

Toward Democratic Laboratory: LLM-based Coaching Agent for Promoting Students' Autonomy

Ekai Hashimoto^{1,*†}, Shun Shiramatsu^{1,†}

¹Nagoya Institute Of Technology

Abstract

The trend towards democratic, flat management structures is increasing, yet active participation often falls short. This study aims to build a prototype coaching system to boost individual autonomy, thus facilitating democratic management within a university research lab. Despite our lab's intent to function democratically, low participation levels have hindered democratization. Our research objective is to foster engagement via a coaching system that leverages self-efficacy and feedback mechanisms with ChatGPT and GPT-4. Experimental results showed improved self-efficacy and positive feedback on usability, although the system struggled with session closure and lacked natural progression in the coaching process.

Keywords

Large-scale Language Model, Organization Management, Coaching Agent, Self-efficacy, Well-being,

1. Introduction

In recent years, there has been an increasing movement from traditional top-down organizational management toward flat, democratic organizational management in which members operate more autonomously. Examples include teal organizations and holacracy-type organizations. Such flat organizational management is expected to increase subjective well-being because it emphasizes member autonomy[1]. Although members' active engagement is required for democratic organizational management, it tends to be insufficient[2].

Therefore, this study aims to develop a prototype coaching system to enhance individual autonomy and encourage active participation in democratic organizational management. The coaching system is implemented by providing specific prompts to a large-scale language model, ChatGPT¹. For this purpose, we need to consider the following requirements.

First, in flat organizational management, autonomy should be more important. People with high levels of autonomy typically have high levels of self-efficacy. A high level of self-efficacy improves the sense of control over one's life, which is expected to improve well-being[3]. Developing such coaching system has been challenging due to the need to accommodate various

DemocrAI 2023: The 2nd International Workshop on Democracy and AI, August 20, 2023, Macao

*Corresponding author.

†These authors contributed equally.

✉ e.hashimoto.611@stn.nitech.ac.jp (E. Hashimoto); siramatu@nitech.ac.jp (S. Shiramatsu)

🌐 <http://www.srmt.nitech.ac.jp/> (S. Shiramatsu)

🆔 0000-0002-6357-3207 (S. Shiramatsu)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

📄 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://openai.com/blog/chatgpt>

utterances. However, with the advent of GPT-4[4] in 2022 , the possibility of implementing such a system that meets such requirements has become possible.

In particular, this paper targets democratic organizational management for university laboratories. Our laboratory has been aiming at an organization in which the management of the laboratory is student-driven and democratic. However, it has been difficult to promote democratization due to the low level of willingness of laboratory members to participate in the management of the laboratory. Even if members showed a willingness to participate, they often found it challenging to generate concrete rule proposals for democratic laboratory management. Further, when they did come up with a proposal, articulating it appropriately posed another challenge. These problems can be solved with the support of verbalization by the coaching system, and democratic management of the organization becomes possible.

2. Previous Studies

Coaching is recognized in both theory and practice of Human Resource Development as an important tool to promote both individual personal growth and improved organizational outcomes[5]. Taking this background into account, exploring how AI can contribute to the coaching process is an important research question. One important concern about AI coaching is that it cannot provide a work alliance for the user[6]. Work alliance is one of the central elements for predicting coaching outcomes[7].

In this study, we expected to provide appropriate feedback in coaching to build trust and work alliance with the user.

3. Method

3.1. Experimental Design

You are a professional coach. Please fully understand the following 19 tips.
1.Sharing clear goals: Users and coaches need to clearly share the purpose of coaching and expected outcomes before beginning coaching.
19.Utilizing tips: You should be conscious of using the right tips at the right time to enhance the quality of coaching.
Thinking the above 19 tips, please coach me (the user).
Please output as follows:
Feedback: (greeting only initially) Question:
Information about the user: Reason for the question:
Coaching tip: (Please indicate which tip you used. No.1)

Figure 1: Initial Prompt
(translated from Japanese)

Until now, the order of presentations in the seminar was determined by the professor, but we were told to decide it among the students from now on.
Feedback: This represents a new responsibility and an excellent opportunity to polish new skills. I understand that you are facing this new situation.
Question: How do you feel about this new change? Also, what do you think are the main challenges you will face when deciding the order of presentations among students?
User information: The user is a student and is facing a new situation where the order of presentations in the seminar is decided among the students.
Reason for the question: This will help me to deeply understand the problem you are facing and to co-construct strategies to solve that problem. Also, by understanding your feelings and challenges, I can support you more effectively.
Coaching tips: No.4, No.5, No.12
I don't feel it's a major change, but I think we need to think about how to decide the order of presentations.
Feedback: I can understand that perspective. The process of deciding the order of presentations is important for ensuring team harmony and fairness.
Question: What is the ideal presentation order decision process that you imagine? Also, what do you think is necessary to realize that process?
User information: The user feels the need to consider the process of deciding the order of presentations.
Reason for the question: Through this question, we can identify the ideal outcome you want and the actions or resources needed to achieve it.
Coaching tips: No.7, No.13, No.11

Figure 2: Dialogue log
(translated from Japanese)

The proposed coaching system is realized by giving specific prompts to ChatGPT. In this research, we utilized GPT-4, which possesses superior natural language processing capabilities compared to GPT-3. The version of GPT-4 used was as of May 12 and 24, 2023.

