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

Abstract

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. People with high levels of autonomy typically have high levels of self-efficacy. A high sense of self-efficacy can enhance the feeling of being in control of one's life, leading to improved well-being. Although members' active engagement is required for democratic organizational management, it tends to be insufficient. Therefore, this study aims to develop a prototype coaching system to enhance individual autonomy and encourage active participation in democratic organizational management. In this paper, we aim for democratic organizational management, using a university research laboratory as the target. Our lab has always aimed to be an organization where students autonomously run the lab democratically. However, it has been difficult to democratize due to the low desire among lab members to participate in lab management. The objective of this research is to develop a coaching system that enhances engagement and aids in facilitating democratic organizational management. In this paper, we focus on user self-efficacy and feedback, and develop a method using ChatGPT and GPT-4. To verify the significance of the developed coaching system, we conducted experiments using the coaching system. Through the coaching system, improvements in self-efficacy were observed and some positive feedback on usability was obtained. However, the coaching system struggled to indicate the end of sessions and failed to establish a natural progression in the coaching process. Future work will focus on enhancing the coaching system and further validating its significance.

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, ChatGPT. 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 sense of self-efficacy improves the sense of control over one's life, which is expected to improve well-being[3].

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

© 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)

In general, when running a democratic organization, the opinions of all members and their backgrounds are not shared throughout the organization. Therefore, if the coaching system can embody individual opinions, it may be possible to realize a medium to recognize and share those opinions with the whole group[4]. Implementing such a coaching system has been challenging in the past. However, with the advent of GPT-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

2.1. Large-scale Language Model

ChatGPT, released by OpenAI in December 2022, has been highly evaluated for its general-purpose performance since its initial release. Subsequently, GPT-3.5, a fine-tuned model based on GPT-3 with Reinforcement Learning from Human Feedback[5], was developed. However, GPT-3.5 has a problem that it generates fake information including errors in proper nouns when asked to provide factual information such as real-world examples. In March 2023, OpenAI released GPT-4[6], which improves the problem and increases reliability. As of May 2023, GPT-4 is capable of highly accurate automatic processing without re-training, simply by in-context learning of natural language work instructions to input prompts. In this paper, we test a prototype coaching system using GPT-4 and confirm its performance.

2.2. AI and Coaching

Coaching is recognized in both theory and practice of Human Resource Development (HRD) as an important tool to promote both individual personal growth and improved organizational outcomes[7]. 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[8]. Work alliance is one of the central elements for predicting coaching outcomes[9]. In this study, we expected to provide appropriate feedback in coaching to build trust and work alliance with the user.

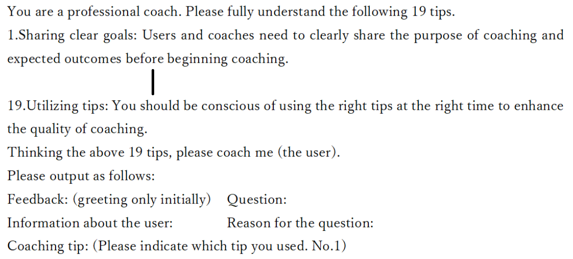
3. Method

3.1. Design of the Proposed Method

The proposed coaching system is realized by giving specific prompts to the large-scale language model 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, 2023.

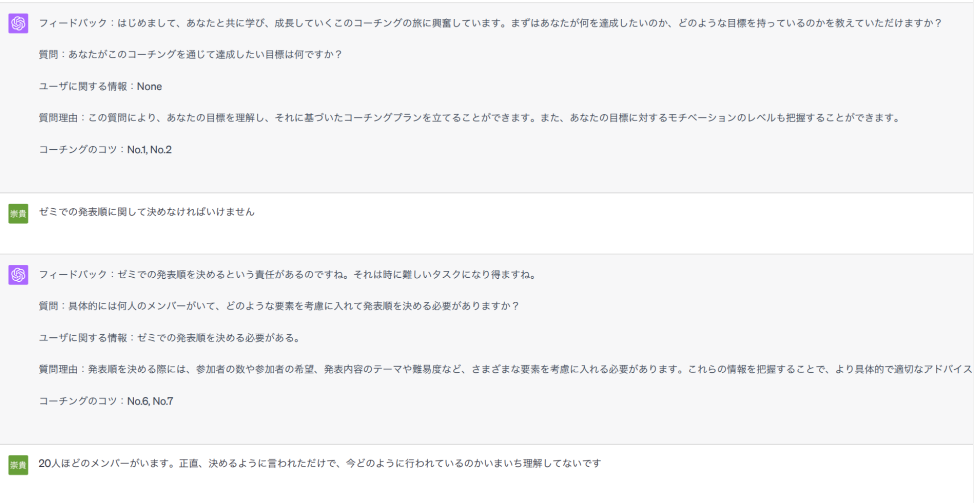
The initial prompt is shown in Figure 1. When using ChatGPT as a system with a specific role, it is advised to make ChatGPT recognize that it is an expert in that role. It is also known that performance can be improved by adopting a sequential process, such as by including the phrase "step by step" [10]. In the proposed method, a similar effect is aimed for by explicitly stating the reason for asking a question.

Therefore, 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.



