

Case Report

How does the Language Prediction Model, the ChatGPT Evaluate Negative Emotions?

Toshimi Nakanishi^{1*}, Akane Tokuda², Tsukasa Arai³, Yoshihiro Sugiura⁴ and Tatsuya Hayasaka⁵

¹General Medical Education Center, Faculty of Medicine, Yamagata University, Yamagata City, Japan

²Yamagata University Medical School, Japan

³Yamagata University Medical School, Japan

⁴Otaki Hospital, Japan

⁵Department of Anesthesiology, Yamagata University, Yamagata City, Japan

*Corresponding author: Toshimi Nakanishi, General Medical Education Center, Faculty of Medicine, Yamagata University, Yamagata City, Japan

Received: December 05, 2023; Accepted: December 08, 2023; Published: December 12, 2023

Summary

The Generative AI (Artificial Intelligence: AI), the ChatGPT (Generative Pretrained Transformer: ChatGPT), is a language prediction model that generates sentences based on word frequencies and interrelationships. In this study, we evaluated how the ChatGPT, a generative AI, performs in cognitive conflicts (conflicts) between healthcare professionals and patients/families encountered in healthcare settings, using dialogue transcripts of licensed medical mediations (with joint decision making), in which the ChatGPT is said to convey limitations and misinformation regarding negative emotions. We report the results of our study of the ChatGPT's negative emotion evaluation, comparing it with human evaluations.

Abstract

We investigated that how the ChatGPT (Version 3.5), a generative AI, evaluated negative emotions in narratives of cognitive discrepancies (conflicts) between medical professionals and patients/families encountered in the medical field. As a result, negative emotion evaluation by the ChatGPT did not reach the level that people do. It can be inferred that there are limitations to negative emotion evaluation by the ChatGPT at this time.

Introduction

ChatGPT (Generative Pretrained Transformer: ChatGPT) is an artificial intelligence (AI), neural network-based language prediction. It is one of the models that generate sentences based on word frequencies and their interrelationships. This is said to cause so-called hallucination (hallucination), which is the conveyance of incorrect information, due to the limitations of human emotion processing that depends on context and situation [1,2]. Conflicts between medical professionals and patients/families encountered in the medical field are always accompanied by negative emotions. How does the ChatGPT, a generative AI, evaluate such negative emotions? There do not seem to be any evaluation reports on this issue. Therefore, we conducted a human evaluation of the ChatGPT's verbal assessment of negative emotions using recorded dialogue data from past medical mediations (a concept with a dialogue process involving collaborative decision making [3] and investigated the rate of agreement.

Case Presentation

The purpose of the study was to determine whether “The ChatGPT (Version3.5), a generative AI, can capture negative emotions from dialogue narratives.”

The overall flow of the research methodology is shown in Figure 1.

The evaluation period was from August 31, 2023 to September 30,

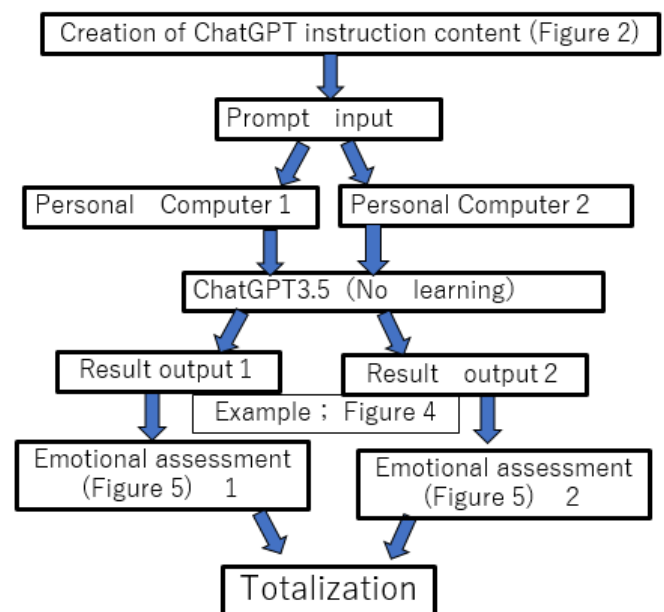


Figure 1: Research Methods.

