5-2-2023

WHERETO FOR AUTOMATED COACHING CONVERSATION: STRUCTURED INTERVENTION OR ADAPTIVE GENERATION?

Andreas Göldi
University of St.Gallen, andreas.goeldi@unisg.ch

Roman Rietsche
University of St.Gallen, roman.rietsche@unsig.ch

WHERETO FOR AUTOMATED COACHING CONVERSATION: STRUCTURED INTERVENTION OR ADAPTIVE GENERATION?

Research in Progress

Göldi, Andreas, University of St.Gallen, andreas.goeldi@unsig.ch,

Rietsche, Roman, University of St.Gallen, roman.rietsche@unsig.ch,

Abstract

In an age of life-long learning, it is important that adult learners can effectively use their motivation and resources to reach their learning goals. In conversation, coaches can intervene to promote learning goal attainment by using behavioural change techniques (BCTs). In a coaching chatbot, such techniques can be ordered in an established, structured way to good effect. With recent technological advances, chatbot responses can be generated adaptively; this means that BCTs can be applied in an adaptive but less structured way. It is yet unclear whether this latter form of configuring coaching interventions is viable, how they compare to more established structured interventions, and whether both methods can be combined. For the purpose of answering this, we propose a 2x2 experimental design with the two intervention types as factors and goal attainment as the dependent variable. Results will indicate avenues for automating skilled conversation including choice of technology.

Keywords: Coaching, Conversational AI, Chatbots, Adult Learning.

1 Introduction

Recent research has demonstrated evidence for parity between human and chatbot coaches in adult education (Terblanche et al., 2022). With the growing importance of life-long learning (Ates and Alsal, 2012), the share of adults in education is likely to co-expand. While adult learners are relatively self-directed (Knowles, 1984), and thus traditionally not provided with close teacher-student supervision, they are known to lack social support and may thus be in need of professionals providing it (Lundberg, McIntire, and Creasman, 2008). The optimal teacher-student ratio in adult education has been variously assessed, e.g., at 1:4 (Dubrowski and MacRae, 2006) or about 1:9 (Mukhopadhyay, Sudarsan, and Tapaswi, 2020). The actual number is at 1:15 or 1:17 in public and private institutions, respectively (OECD, 2022). This issue is not transitory. As ever more education is moving online (Maatuk et al., 2022), the intrinsic nature of adult learners learning motivation (Cox, 2015) fails to ensure that they fulfill their intent: Non-completion is a persistent problem in online higher education (Delnoij et al., 2020).

Coaching is a professional helping relationship based on dialogue that is supposed to restructure a clients thoughts, values, and feelings in accordance with their personal aspirations (Passmore and Lai, 2020). To keep learners enrolled and promote academic performance, it is beneficial that they have developed goals (Harackiewicz et al., 2000). Goals are cognitive representations of desired behaviour outcomes (Moskowitz and H. Grant, 2009). They serve an important role in guiding sustained action, outcomes tracing what was resolved to be achieved (Latham and Kinne, 1974). Setting goals in the company of a supportive other leads people to aim higher (Latham and Saari, 1979). Simply monitoring goal progress leads to higher attainment; an effect that is even more pronounced if shared with others (Harkin et al., 2016).

To achieve better outcomes in goal attainment, human coaches on the one hand use evidence-based dialogue recipes (A. M. Grant, 2017). They also talk with, and not at, their clients (Armstrong, 2012): Goal pursuit coaching may be effective by following rules or adapting to circumstance. In both cases, goal-directed behaviour patterns are promoted. This is possible using different so-called behaviour change techniques (Michie et al., 2013). Two pathways are open to behaviour change and the resulting goal attainment. They may be characterized in terms of their makeup: Long, evidence-based recipes are structured interventions, while the more adaptive strategy may use only few or single utterances. Naming them after this, they are nano-interventions. Both structured and nano-interventions promote the same aim. However, the behavioural change techniques employed therein are differently configured: structured or adaptive.

Coaching chatbots are mainly researched in the health and education domain, where interventions aim to produce beneficial changes in users (Pande, Fill, and Hinkelmann, 2022). In the health domain, behaviour change models are focused mainly on how relational and persuasive capacity can be achieved (J. Zhang et al., 2020). As a review notes, while chatbots in education incorporate adaptation to user traits, most do not generate responses adaptive to the prior conversation (Wollny et al., 2021); furthermore, most chatbots aim not at a coaching but mentoring role, which is less formalized (Hastings and Kane, 2018) and therefore does not rely as much on intervention capabilities.

Chatbots differ in how conversation flows are controlled (Rasa, 2021) and whether responses are retrieved or generated (Winkler and Soellner, 2018). More recent work used a generative approach to coaching chatbots that successfully used structured intervention knowledge in the form of scenarios to generate responses conditionally (Zorrilla and Torres, 2022). Recent advances in conversational AI enable even better conditioned response generation (Peng et al., 2022). Resultingly, the difference between structured and nano-interventions becomes a matter of great importance for any chatbot project attempting to intervene in human experience and behaviour. If one configuration of intervention techniques is more suitable to the task, this determines the technological implementation.

This results in the following two research questions:

RQ1: *Is adaptive response generation to apply BCTs in nano-interventions viable to improve goal attainment?*

RQ2: *How do adaptive nano-interventions compare to structured interventions for promoting goal attainment?*

RQ3: *How does the combination of nano- and structured interventions compare to each method used separately?*

After discussing the automation of coaching, and giving a narrative review on coaching practice and research, the dialogue necessary for coaching interventions will be theoretically derived. Following this, current trends in natural language processing are discussed in their applicability to the problem. And lastly, the development and evaluation of a coaching chatbot artifact is outlined.