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)



フィードバック：はじめまして、あなたと共に学び、成長していくこのコーチングの旅に興奮しています。まずはあなたが何を達成したいのか、どのような目標を持っているのかを教えていただけますか？

質問：あなたがこのコーチングを通じて達成したい目標は何ですか？

ユーザに関する情報：None

質問理由：この質問により、あなたの目標を理解し、それに基づいたコーチングプランを立てることができます。また、あなたの目標に対するモチベーションのレベルも把握することができます。

コーチングのコツ：No.1, No.2

ゼミでの発表順に関して決めなければいけません

フィードバック：ゼミでの発表順を決めるという責任があるのですね。それは時に難しいタスクになり得ますね。

質問：具体的には何人のメンバーがいて、どのような要素を考慮に入れて発表順を決める必要がありますか？

ユーザに関する情報：ゼミでの発表順を決める必要がある。

質問理由：発表順を決める際には、参加者の数や参加者の希望、発表内容のテーマや難易度など、さまざまな要素を考慮に入れる必要があります。これらの情報を把握することで、より具体的で適切なアドバイス

コーチングのコツ：No.6, No.7

20人ほどのメンバーがいます。正直、決めるように言われただけで、今どのように行われているのかいまいち理解してないです

Figure 2: Dialogue log

3.1.1. Creating Coaching Tips

To improve performance, we had ChatGPT enumerate coaching tips. Before designing the initial prompt, we categorized the coaching tips into three categories: general coaching tips, tips for building a trusting relationship, and tips for maintaining user engagement in coaching. From these, we eliminated tips with the same meaning and obtained 18 tips. We added a 19th tip:"consciously coaching with these tips in mind". As shown in Figure 1, these 19 tips were input as part of the prompt of the proposed method.

The 19 coaching tips were classified into four categories: tips for questions, tips for feedback, tips for agreement on goals, and tips for building trust relationships. The tips were also partially revised to adjust the output.

3.2. Evaluation Method

The purpose of this study is to investigate how the use of the system influences students' self-efficacy and willingness to participate in organizational operations. To achieve this purpose, we used four students from the laboratory to which the authors belong as subjects.

All questionnaires were rated on a seven-point scale.

As shown in Table 1, we utilized a translated version of eight out of the nine items from the Motivated Strategies for Learning Questionnaire (MSLQ) by PINTRICH and DE GROOT (1990)[11]. 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 having feedback would make it easier to build trust and increase trust in the system. For a controlled experiment, we prepared two types of systems: the proposed method with a feedback function and a baseline method which is the proposed method without the feedback function and tips. 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.

After using the system, we conducted a questionnaire about the performance of each method.

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

Table 2
Questionnaire for Performance (translated from Japanese)

No.	Item Content
1.	Were you convinced by the result of this coaching?
2.	Do you think your understanding of your thoughts and yourself deepened through this coaching?
3.	Do you think you could trust this coach?

4. Results

4.1. The User's Self-efficacy

The difference in self-efficacy before and after the experiment is shown in Table 3. 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 particularly notable in terms of ability.

Table 3
Improvements Points of Self-efficacy and Evaluations of the System

Method	Knowledge	Ability	Confidence	No.1	NO.2	No.3	Average
Proposed	+1.8	+1.3	+1.5	5.0	5.0	4.3	4.8
Base	+1.7	+2.8	+2.2	6.5	5.8	6.0	6.1

4.2. The Evaluation of The System

Table 3 shows the evaluation of the system's performance. The base method is rated higher overall than the proposed method. In particular, the base method has a high level of satisfaction and trust in the coach. The hypothesis that the inclusion of a feedback function could improve trust was not supported by the results.

4.3. Coaching Log

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. It is known that the bond with the system becomes stronger when it is used continuously[12]. Therefore, we want to measure the sense of trust when used continuously.

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. 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, they realized that they lack knowledge and abilities through they used the system and thought more deeply about the topic.

Table 4 shows the results with the outlier removed. The figures for the proposed method significantly improved.

Table 4
Improvement Points

Method	Knowledge	Ability	Confidence
Proposed	+3.0	+2.3	+2.4
Base	+1.7	+2.8	+2.2

Considering the results , 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 a third-party 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

Since the coaching system looks at things from multiple angles and ensures nothing is overlooked, it is helpful when setting up the finer details. For example, it would be very useful in managing laws, rules, and handling exceptions. 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.

Also, from the survey results, we infer that it is suitable for introspection and self-analysis. For example, job seekers and students who are unsure about their career path may use it to find out the career they truly want.

6. Conclusion

In this study, we aimed to improve the self-efficacy of members within an organization and discussed a coaching method using ChatGPT.

In the experiment using the coaching system, it was possible to improve self-efficacy and gain a certain evaluation of the system from the participants. Although it is a base method, it was found that the coaching system is useful when thinking about topics. By receiving coaching, users can analyze their abilities and express their opinions. From this, it was found that this system is useful for democratizing organizations.

However, in this system, users ended up receiving coaching while feeling anxious and dissatisfied. Therefore, we plan to refine the coaching method in the future to minimize user anxiety. The refinement of the system to address user concerns is a key goal. There were also many points that need to be improved in the system, so we would like to improve the corrections and verify the significance.

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, The International Journal of Human Resource Management 21 (2010) 2218–2233.

[4] D. I. Macleod, N. Clarke, Engaging for success: enhancing performance through employee engagement, a report to government, 2011.

[5] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, R. Lowe, Training language models to follow instructions with human feedback, 2022. `arXiv:2203.02155`.

[6] OpenAI, Gpt-4 technical report, 2023. `arXiv:2303.08774`.

[7] 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.

[8] 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.

[9] C. Graßmann, F. Schölmerich, C. C. Schermuly, The relationship between working alliance and client outcomes in coaching: A meta-analysis, Human Relations 73 (2020) 35–58.

[10] 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, in: A. H. Oh, A. Agarwal, D. Belgrave, K. Cho (Eds.), Advances in Neural Information Processing Systems, 2022.

[11] P. R. Pintrich, E. V. d. Groot, Motivational and self-regulated learning components of classroom academic performance., Journal of Educational Psychology (1990).

[12] T. W. Bickmore, R. W. Picard, Establishing and maintaining long-term human-computer relationships, ACM Trans. Comput.-Hum. Interact. 12 (2005) 293–327.