The initial prompt is shown in Figure 1. It is known that the performance of ChatGPT can be improved by following a sequential process[8]. In the proposed method, explicitly stating the reason of asking a question is aiming for a similar effect.

In the prompt of the proposed method, we assigned the role of a professional coach to ChatGPT, made it understand the enumerated coaching tips, and then conducted coaching. Also, to ensure consistent output, specific settings were applied to the output parameters.

3.2. Evaluation Method

To achieve this aim, we selected eight students from our own laboratory as subjects. This selection was made in an attempt to address issues such as a lack of willingness to participate present in our lab. All questionnaires were rated on a seven-point scale.

Table 1

Questioner for Self-efficacy (translated from Japanese)

No.	Item Content	Category
1.	I think I know the background to the topic	Knowledge
4.	I think I know the methods of the topic	Knowledge
6.	I think I understand the necessity and impact of my topic	Knowledge
2.	I think I can propose a better topic than others	Ability
5.	I think my ability to consider and propose topics is outstanding compared to others	Ability
3.	I think I can contribute when discussing a topic	Confidence
7.	I think I can propose a good new plan when the topic should be changed	Confidence
8.	I think I can propose appropriately in a discussion about the topic	Confidence

As shown in Table 1, we utilized a translated version of eight out of the nine items from the Motivated Strategies for Learning Questionnaire [9]. This questionnaire uses a scale for measuring self-efficacy. We further classified these eight items into three categories related to knowledge, ability, and confidence-based self-efficacy.

We hypothesized that providing feedback would facilitate the enhancement of users' self-efficacy. For the controlled experiment, we prepared two variants of the system: one featuring a feedback function (proposed method), and the other without it (baseline method). Subjects used each system for two designated topics. The topics were to consider the order of presentations in the laboratory and to think about rule-making in the laboratory. Through these procedures, we investigated how the use of the system affects self-efficacy.

4. Results

4.1. The User's Self-efficacy

The difference in self-efficacy before and after the experiment is shown in Table 2. Both the proposed method and the baseline method improved overall self-efficacy. However, the baseline method led to a greater increase in user's self-efficacy compared to the proposed method. This difference was notably more significant in terms of ability than anything else.

Table 2

Improvements Points of Self-efficacy

Method	Knowledge	Ability	Confidence
Proposed	+1.9pt (p<0.01)	+1.9pt (p<0.01)	+1.1pt (p<0.025)
Base	+1.4pt (p<0.025)	+2.3pt (p<0.01)	+1.5pt (p<0.01)

4.2. Coaching Log

Figure 2 shows part of dialogue. On average, one session took 34.5 minutes. While this is average for general coaching, in free answers it was found that users felt it was very long. This is thought to be because similar questions were repeated many times. However, such questions tend to arise when the user gives evasive responses.

The system tends to view things in a comprehensive and leak-free way. The system asks questions to get specifics, and often pursues parts that lack concreteness in the user's responses. While it doesn't ask exactly the same question, it often asks about the same topic or detailed events within the topic.

In human conversation, there may be parts that end vaguely. However, the system asks more than usual about the same matter in order to draw out all the specific parts. The user did not refuse to answer even when the system was asking about parts that the user did not consider important. It was found from the free answer questionnaire that the users became anxious as to whether the coaching was going off track as similar questions were repeated.

The system usually responds favorably to the user's answers and proposals. The system checks whether the user's response is logically correct. If there is a problem even if it is correct, it points it out with reasons.

5. Discussion

5.1. Self-efficacy of Users

Our hypothesis, which stated that the system's feedback could maintain users' motivation and enhance their sense of self-efficacy, was not confirmed. However, in the user survey, there was feedback saying, "I didn't feel like I was just mechanically typing characters because there was feedback." Furthermore, there were no negative opinions regarding the feedback feature itself.

Both the proposed method and the base method improved self-efficacy after using the system. From this, using this system encourages users to think more deeply about the topic and improves their self-efficacy.

5.2. Outliers

One participant rated the proposed method much lower in terms of self-efficacy of knowledge. Furthermore, this participant experienced a decrease in self-efficacy across all categories when using proposed model.

The survey revealed that, despite having confidence and high evaluation in their own knowledge before use, they downgraded it. Similarly, this user also downgraded their capabilities

compared to before the experiment. From the results, the participant realized that lacking knowledge and abilities through using the system and thought more deeply about the topic.