2 Is Coaching Amenable to Automation?

Being based on a relationship, coaching may be thought of as fundamentally linked to human coaches. However, far from being a universal benefit, the personality and behaviour of coaches in human-to-human coaching is the greatest barrier to coaching success (Blackman, A., & Carter, A., 2014). Furthermore, even though coaches need to possess a minimum set of core competencies (Crowe, 2017), and are considered professionals (Passmore and Lai, 2020), most coaches lack any psychological training (McKelley and Rochlen, 2007). Even more; there is not even consensus on how to train a coach (Bachkirova, G. Spence, and Drake, 2017). Consequently, conversational AI that is fed from a relevant knowledge base may act closer to evidence-based practice than many human practitioners. To address this, coaches would need to be better educated. Individual coaches improve by attending to teachers and reading gathered knowledge that has been written down. Modern technology is at least as capable of ingesting knowledge; and it can be improved for a longer time, without deterioration. A coaching machine is capable of monotone improvements, while individual human improvement stops at the very least at death.

Besides scalability into the future in terms of monotone improvements, a more mundane way in which coaching chatbots may outperform human coaches is in terms of cost. Coaching is a time-intensive business which ideally requires specially trained, competent individuals. Lömker, Weber, and Moskaliuk (2021) cite Rauen (2020), who estimates the average hourly rates of coaches in Germany in the management domain at a high EUR 291.11, while coaches in other domains still charged EUR 115.50. In order to achieve wide-spread uptake of coaching resources, affordability needs to be improved. Conversational AI tools may help to achieve scalability for coaching (Terblanche et al., 2022). It is unlikely that all coaches will be replaced by automated coaching conversations, yet the employment of doubtfully qualified coaches testifies to a demand for coaching which can only be supplied using scalable solutions.

Coaching dialogues require a mapping of coachee- or user-provided content onto a relatively restricted set of mainly structural coach responses (Bachkirova, G. Spence, and Drake, 2017). Neural networks in AI are function approximators (Hornik, Stinchcombe, and White, 1989), rendering neural AI capabilities apt to exploit mapping of bot responses to user utterances. Conversational AI is the application of natural language technology to conversational problems, say in chatbots (Gao, Galley, and Li, 2018). From their very beginnings, chatbots have been used in a psychological role: going back to the seminal rule-based bot ELIZA (Weizenbaum, 1966). Since then, advances in natural language processing have made the technology useful for coaching conversations in the domain of education (Terblanche et al., 2022).

Lömker, Weber, and Moskaliuk (2021) summarise the state of current research in coaching chatbots. During the Corona pandemic in the years following 2020, anti-virulence measures forced institutions to move on-premises activities, such as education (Maatuk et al., 2022), online. Many psychotherapists were convinced moving some of their activities there (Békés and Aafjes-van Doorn, 2020). This has benefited conversational agents or social robots, which were increasingly used as liaison service providers or well-being coaches (Aymerich-Franch and Ferrer, 2022). An accompanying rise was seen in mental health coaching apps (Lindsay, Baker, and Calder-Dawe, 2022). Furthermore, the pandemic years brought with them a breakthrough for tele-coaching, which has been on the rise since the early 2010s (Lömker, Weber, and Moskaliuk, 2021). The remaining challenge is to get to systems using conversational AI that continue this trend into more advanced channels of intervention, leaving expensive experts behind. Using conversational AI, transformational approaches using technology for new purposes in coaching may be enabled (Cushion and Townsend, 2019). An especially interesting use of technology in the coaching domain is the full automation. Recent research has indicated that coaching chatbots can reach parity with human coaches in terms of long-term goal attainment (Terblanche et al., 2022).

3 Theoretical Background

3.1 Coaching

Coaching differs in the mode of dialogue employed. Most coaching dialogues are solution-oriented. Others are dialogic, meaning they are based on achieving mutual understanding. And yet others are dialectical, following the tradition of the ancients by helping coachees to rethink implicitly held schemata (John L. Bennett and Francine Campon, 2017). While the last way of dialoguing is a useful asset in cognitive behavioural coaching (Neenan, 2008), its focus on conceptual clarification does not generalise well to all aspects of coaching conversations. Coaching is based on the coach-coachee relationship (Passmore and Lai, 2020). This relationship is built on the exploration of a shared understanding that is as much affective as it is conceptual. For example, appreciation may help professionals enact positive change with their interlocutors (Rogers, 1951). In psychotherapy research, factors common to all schools and procedures have been identified, among them the helping relationship (Gelo, Pritz, and Rieken, 2015).

Verbal behaviour such as validating patients' experience or making positive comments strengthens such relationships (Duff and Bedi, 2010).

A second informative taxonomy groups coaching as a discipline into four discourses (Western, 2017): there is the coach as a soul guide, the network coach, the managerial coach, and the psychological expert. According to Bachkirova (2011), 80% of the world's population have a deep interest in the spiritual. Members of the relatively novel profession of network coaches work with their client on achieving an overview and analysis of the coachee's personal network. Managerial coaches help to handle and improve performance in the different roles management personnel needs to inhabit. The fourth and last discourse psychological coaching which largely focuses on behaviour, leaving aside aspects of spirituality or self. These discourses are not meant to be exclusionary; actual practice can be eclectic in its adherence to the various conceptualisations of coaching delineated herein. The last discourse from the coaching literature mentioned is most amenable to "evidence-based" (Sackett et al., 1996) methodology. Evidence-based coaching was first used as a term describing the use of theory and an empirical basis in contrast to coaching derived from pop psychology concepts (A. M. Grant, 2017). It has since been used to distinguish coaching that uses state-of-the-art empirical knowledge to inform its practice.

