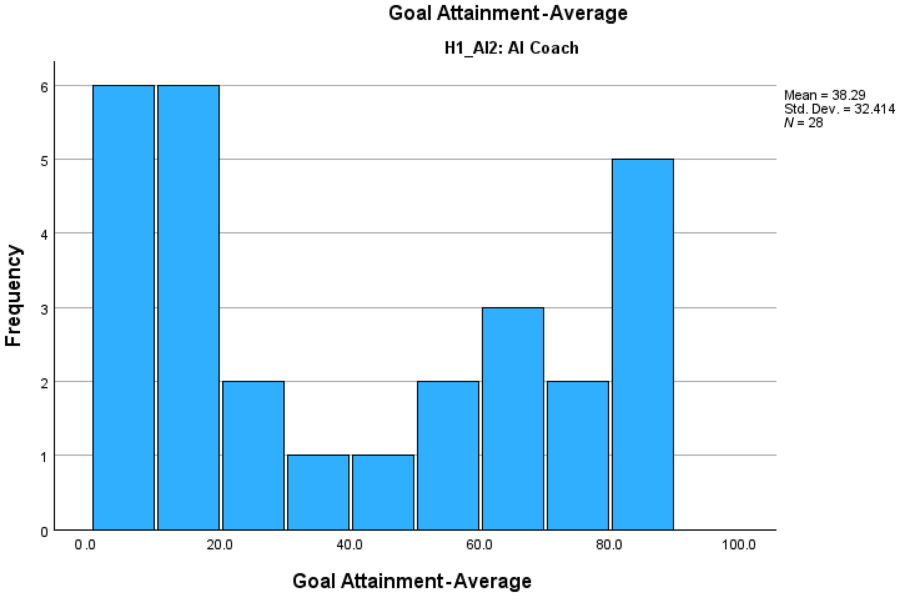

Figure 2. Histogram AI Coachbot – Goal attainment. Source: figure created by author

by participants of their experience with human coaches. However, for others, the AI experience was rated as significantly lower than participant ratings of the human coach experience. Extreme caution is needed, given the small sample size and the use of a custom questionnaire employed in this exploratory study. Alternative explanations may include: client