Table 3
Improvement Points of Self-efficacy without Outliers

Method	Knowledge	Ability	Confidence
Proposed	+2.5pt (p<0.01)	+2.5pt (p<0.01)	+1.5pt (p<0.01)
Base	+1.1pt (p<0.1)	+2.0pt (p<0.1)	+1.3pt (p<0.1)

Table 3 shows the results with the outlier removed. The figures for the proposed method significantly improved. In terms of knowledge, improvement points of proposed method are 1.4pt higher than base method(p<0.1).

The system is expected to be beneficial as a method to enhance self-understanding and self-recognition. Some people discover that their self-understanding is lacking as they dive further into themselves. Others make new discoveries and raise their self-efficacy. Therefore, we want to explore additional evaluation metrics.

5.3. Coaching Logs and Flow

Users often feel frustrated when they are pressed for specifics in areas that are not essential to their answers or thoughts. From the perspective, looking at the dialogue logs, we can see situations where the user abandons things that should be considered. However, in defining large frameworks like the topic of this study, it should be deferred unless there is a significant problem. Therefore, it seems that users' dissatisfaction accumulated as the session progressed.

The system proceeds to a new question once it has shown reasons why the user doesn't need to answer. However, unlike when speaking to a human, the user didn't doubt the query's intent or decline to respond. If the interlocutor is a trusted figure such as a thesis advisor or an experienced coach, the user will answer without hesitation even if asked a question unrelated to the main point.

5.4. About Using the Coaching System

As shown in 5.3, when grasping and setting up a larger framework, adjustments are needed. There's a lot of waste in system coaching to achieve the user's desired goal. Establishing an early knowledge of the procedures that will be done to achieve the user's goal with them is crucial in order to overcome the issues. The system ensure the user's psychological safety by showing where the session is in progress.

6. Conclusion

In this study, we aimed to enhance self-efficacy within organizational members by exploring a coaching method using ChatGPT. Our experiment with the coaching system demonstrated improvements in self-efficacy and received positive feedback from the participants. Despite being a fundamental method, the system proved useful for tackling specific topics. Furthermore, as

shown in 5.2, the utilization of this system deepened user self-understanding and self-cognition, which in turn contributed to the democratization within the organization.

However, we acknowledge that the sample size of eight participants is relatively small and these findings are specific to our laboratory. It would be necessary to measure the system's effectiveness in other settings as well. Moreover, in the current system, users ended up receiving coaching while feeling anxious and dissatisfied. Therefore, a refinement of the coaching method is necessary to minimize such user anxiety. Additionally, we believe it's essential to evaluate the sense of self-efficacy when the system is used continuously [10].

Refining and improving the system to address user concerns is a primary objective of our research. The system had multiple areas requiring enhancement. We plan to make the necessary improvements and verify their significance.

Acknowledgments

This study is partially supported by JST CREST (JPMJCR20D1) and NEDO (JPNP20006). We greatly appreciate to kind advices by Dr. Minoru Mitsui and his colleagues in mitemo Co., Ltd.

References

- [1] J. R. Foley, M. Polanyi, Workplace democracy: Why bother?, *Economic and Industrial Democracy* 27 (2006) 173–191.
- [2] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, *Advances in Behaviour Research and Therapy* 1 (1978) 139–161.
- [3] T.-S. Han, H.-H. Chiang, A. Chang, Employee participation in decision making, psychological ownership and knowledge sharing: Mediating role of organizational commitment in taiwanese high-tech organizations 21 (2010) 2218–2233.
- [4] OpenAI, Gpt-4 technical report, 2023. [arXiv:2303.08774](https://arxiv.org/abs/2303.08774).
- [5] R. S. Beattie, S. Kim, M. S. Hagen, T. M. Egan, A. D. Ellinger, R. G. Hamlin, Managerial coaching: A review of the empirical literature and development of a model to guide future practice, *Advances in Developing Human Resources* 16 (2014) 184–201.
- [6] G. Bozer, M. Delegach, Bringing context to workplace coaching: A theoretical framework based on uncertainty avoidance and regulatory focus, *Human Resource Development Review* 18 (2019) 376–402.
- [7] C. Graßmann, F. Schölmerich, C. C. Schermuly, The relationship between working alliance and client outcomes in coaching: A meta-analysis, *Human Relations* 73 (2020).
- [8] J. Wei, X. Wang, D. Schuurmans, M. Bosma, brian ichter, F. Xia, E. H. Chi, Q. V. Le, D. Zhou, Chain of thought prompting elicits reasoning in large language models, 2022.
- [9] P. R. Pintrich, E. V. d. Groot, Motivational and self-regulated learning components of classroom academic performance., *Journal of Educational Psychology* (1990).
- [10] T. W. Bickmore, R. W. Picard, Establishing and maintaining long-term human-computer relationships, *ACM Trans. Comput.-Hum. Interact.* 12 (2005) 293–327.