All four discourses have their strengths. If coaching is centered on adult learners as recipients, managerial and network considerations are less directly relevant than spiritual and psychological coaching, which target the person. While the specific aims of spiritual coaching are not well defined, it puts the inner life of the coachee at the center. In combination with dialogic practice, this discourse may inform design features linked to common factors from psychotherapy research as well as less clinically dominated literatures such as flourishing or self-reflection. Psychological coaching on the other hand is well suited to be used in designing specific solution-oriented conversation elements, i.e., interventions.

In summary, coaching dialogues may be held in a dialogic or solution-oriented mode. As coaching is built on a relationship that is established largely with dialogic methods, both modes ought to be combined. Both a holistic-existentialist discourse and an evidence-based psychological discourse may inform coaching dialogues. The latter especially is relevant to locate evidence-based intervention techniques. The remaining question is how these interventions should be constituted. This can be further narrowed down by turning to dialogue, and coaching dialogues in particular.

3.2 Dialogue

Coaching is practised using dialogue (Armstrong, 2012). In dialogue, what is done with talk is more important than what the talk is about (Schegloff, 2007). Utterances raise the expectation of an appropriate utterance in response. To get it, supplementary utterances before, between and and after the base utterance can be used. This has empirical support and seems to be language-universal (Kendrick et al., 2020). Properties of this sequence organisation largely generalise to text-based chats as well (Tudini, 2015).

3.3 Promoting Goal Attainment with Dialogue

For coachees, the most frequently used utterances involve self-disclosure and delivering case specific information (Gray, 2017): 70% of coach utterances are questions. Coaches paraphrase coachees, for example to signal active listening (Weger et al., 2014). A part of the conversation will have to involve answers to coachee's questions. Some utterances may be purely structural (Schegloff, 2007), for example aiming to repair conversations. Of course, all these utterances can be thought of as action (Austin, Urmson, and Sbisà, 1975). However, what is left is action through speech more narrowly defined: namely, utterances can be part *or* parcel of an intervention in the sense of psychological science: they can introduce changes to the situation to affect outcomes. In this case, they ought to affect goal attainment. If these utterances are not part of a structured intervention, they may be thought of as psychological nano-interventions: e.g., setting tasks for the coachee or otherwise attempting to cause change in the interlocutor.

As coaching dialogue is supposed to help coachees in their pursuits, this pursuit needs to be affected by it. Besides activation of preexisting resources through questioning, offering of novel insights by answering questions, and supporting reflection by paraphrasing, coaching dialogues can also directly intervene to promote goal-directed behaviour or a change in behaviour that is beneficial for the attainment of a goal. These interventions fall into two groups: one is structured, and follows an empirically validated recipe, such as mental contrasting with implementation intents (MCII) (Duckworth et al., 2011), where imagery of goal states and reality is contrasted before concretely defining and committing to the resolution of this discrepancy in terms of if-then clauses. The other pathway is to rely on a suite of behavioural change techniques to be applied according to a perceived need expressed in the conversation. This pathway is to be called one of nano-intervention to underline that it takes place in potentially single utterances, in contrast to structured conversations, which rely on a predefined conversation flow.
To guide the design of coaching conversations, a blueprint model developed for a health coaching chatbot can be adapted (Beinema et al., 2022). This model describes a hierarchical way of thinking about bot utterances to be generated or retrieved. At a highest level, the domain is specified. In the current project, this is goal pursuit coaching. At the next level down, there is a choice among discourse topics. For example, goal setting or goal pursuit check-ins can be understood from this lens: meaning courses-of-action. The third highest level is that of actions, e.g., getting acquainted talk. Most actions presented by Beinema et al., 2022 are behavioural change techniques taken from the taxonomy of (Michie et al., 2013).
In summary, dialogue is structured in a way to raise and fulfill expectations. In coaching dialogues, what is to be achieved is defined in terms of the coachees aspirations. Within goal pursuit coaching, this specifically pertains to behaviours promoting the attainment of a chosen goal.

3.4 Current Trends in Conversational AI and Natural Language Processing

A common industry standard in conversational AI is a focus on intent, i.e., what is to be achieved with the conversation (Williams, 2018). Here, user intents are in focus. Machine learning can be used to classify any user input to a predefined set of intents, which in turn are linked to responses, which are often predefined as well. This way of driving conversation has had great success; however, its limitations have led to attempts at breaking the corset of this dialogue management approach. One way to go about it is to skip the intermediary step of intent classification (Rasa, 2021). Besides intent classification, the second part of the by-now traditional formula, namely response selection, may be partly replaced as well. Advances in natural language processing, namely transformer models (Vaswani et al., 2017), have enabled the use of text generation. Very large generative models, such as GPT-3 (Brown et al., 2020) are particularly successful at generating well-conditioned text. Soon after their wide-spread introduction, Large Language Models (LLMs) were fine-tuned to generate conversational responses (Y. Zhang et al., 2019). This approach has been variously developed, including generation based on more representative data (Roller et al., 2020; Thoppilan et al., 2022) or models improving with use (Shuster et al., 2022).
A possibility of combining the mature technology of intent-based dialogue systems with the emerging technology of LLMs is to use an open domain chatbot as a fallback option where no intents can be classified with high enough confidence (Young et al., 2022).
The usefulness of LLMs in conversational AI is not confined to their use as response generators. Intent-based systems rely on data that is traditionally hand-written. The capability of LLMs to extrapolate a pattern of text from a few examples can help to populate data structures used for classification and selection as well. An interesting development in this regard is the chaining of LLM generations (Wu, Terry, and Cai, 2022). Chatbots could be used in the educational context such as for providing formative feedback (Rietsche, Duss, et al., 2018), in review generation (Rietsche, Frei, et al., 2019) or peer-feedback creation (Rietsche and Söllner, 2019).
In summary, incumbent intent classification and response selection is being replaced by text-to-response mapping, especially in generative models fine-tuned or explicitly trained on dialogue data. The mere possibility of an alternative should not be enough to discard proven technology, however. Rather, an

important task for researchers in this area is to identify and contour where which method or combination is superior.