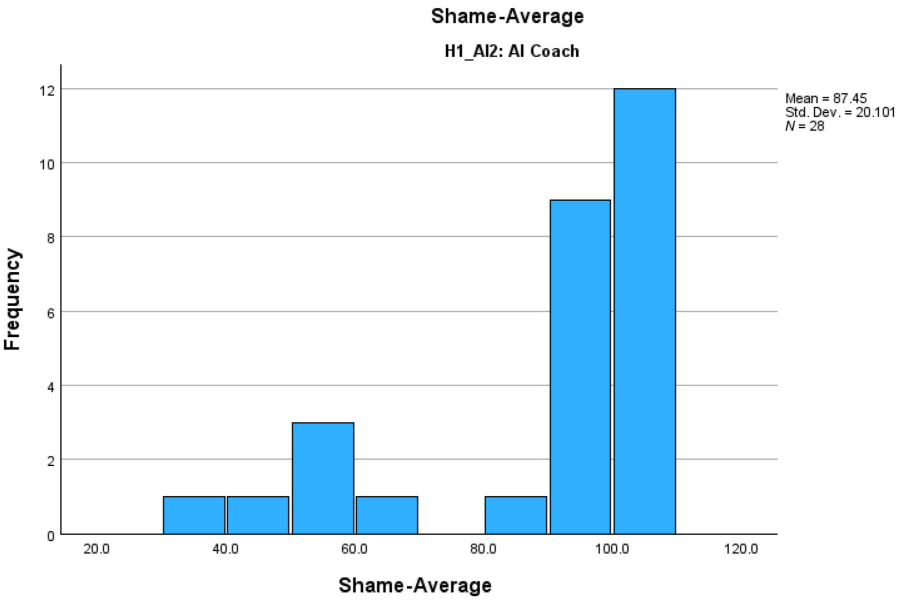

Figure 3. Histogram AI Coachbot – AI- Shame. Source: figure created by author

factors such as age (younger participants are more open to AI technology than older users): technology confidence (Terblanche and Cilliers, 2020), with more confident technology users being more engaged and giving higher ratings than less confident users: job role (data roles being more engaged with higher ratings than users from roles focussing on interpersonal relationships): or personality preferences (with introversion or neurodiverse characteristics contributing to higher engagement); as well as the items used in the questionnaire. Furthermore, the text-based nature of the AI coaching agent used in this study required text-to-text communication, encouraging the client to crystallise their thinking prior to committing it to text. Thoughts in human coaching can sometimes be less considered and more emergent, which may have influenced these results.

As AI capability accelerates, further research is needed to explore voice-to-voice coachbots, avatar-based coachbots and coachbots that can meaningfully read client micro-expressions and use this data to shape how they coach. Our hypothesis is that the greater the anthropomorphising of the tool, such as through voice-to-voice (with options on different accents and languages), the use of a 2D or 3D photorealistic codec avatar and other humanising features, each will contribute to higher levels of engagement, a stronger client sense of psychological safety and higher levels of client rating.

The intriguing nature of the results in this study brings into question whether the outcomes from previous student-based and goal attainment coaching studies can be directly transferred to workplace coaching. Further research is needed to explore this question. This study also suggests that our relationship with AI coaches may be more complex than initially assumed, and that AI coaching providers may have a different experience than human coaching, which requires different expectations and assumptions, rather than direct comparison with the human coaching experience.

Practical implications

To date, studies exploring the potential of AI coaches have focused on student populations and goal attainment or used human masking as an AI tool ("Wizard of Oz" designs). This study

employed a commercially available AI coaching agent with real workplace coaching clients exploring real-world issues. It sought to evaluate their engagement beyond goal attainment, considering factors such as perceptions of trust, confidentiality and insight. The study's results suggest that while AI coaches, as of 2024, may have been capable of holding coach-like conversations (Passmore *et al*., 2025c), users, on average, preferred a human experience, with a portion of participants rating the experience as similar to those rating a human coach, and others rating it significantly worse. As technology continues to evolve, with the development of voice-to-voice and avatar-based tools and as user expectations evolve, different outcomes may emerge.

Limitations

The primary limitation of the study was the introduction of a custom measure. Although the items were a subset selected from the existing Executive Coaching Behaviour Observation Scales (EXBOS) instrument (Seiler, 2021), the study sought to measure six variables using an 11-item scale and did so as a pilot without conducting any factor analysis. This might account for the high SD values obtained in the dataset. Thus, care must be taken in generalising these findings. A second limitation was the diversity of the samples. While both participant groups were white-collar professionals, the individuals were drawn from different organisations. A range of factors, not least organisational culture, may have impacted the results. Thirdly, individuals were volunteers who chose to engage with either a human or AI coach, as opposed to being randomly allocated and instructed to participate. Factors such as user perception may have impacted the results. Given the relative novelty of AI coaching, those who were offered the AI coach may have seen this as an opportunity to test out this new technology, rather than to engage in meaningful coaching. In contrast, those choosing human coaches may have been aware that human coaching is a well-established organisational intervention. Fourth, the AI coaching condition participants were advised that they could either delete or save the conversation. If they agree to save it, it would then be reviewed by the researchers. Thus, participants were aware that humans would also be viewing their content in detail. This may have impacted on views about the confidentiality of the conversations and shame. A fifth limitation was the relatively small sample size, which allowed for comparative analysis between the AI and human coaching but did not allow for exploration of differences in age, gender, race, or differentiating factors. A further limitation was that the evaluation was based on a one-off (single), coaching session. Attitudes may change if the assessment takes place after, say, the 4th or 6th session. In addition, there was no guidance or requirement for the human coaches to adopt any given approach, structure, or tool, meaning there would have been some variation in the coaching experience for participants in the human coaching condition. Some human coaches may have slipped into "tell" mode, some may have rigidly adhered to prescribed coaching frameworks, some may have detoured into orthodox counselling territory and so on, with this variance making it difficult to treat the human coaching intervention as a meaningful construct for research purposes. Finally, the sample size in this small-scale study was insufficient to detect anything other than large effects. Future replication studies would be strengthened by featuring larger, more diverse samples.

Conclusion

The results from this study challenge previous research findings, which have claimed similar outcomes from engagement with AI and human coaches (Hassoon *et al*., 2021; Terblanche *et al*., 2022a). We suggest that greater caution is needed in claiming the comparability of AI coaches with human coaches in applied settings and that more research is needed to better understand client responses to AI coaching. However, given the significant cost and convenience advantage AI coaching presents compared to human coaching, they are likely to have a future role in workplace learning and development. But in designing these agents, the

desired outcomes and the user experience should be central, with the aim of creating an easy-to-use and flexible tool that generates individual and organisational benefits.

About the authors

Prof Jonathan Passmore is a professor at Henley Business School, UK and Senior VP EZRA, LHH. He has published widely with over 100 peer review papers and 40 books.

Dr David Tee is an academic based at the Institute of Continuing Education, University of Cambridge, UK and also works in private practice.

Dr Gina Palermo is a researcher at EZRA Coaching, LHH and a specialist in psychometric testing and measurement.

Rebecca Rutschmann is a coach, an expert in AI and an educator based at Viva la Coaching Academy, Germany

Ethical statement

The research was approved by Henley Business School–Approval number – SREC-HBS-20240209-JOPA9549.

Supplementary material

The supplementary material for this article can be found online.

References

Asay, T.P. and Lambert, M.J. (1999), "The empirical case for the common factors in therapy: quantitative findings", in Hubble, M.A., Duncan, B.L. and Miller, S.D. (Eds), *The Heart and Soul of Change: What Works in Therapy*, American Psychological Association, pp. 23-55, doi: 10.1037/11132-001.

Athanasopoulou, A. and Dopson, S. (2018), "A systematic review of executive coaching outcomes: is it the journey or the destination that matters the most?", *The Leadership Quarterly*, Vol. 29 No. 1, pp. 70-88, doi: 10.1016/j.leaqua.2017.11.004.

Audet, J. and Couteret, P. (2012), "Coaching the entrepreneur - features and success factors", *Journal of Small Business and Enterprise Development*, Vol. 19 No. 3, pp. 515-531, doi: 10.1108/14626001211250207.

Bachkirova, T. and Kemp, R. (2024), "'AI coaching': democratising coaching service or offering an ersatz?", *Coaching: An International Journal of Theory, Research and Practice*, Vol. 18 No. 1, pp. 27-45, Pre-print, doi: 10.1080/17521882.2024.2368598.

Balasubramanian, S. (2023), "Exploring the capabilities of ChatGPT in natural language process tasks", *Journal of Artificial Intelligence and Machine Learning*, Vol. 2 No. 1, pp. 7-17, doi: 10.17605/OSF.IO/XJYMQ.

Bamber, A. (2025), "Artificial intelligence vs. human coaches: examining the development of working alliance in a single session", *Frontiers in Psychology*, Vol. 15, 1364054, doi: 10.3389/fpsyg.2024.1364054.

Baron, L. and Morin, L. (2009), "The coach–coachee relationship in executive coaching: a field study", *Human Resource Development Quarterly*, Vol. 20 No. 1, pp. 85-106, doi: 10.1002/hrdq.20009.

Boszoremyi-Nagy, I.K. (2013), *Between Give and Take: A Clinical Guide to Contexural Therapy*, Routledge.

Bouwer, J. and van Egmond, J. (2012), "Moderating factors of the van Egmond coaching model (VECM)", *International Coaching Psychology Review*, Vol. 7 No. 1, pp. 55-63, doi: 10.53841/bpsicpr.2012.7.1.55.