4 Research Plan

Besides coaching in a purely dialogic mode, incorporating questions, answers, and paraphrases, coach utterances may have an explicit psychological intervention aim too. Some interventions are structured, others can be referred to as nano-interventions, and are based on potentially single utterances. Combining this with an eye on informing automated coaching conversations, two questions arise: whether interventions have an effect beyond the provision of a responsive self-reflection opportunity, and whether structured or nano-interventions are more effective.

4.1 Falsification Plan

The research will be conducted with consenting adults who are taking part or have over the course of their lives taken part in a course or program of further education. It will use a 2x2 longitudinal design (see left column in Figure 1 for the 4 groups) and require 140 participants to achieve 80% power to detect the smallest effect of interest, which is defined as a pre-post effect in goal attainment of at least a medium effect ($d = .3$) for the double intervention group. This is based on a meta-analysis for the structured intervention, which reports an average effect of $g = .277$ with documents and .465 with human experimenters (Wang, 2021).

4.2 Hypotheses

The operational hypothesis is that there are group differences in goal attainment. As structured interventions have a better evidence base but adaptive nano-interventions more closely correspond to natural coaching conversations, it is expected that combining both structured and adaptive nano-interventions (Group 4) will lead users to outperform users exposed to any one of them (Groups 2 and 3, respectively), while all three are expected to outperform a chitchat control group (Group 1) in goal attainment:

- H1: The goal attainment of Group 4 will be significantly higher than that of Groups 1-3.
- H2: The goal attainment of Groups 2-4 will be significantly higher than that of Group 1.
- Exploratory: Is goal attainment higher in group 2 or 3?

4.3 Timeline and Independent Variables

The timeline of the experiment is as follows: Participants are screened to have a personally important learning goal they pursue over the course of months. They are then confronted with standardised questions: First on demographics and level of education, then on what their goal is, what their progress is in terms of Goal Attainment Scaling (G. B. Spence, 2007), whether they are intrinsically motivated towards their goal and how helpful they expect the coaching sessions to be. After this, the conversation changes depending on randomized group allocation. The available groups are described below. After two weeks, participants are invited to a second session that starts with a goal pursuit check in, again measuring goal attainment; thereafter group allocation determines the conversation. In week 4, participants are given access to the chatbot but not otherwise incentivised to use it. In week 6, they are invited to a follow-up survey.

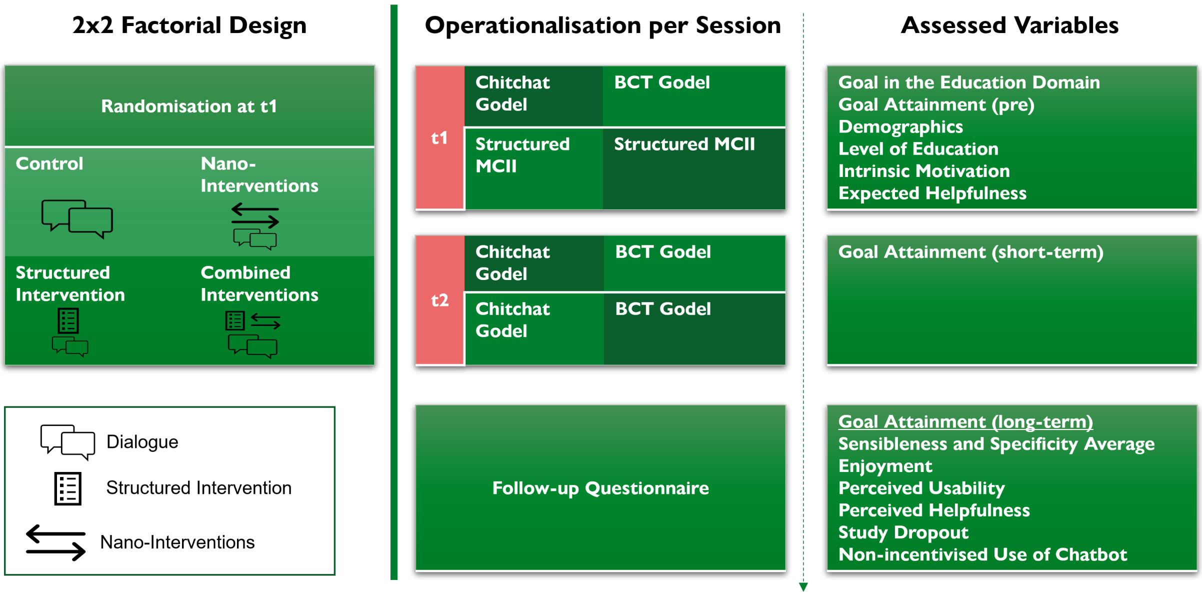

Figure 1. Experimental Design Overview. BCT = Behavioural Change Techniques. MCII = Mental Contrasting with Implementation Intentions. Godel = Grounded Open Dialogue Language Model (Peng et al., 2022). The left column shows the groups of the 2x2 factorial design. The middle column informs on the technological operationalisation per session. The right column gives an overview over collected data per session. The main hypotheses are that a combined intervention leads to better goal attainment than the control group and each separate intervention experiences. The minimal hypothesis is that at least one non-control group experiences improved goal attainment. The main exploratory focus lies on comparing nano-interventions and structured intervention groups.