Brennan, D. and Wildflower, L. (2018), "Ethics and coaching", in Bachkirova, T., E Cox, E. and Clutterbuck, D. (Eds), *The Complete Handbook of Coaching*, 3rd ed., Sage Publications, pp. 500-517.

de Haan, E., Molyn, J. and Nilsson, V.O. (2020), "New findings on the effectiveness of the coaching relationship: time to think differently about active ingredients?", *Consulting Psychology Journal: Practice and Research*, Vol. 72 No. 3, pp. 155-167, doi: 10.1037/cpb0000175.

Dell'Acqua, F. ,McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R., (2023), Navigating the jagged technological frontier: field experimental evidence of the effects of AI on knowledge worker productivity and quality (september 15, 2023). Harvard Business School Technology and Operations Mgt. Unit Working Paper No. 24-013, The Wharton School Research Paper, doi: 10.2139/ssrn.4573321, SSRN, available at: https://ssrn.com/abstract=4573321.

DeRubeis, R.J., Brotman, M.A. and Gibbons, C.J. (2005), "A conceptual and methodological analysis of the non-specifics argument", *Clinical Psychology: Science and Practice*, Vol. 12 No. 2, pp. 174-183, doi: 10.1093/clipsy.bpi022.

Diller, S.J., Stenzel, L.-C. and Passmore, J. (2024), "The coach bots are coming: exploring coaches' attitudes and responses to the threat of AI coaching", *Human Resource Development International*, Vol. 27 No. 4, pp. 597-621, doi: 10.1080/13678868.2024.2375934.

Ebrahimi, M., Akbari Asbagh, F., Davari Tanha, F., Amirkhanloo, F., Sahraiyan, G., Feizabad, E. and Lotfi, S. (2024), "The pulse of connection: trust and confidentiality in internal coaching dynamics", Vol. 181, pp. 55-59, doi: 10.5590/jsbhs.2024.18.1.22, available at: https://ssrn.com/abstract=5055315

Ellis-Brush, K. (2021), "Augmenting coaching practice through digital methods", *International Journal of Evidenced based Coaching and Mentoring*, Vol. S15, pp. 187-197, doi: 10.24384/ER2P-4857.

Graßmann, C., Scholmerich, F. and Schermuly, C.C. (2020), "The relationship between working alliance and client outcomes in coaching: a meta-analysis", *Human Relations*, Vol. 73 No. 1, pp. 35-58, doi: 10.1177/0018726718819725.

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. and Appel, L.J. (2021), "Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors", *npj Digital Medicine*, Vol. 4 No. 1, p. 168, doi: 10.1038/s41746-021-00539-9.

International Coaching Federation (2019), "Coach competencies", available at: https://coachingfederation.org/credentialing/coaching-competencies/icf-core-competencies/

Kaplain, A. and Haenlein, M. (2019), "Siri, Siri in my hand who is the fairest in the land: on the Interpretation and implications of Artificial Intelligence", *Business Horizons*, Vol. 62 No. 1, pp. 15-25, doi: 10.1177/0008125619864925.

Passmore, J. and Tee, D. (2024), "Can Chatbots like GPT-4 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 von Bartheld, J. (2024), *Becoming a Credentialed ICF Coach: The Journey to ACC, PCC and MCC – Practice Insights from ICF Assessors*, Libri Publishing.

Passmore, J., Tee, D. and Gold, R. (2025a), "Team coaching using LSP and team facilitation: a randomized control trial study measuring team cohesion and psychological safety", *Journal of Work-Applied Management*, Vol. 17 No. 1, pp. 119-132, doi: 10.1108/JWAM-12-2023-0137.

Passmore, J., Olafsson, B. and Tee, D. (2025b), "A systematic literature review of artificial intelligence (AI) in coaching: insights for future research and product development", *Journal of Work – Applied Management*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/JWAM-11-2024-0164.