4.3.1 Group 1 (Control): Over-the-counter Open Domain LLM

After the initial dialogue assessing the variables in 1, the Grounded Open Dialogue Language Model (Godel) (Peng et al., 2022) is added in chitchat mode as a way to keep participants engaged for the duration of about 8 minutes in the first and second sessions. Unincentivised use opportunity and follow-up survey follow as described.

4.3.2 Group 2 (Nano-Interventions): Behavioural Change Techniques Godel

Group 2 differs from group 1 in that chitchat mode is exchanged for nano-interventions pertaining to the attainment of their goal. Such responses may take the form "Can you tell me more about what facilitators you've found helpful in pursuing your goal?". To test the effect of nano-interventions, selected behavioural change technique titles from Michie et al. (2013) are used in a prompt of the form "Given a topic, generate evidence-based knowledge on that topic.\nTopic: TITLE\nKnowledge:" and put into the GPT-3 model davinci-003 (Brown et al., 2020). This produces contextual knowledge for a grounded response, which is triggered as a fallback if none of the intents is detected. Besides the contextual knowledge, the sessions' complete dialogue history and an instruction of the form "Instruction: Given dialogue history and relevant knowledge, DESCRIPTION" is used with the description of the behavioural change technique. This is meant to engage the participants for about 8 minutes. The second session functions equivalent.

4.3.3 Group 3 (Structured Intervention): Mental Contrasting With Implementation Intentions

Group 2 differs from group 1 in that participants are exposed to an example for a structured intervention with good evidence, namely to MCII (Duckworth et al., 2011). For this purpose, bot utterances are defined using the procedure outlined in (Fritzsche et al., 2016). They are retrieved using story-blocks [1] if users fulfill the task by providing input that is classified with intents such as "answer question on positive aspects of goal attainment". Initial intent expression examples are derived from GPTJ Wang, 2021 using the prompt "Given an intent, generate utterances conveying this intent\nIntent:". Further intent expression examples are generated using GPTJ with the prompt "Given an intent, generate utterances with this intent.\nIntent: affirm Utterances:\n- EXAMPLE 1\n- EXAMPLE 2\n- EXAMPLE 3". Of all expressions, an additional 4 paraphrases each are generated using pegasus Y. Zhang et al., 2019. In the second session, the procedure for Group 1 is used.

4.3.4 Group 4 (Double Intervention): Both nano-interventions and MCII

For this group, all data generated for groups 2 and 3 are used in combination, which means that participants ought to receive a structured intervention in session 1 and nano-interventions in session 2.

4.4 Dependent Variables

The primary variable of interest is goal attainment in a follow-up. As secondary variables, the Sensibleness and Specificity Average (Adiwardana et al., 2020), enjoyment, perceived usability, perceived helpfulness, study dropout, and voluntary use of the chatbot after follow-up are used. Study dropout being predicted to be negatively, and all others positively affected by belonging to groups 2 through 4, with the largest effect predicted for group 4.

5 Conclusion

Adult learning is a growing field with a relative dearth of professional helping relationships built into established institutions. As conversational AI is maturing, it may capture some of this growth, which adds demand to the market for skilled conversation. Coaching chatbots can supply this demand. On the level of coaching conversation, interventions aiming at improving coachees' goal attainment may be structured or adaptive nano-interventions. On the level of enabling technology, these approaches may be transported by story-based response selection or prompting-derived generative methods. The proposed experimental design strives to answer which approach is preferable for use in goal pursuit coaching for adult learners. Its results may inform the more general questions of how dialogue is best able to transport change techniques, and which technological foundation or combination thereof is to be further developed to enable the automation of skilled conversation.

References

Adiwardana, Daniel, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le (2020). *Towards a Human-like Open-Domain Chatbot*. DOI: 10.48550/arXiv.2001.09977.

Armstrong, Hilary (2012). "Coaching as dialogue: Creating spaces for (mis) understandings." *International Journal of Evidence Based Coaching and Mentoring* 10 (1), 33–47.

Ates, Haydar and Kadir Alsal (2012). "The Importance of Lifelong Learning has been Increasing." *Procedia - Social and Behavioral Sciences* 46, 4092–4096. ISSN: 18770428. DOI: 10.1016/j.sbspro.2012.06.205.

[1] For the retrieval mechanism, see rasa.com; see github.com/*REDACTED*/botgroup3/data/stories.yml for story blocks, including blocks redirecting to recover the course-of-action

Austin, J. L., J. O. Urmson, and Marina Sbisà (1975). *How to do things with words*. 2nd ed. / edited by J.O. Urmson and Marina Sbisà. Vol. 1955. William James lectures. Oxford: Clarendon Press. ISBN: 9780198245537.

Aymerich-Franch, Laura and Iliana Ferrer (2022). "Liaison, safeguard, and well-being: Analyzing the role of social robots during the COVID-19 pandemic." *Technology in society* 70, 101993. ISSN: 0160-791X. DOI: 10.1016/j.techsoc.2022.101993.

Bachkirova, Tatiana (2011). *Developmental coaching: Working with the Self*. Maidenhead: Open University Press. ISBN: 9780335238569.

Bachkirova, Tatiana, Gordon Spence, and David B. Drake, eds. (2017). *The SAGE handbook of coaching*. SAGE reference. Los Angles: SAGE reference. ISBN: 978-1-4739-1653-1.

Beinema, Tessa, Harm op den Akker, Hermie J. Hermens, and Lex van Velsen (2022). "What to Discuss?—A Blueprint Topic Model for Health Coaching Dialogues With Conversational Agents." *International Journal of Human–Computer Interaction*, 1–19. ISSN: 1044-7318. DOI: 10.1080/10447318.2022.2041884.

Békés, Vera and Katie Aafjes-van Doorn (2020). "Psychotherapists' attitudes toward online therapy during the COVID-19 pandemic." *Journal of Psychotherapy Integration* 30 (2), 238–247. ISSN: 1053-0479. DOI: 10.1037/int0000214.

Blackman, A., & Carter, A. (2014). *Initial Findings from the Coaching for Effectiveness Survey*.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei (2020). *Language Models are Few-Shot Learners*. URL: https://arxiv.org/pdf/2005.14165.

Cox, Elaine (2015). "Coaching and Adult Learning: Theory and Practice." In: *Transforming adults through coaching*. Ed. by James P. Pappas, Jerry Jerman, Susan Imel, and Jovita M. Ross-Gordon. New Directions for Adult and Continuing Education. San Francisco, California: Jossey-Bass, pp. 27–38. ISBN: 978-1119215080.

Crowe, Trevor (2017). "Coaching and Psychotherapy." In: *The SAGE handbook of coaching*. Ed. by Tatiana Bachkirova, Gordon Spence, and David B. Drake. SAGE reference. Los Angles: SAGE reference. ISBN: 978-1-4739-1653-1.

Cushion, Christopher J. and Robert C. Townsend (2019). "Technology-enhanced learning in coaching: a review of literature." *Educational Review* 71 (5), 631–649. ISSN: 0013-1911. DOI: 10.1080/00131911.2018.1457010.

Delnoij, Laurie E.C., Kim J.H. Dirkx, José P.W. Janssen, and Rob L. Martens (2020). "Predicting and resolving non-completion in higher (online) education – A literature review." *Educational Research Review* 29, 100313. ISSN: 1747938X. DOI: 10.1016/j.edurev.2020.100313.

Dubrowski, Adam and Helen MacRae (2006). "Randomised, controlled study investigating the optimal instructor: student ratios for teaching suturing skills." *Medical education* 40 (1), 59–63. ISSN: 0308-0110. DOI: 10.1111/j.1365-2929.2005.02347.x.

Duckworth, Angela Lee, Heidi Grant, Benjamin Loew, Gabriele Oettingen, and Peter M. Gollwitzer (2011). "Self–regulation strategies improve self–discipline in adolescents: benefits of mental contrasting and implementation intentions." *Educational Psychology* 31 (1), 17–26. ISSN: 0144-3410. DOI: 10.1080/01443410.2010.506003.

Duff, Carlton T. and Robinder P. Bedi (2010). "Counsellor behaviours that predict therapeutic alliance: From the client's perspective." *Counselling Psychology Quarterly* 23 (1), 91–110. ISSN: 0951-5070. DOI: 10.1080/09515071003688165.

Fritzsche, Anja, Björn Schlier, Gabriele Oettingen, and Tania M. Lincoln (2016). "Mental Contrasting with Implementation Intentions Increases Goal-Attainment in Individuals with Mild to Moderate

Depression." *Cognitive Therapy and Research* 40 (4), 557–564. ISSN: 0147-5916. DOI: 10.1007/s10608-015-9749-6.

Gao, Jianfeng, Michel Galley, and Lihong Li (2018). "Neural Approaches to Conversational AI." In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Ed. by Kevyn Collins-Thompson, Qiaozhu Mei, Brian Davison, Yiqun Liu, and Emine Yilmaz. New York, NY, USA: ACM, pp. 1371–1374. ISBN: 9781450356572. DOI: 10.1145/3209978.3210183.

Gelo, Omar C.G., Alfred Pritz, and Bernd Rieken, eds. (2015). *Psychotherapy Research*. Vienna: Springer Vienna. ISBN: 978-3-7091-1381-3. DOI: 10.1007/978-3-7091-1382-0.

Grant, Anthony M. (2017). "Coaching as Evidence-Based Practice: The View Through a Multiple-Perspective Model of Coaching Research." In: *The SAGE handbook of coaching*. Ed. by Tatiana Bachkirova, Gordon Spence, and David B. Drake. SAGE reference. Los Angles: SAGE reference. ISBN: 978-1-4739-1653-1.

Gray, David E. (2017). "Towards a Systemic Model of Coaching Supervision." In: *The SAGE handbook of coaching*. Ed. by Tatiana Bachkirova, Gordon Spence, and David B. Drake. SAGE reference. Los Angles: SAGE reference. ISBN: 978-1-4739-1653-1.

Harackiewicz, Judith M., Kenneth E. Barron, John M. Tauer, Suzanne M. Carter, and Andrew J. Elliot (2000). "Short-term and long-term consequences of achievement goals: Predicting interest and performance over time." *Journal of Educational Psychology* 92 (2), 316–330. ISSN: 0022-0663. DOI: 10.1037/0022-0663.92.2.316.

Harkin, Benjamin, Thomas L. Webb, Betty P. I. Chang, Andrew Prestwich, Mark Conner, Ian Kellar, Yael Benn, and Paschal Sheeran (2016). "Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence." *Psychological bulletin* 142 (2), 198–229. DOI: 10.1037/bul0000025.

Hastings, Lindsay J. and Cindy Kane (2018). "Distinguishing Mentoring, Coaching, and Advising for Leadership Development." *New directions for student leadership* 2018 (158), 9–22. DOI: 10.1002/yd.20284.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989). "Multilayer feedforward networks are universal approximators." *Neural Networks* 2 (5), 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8. URL: https://www.sciencedirect.com/science/article/pii/0893608089900208.

John L. Bennett and Francine Campon (2017). "Coaching and Theories of Learning." In: *The SAGE handbook of coaching*. Ed. by Tatiana Bachkirova, Gordon Spence, and David B. Drake. SAGE reference. Los Angles: SAGE reference. ISBN: 978-1-4739-1653-1.