Passmore, J., Tee, D. and Rutschman, R. (2025c), "Getting better all the time: using professional coach competencies to evaluate the quality of AI Coach bot performance", *Coaching: An International Journal of Theory, Research and Practice: An International Journal of Theory, Practice and Research*, Pre-print, doi: 10.1080/17521882.2025.2510643.

Passmore, J. and Sinclair, T. (2023), *Becoming a Coach: The Essential ICF Guide*, 2nd ed., Pavilion.

Passmore, J. and Whybrow, A. (2019), "Motivational interviewing", in Palmer, S. and Whybrow, A. (Eds), *Handbook of Coaching Psychology*, Routledge, pp. 144-153.

Peterson, D.B. (1996), "Executive coaching at work: the art of one-on-one change", *Consulting Psychology Journal: Practice and Research*, Vol. 48 No. 2, pp. 78-86, doi: 10.1037/1061-4087.48.2.78.

Rayiu, L. and Baban, A. (2012), "Executive coaching as a change process", *Brain behavior*, Vol. 16 No. 10, pp. 139-145.

Schiemann, S.J., Mühlberger, C., Schoorman, F.D. and Jonas, E. (2019), "Trust me, I am a caring coach: the benefits of establishing trustworthiness during coaching by communicating benevolence", *Journal of Trust Research*, Vol. 9 No. 2, pp. 164-184, doi: 10.1080/21515581.2019.1650751.

Seay, S. and Muscarella, A. (2024), "The effectiveness of executive coaching: executive views and metrics", *Journal of Human Resource and Sustainability Studies*, Vol. 12 No. 02, pp. 371-388, doi: 10.4236/jhrss.2024.122020.

Seiler, H. (2021), *Using Client Feedback in Executive Coaching*, Oxford University Press.

Smith, W.-A. and Arnold, C. (2023), "Ethical coaching and its relationships", in Smith, W.-A., Passmore, J., Turner, E., Lai, T.-L. and Clutterbuck, D. (Eds), *The Ethical Coaches' Handbook*, Routledge, pp. 3-21.

Suler, J.R. (2004), "The online disinhibition effect", *CyberPsychology and Behavior*, Vol. 7 No. 1, pp. 321-326, doi: 10.1089/1094931041291295.

Suler, J.R. (2016), *Psychology of the Digital Age: Humans Become Electric*, Cambridge University Press.

Tee, D., Misra, K., Roderique-Davies, G. and Shearer, D. (2022), "A systematic review of coaching client characteristics", *International Coaching Psychology Review*, Vol. 17 No. 1, pp. 50-74, doi: 10.53841/bpsicpr.2022.17.1.50.

Terblanche, N. and Cilliers, D. (2020), "Factors that influence users' adoption of being coached by an Artificial Intelligence Coach", *Philosophy of Coaching: International Journal*, Vol. 5 No. 1, pp. 61-70, doi: 10.22316/poc/05.1.06.

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.

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.

Terblanche, N., Molyn, J., Williams, K. and Maritz, J. (2023), "Performance matters: students' perceptions of Artificial Intelligence Coach adoption factors", *Coaching: An International Journal of Theory, Research and Practice*, Vol. 16 No. 1, pp. 100-114, doi: 10.1080/17521882.2022.2094278.

Terblanche, N.H.D., Van Heerden, M. and Hunt, R. (2024), "The influence of an artificial intelligence chatbot coach assistant on the human coach-client working alliance", *Coaching: An International Journal of Theory, Research and Practice*, Vols 1-18 No. 2, pp. 189-206, doi: 10.1080/17521882.2024.2304792.

Wang, Q., Lai, L.Y.-L., Xu, X. and McDowall, A. (2022), "The effectiveness of workplace coaching: a meta-analysis of contemporary psychologically informed coaching approaches", *Journal of Work -Applied Management*, Vol. 14 No. 1, pp. 77-101, doi: 10.1108/JWAM-04-2021-0030.

Corresponding author
Jonathan Passmore can be contacted at: jonathancpassmore@yahoo.co.uk