Kendrick, Kobin H., Penelope Brown, Mark Dingemanse, Simeon Floyd, Sonja Gipper, Kaoru Hayano, Elliott Hoey, Gertie Hoymann, Elizabeth Manrique, Giovanni Rossi, and Stephen C. Levinson (2020). "Sequence organization: A universal infrastructure for social action." *Journal of Pragmatics* 168, 119–138. ISSN: 03782166. DOI: 10.1016/j.pragma.2020.06.009.

Knowles, Malcolm S. (1984). *Andragogy in action*. The Jossey-Bass higher education series. San Francisco and London: Jossey-Bass. ISBN: 0875896219.

Latham, Gary P. and Sydney B. Kinne (1974). "Improving job performance through training in goal setting." *Journal of Applied Psychology* 59 (2), 187–191. ISSN: 0021-9010. DOI: 10.1037/h0036530.

Latham, Gary P. and Lise M. Saari (1979). "Importance of supportive relationships in goal setting." *Journal of Applied Psychology* 64 (2), 151–156. ISSN: 0021-9010. DOI: 10.1037/0021-9010.64.2.151.

Lindsay, Neil, Tom Baker, and Octavia Calder-Dawe (2022). "Mental coaching through crisis: digital technologies and psychological governance during COVID-19." *Critical Public Health* 32 (1), 104–115. ISSN: 0958-1596. DOI: 10.1080/09581596.2021.1965543.

Lömker, Malte, Ulrike Weber, and Johannes Moskaliuk (2021). *Chatbots im Coaching: Chancen im lösungs-fokussierten Coaching*. 1st ed. essentials. Wiesbaden: Springer Fachmedien. ISBN: 9783658328306. DOI: 10.1007/978-3-658-32830-6.

Lundberg, Carol A., David D. McIntire, and Caroline T. Creasman (2008). “Sources of Social Support and Self-Efficacy for Adult Students.” *Journal of College Counseling* 11 (1), 58–72. ISSN: 10990399. DOI: 10.1002/j.2161-1882.2008.tb00024.x.

Maatuk, Abdelsalam M., Ebitisam K. Elberkawi, Shadi Aljawarneh, Hasan Rashaideh, and Hadeel Alharbi (2022). “The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructors.” *Journal of computing in higher education* 34 (1), 21–38. ISSN: 1042-1726. DOI: 10.1007/s12528-021-09274-2.

McKelley, Ryan A. and Aaron B. Rochlen (2007). “The practice of coaching: Exploring alternatives to therapy for counseling-resistant men.” *Psychology of Men & Masculinity* 8 (1), 53–65. ISSN: 1524-9220. DOI: 10.1037/1524-9220.8.1.53.

Michie, Susan, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P. Eccles, James Cane, and Caroline E. Wood (2013). “The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions.” *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine* 46 (1), 81–95. DOI: 10.1007/s12160-013-9486-6.

Moskowitz, Gordon B. and Heidi Grant (2009). *The psychology of goals*. The psychology of goals. New York, NY, US: Guilford Press. ISBN: 978-1-60623-029-9.

Mukhopadhyay, Pranab, P. K. Sudarsan, and M. P. Tapaswi (2020). “Optimal Ph D student-teacher ratio for Indian research universities: a production function approach.”

Neenan, Michael (2008). “From Cognitive Behaviour Therapy (CBT) to Cognitive Behaviour Coaching (CBC).” *Journal of Rational-Emotive & Cognitive-Behavior Therapy* 26 (1), 3–15. ISSN: 0894-9085. DOI: 10.1007/s10942-007-0073-2.

OECD (2022). *Education at a Glance 2022*. DOI: 10.1787/3197152b-en. URL: https://www.oecd-ilibrary.org/content/publication/3197152b-en.

Pande, Charuta, Hans-Georg Fill, and Knut Hinkelmann (2022). “A Computational Literature Analysis of Conversational AI Research with a Focus on the Coaching Domain.” In: *EPiC Series in Computing*. EasyChair. DOI: 10.29007/lh9r.

Passmore, Jonathan and Yi–Ling Lai (2020). “Coaching Psychology.” In: *Coaching Researched*. Ed. by Jonathan Passmore and David Tee. Wiley, pp. 3–22. ISBN: 9781119656913. DOI: 10.1002/9781119656913.ch1.

Peng, Baolin, Michel Galley, Pengcheng He, Chris Brockett, Lars Liden, Elnaz Nouri, Zhou Yu, Bill Dolan, and Jianfeng Gao (2022). “GODEL: Large-Scale Pre-Training for Goal-Directed Dialog.” DOI: 10.48550/arXiv.2206.11309.

Rasa (2021). *We’re a step closer to getting rid of intents*. URL: https://rasa.com/blog/were-a-step-closer-to-getting-rid-of-intents/.

Rauen, C. (2020). *RAUEN Coaching-Marktanalyse 2020*. rauen.de. URL: https://www.rauen.de/cma/.

Rietsche, Roman, Kevin Duss, Jan Martin Persch, and Matthias Soellner (2018). “Design and Evaluation of an IT-based Formative Feedback Tool to Foster Student Performance.” In: *Proceedings of the International Conference on Information Systems (ICIS)*. San Francisco, CA, USA.

Rietsche, Roman, Daniel Frei, Emanuel Stoeckli, and Matthias Soellner (2019). “Not all Reviews are Equal - a Literature Review on Online Review Helpfulness.” In: Stockholm, Sweden: Proceedings of the European Conference on Information Systems (ECIS).

Rietsche, Roman and Matthias Söllner (2019). “Insights into Using IT-Based Peer Feedback to Practice the Students Providing Feedback Skill.” In: Maui, HI, USA: Proceedings of the Hawaii International Conference on System Sciences (HICSS).

Rogers, Carl R. (1951). *Client-centered therapy: Its current practice, implications and theory*. S.l.: Constable. ISBN: 978-0094539907.

Roller, Stephen, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, and Jason Weston (2020). *Recipes for building an open-domain chatbot*. DOI: 10.48550/arXiv.2004.13637.

Sackett, D. L., W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson (1996). "Evidence based medicine: what it is and what it isn't." *BMJ (Clinical research ed.)* 312 (7023), 71–72. ISSN: 0959-8138. DOI: 10.1136/bmj.312.7023.71.

Schegloff, Emanuel A. (2007). *Sequence organization in interaction: A Primer in Conversation Analysis I*. A primer in conversation analysis : v. 1. Cambridge: Cambridge University Press. ISBN: 9780521532792.

Shuster, Kurt, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, Morteza Behrooz, William Ngan, Spencer Poff, Naman Goyal, Arthur Szlam, Y-Lan Boureau, Melanie Kambadur, and Jason Weston (2022). *BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage*. DOI: 10.48550/arXiv.2208.03188.

Spence, G. B. (2007). *GAS powered coaching: Goal Attainment Scaling and its use in coaching research and practice*. URL: https://www.researchgate.net/profile/jonathan-passmore/publication/47528957_the_state_of_executive_coaching_research_what_does_the_current_literature_tell_us_and_what's_next_for_coaching_research/links/5730d1e108ae08415e6a7fa5/the-state-of-executive-coaching-research-what-does-the-current-literature-tell-us-and-whats-next-for-coaching-research.pdf#page=43.

Terblanche, Nicky, Joanna Molyn, Erik de Haan, and Viktor O. Nilsson (2022). "Comparing artificial intelligence and human coaching goal attainment efficacy." *PloS one* 17 (6), e0270255. DOI: 10.1371/journal.pone.0270255.

Thoppilan, Romal, Daniel de Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le (2022). *LaMDA: Language Models for Dialog Applications*. DOI: 10.48550/arXiv.2201.08239.

Tudini, Vincenza (2015). "Extending Prior Posts in Dyadic Online Text Chat." *Discourse Processes* 52 (8), 642–669. ISSN: 0163-853X. DOI: 10.1080/0163853X.2014.969138.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, \L ukasz Kaiser, and Illia Polosukhin (2017). "Attention is All you Need." In: *Advances in Neural Information Processing Systems*. Ed. by Guyon I., Von Luxburg U., Bengio S., Wallach H., Fergus R., Vishwanathan S., and Garnett R. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Wang, Ben (2021). *Mesh-Transformer-JAX: Model-Parallel Implementation of Transformer Language Model with JAX*.

Weger, Harry, Gina Castle Bell, Elizabeth M. Minei, and Melissa C. Robinson (2014). "The Relative Effectiveness of Active Listening in Initial Interactions." *International Journal of Listening* 28 (1), 13–31. ISSN: 1090-4018. DOI: 10.1080/10904018.2013.813234.

Weizenbaum, Joseph (1966). "ELIZA—a computer program for the study of natural language communication between man and machine." *Communications of the ACM* 9 (1), 36–45. ISSN: 0001-0782. DOI: 10.1145/365153.365168.

Western, Simon (2017). "The Key Discourses of Coaching." In: *The SAGE handbook of coaching*. Ed. by Tatiana Bachkirova, Gordon Spence, and David B. Drake. SAGE reference. Los Angles: SAGE reference, pp. 42–61. ISBN: 978-1-4739-1653-1.

Williams, Sam (2018). *Hands-on chatbot development with Alexa skills and Amazon Lex: Create custom conversational and voice interfaces for your Amazon Echo devices and web platforms*. 1st edition. Birmingham, UK: Packt. ISBN: 9781788993487.

Winkler, Rainer and Matthias Soellner (2018). "Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis." *Academy of Management Proceedings* 2018 (1), 15903. ISSN: 0065-0668. DOI: 10.5465/AMBPP.2018.15903abstract.

Wollny, Sebastian, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler (2021). "Are We There Yet? - A Systematic Literature Review on Chatbots in Education." *Frontiers in artificial intelligence* 4, 654924. DOI: 10.3389/frai.2021.654924.

Wu, Tongshuang, Michael Terry, and Carrie Jun Cai (2022). "AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts." In: *CHI Conference on Human Factors in Computing Systems*. Ed. by Simone Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Drucker, Julie Williamson, and Koji Yatani. New York, NY, USA: ACM, pp. 1–22. ISBN: 9781450391573. DOI: 10.1145/3491102.3517582.

Young, Tom, Frank Xing, Vlad Pandelea, Jinjie Ni, and Erik Cambria (2022). "Fusing Task-Oriented and Open-Domain Dialogues in Conversational Agents." *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (10), 11622–11629. ISSN: 2159-5399. DOI: 10.1609/aaai.v36i10.21416.

Zhang, Jingwen, Yoo Jung Oh, Patrick Lange, Zhou Yu, and Yoshimi Fukuoka (2020). "Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint." *Journal of medical Internet research* 22 (9), e22845. DOI: 10.2196/22845.

Zhang, Yizhe, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan (2019). *DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation*. DOI: 10.48550/arXiv.1911.00536.

Zorrilla, Asier López and M. Inés Torres (2022). "A Multilingual Neural Coaching Model with Enhanced Long-term Dialogue Structure." *ACM Transactions on Interactive Intelligent Systems* 12 (2), 1–47. ISSN: 2160-6455. DOI: 10.1145/3487066.