2023. The subjects were the Ethics Committee and the record language of the medical mediator of the first complaint claim submitted with the permission and consent of the patient's family, among the previously resolved complaint cases, as shown in Table 1.

For the content of the ChatGPT instructions for the generated AI, prompts were created based on the Fukatsu-style questioning technique created by Fukatsu [4], as shown in Figure 2. For the constraints and emotions in the prompts, nine emotions [5] were used, including “joy,” “expectation,” “anger,” “disgust,” “sadness,” “surprise,” “fear,” “trust,” and “anxiety,” referring to Pulchick’s Emotion Circle, and responses were also required to provide reasons for their emotion evaluations. Responses were also asked for reasons for their own emotion ratings. The ChatGPT used for the evaluation was Version 3.5, which was set to not perform the learning function. The prompts were then entered and implemented on two PCs, and the interpersonal emotions of the narratives were

estimated for three participants (an ethics mediator and two male and two female medical students in their 20s each, before they left for clinical practice).

For the evaluation of the generated AI ChatGPT emotion output, examples of output results are shown in Table 2; the output results of the 11 cases were individually evaluated by the three aforementioned prompt implementers using the published rating criteria [5,6] scale. The respective emotional evaluation items and evaluation methods are shown in Figure 3. The two evaluations were then aggregated to conduct an overall evaluation of ChatGPT’s emotion evaluation status [7].

Table 1: Complaint status of subject cases.

Case	Patient narratives of situation content
1	The patient's daughter and the outpatient attending physician talk about the delay in seeing the patient for inappropriate medical care.
2	Patient speaks up to health care provider about his dissatisfaction with the treating technician.
3	The head technician and nurse talk about the policy for dealing with problems between patients and technicians in the department in charge.
4	The patient's son and daughter-in-law have doubts about the medical personnel's handling of the sudden change.
5	This is a scene in which a patient and a medical professional are discussing a treatment plan for a nerve palsy that has appeared since the surgery.
6	A patient who was seen for abdominal pain is misdiagnosed initially and speaks with the corresponding physician.
7	A family member of a patient who has accidentally swallowed a partial denture is talking with a medical professional, a nurse.
8	Patient speaks of how nurses treated him during his hospitalization.
9	The patient is just after surgery. He discusses the response of the resident in charge of the patient.
10	A bereaved family member who lost a patient tells the story.
11	A bereaved family member who lost a patient suddenly told the story.

Figure 2: Instructions to the generating AI (ChatGPT).

```
#Instructions
You are a medical mediator (Japanese medical dialogue promoter).
Please analyze and output patient narratives according to the following #constraints.
#Constraints
*Output format must be followed
*Analysis must be based on #patient narratives and #situation content
*Evaluate "emotion". *Emotions are "joy," "anticipation," "anger," "disgust," "sadness,
Surprise, Fear, Trust, or Anxiety. In addition, evaluate them. Only these nine categories should be evaluated, and
emotions other than these nine should not be listed.
The reason for the rating should be listed in the # of the patient's narrative.
*Reasons for the rating should be cited from the patient's narrative.##Situation
# Patient's Narrative
Output format
*"Emotion."
*Reasoning.
```

Figure 3: Evaluation items and methods for the emotional results of the generated AI (ChatGPT).

Assessment Items

1. Is the sentiment consistent with the response regarding " feelings" ? Evaluate with I.
2. Is the rationale for the response regarding " emotion" appropriate? Evaluate with II.
3. Is the reason for the response regarding " emotion" vague? Evaluate with III.
4. In response to the " emotion" question, does the emotion expressed in the rationale match the " emotion" ? Evaluate with I.
5. For the response on "emotion," is the emotion expressed in the rationale appropriate? Evaluate with II.

Evaluation Methods

I: ①100% match ②75% match ③50% match ④25% match ⑤0% match
II: (1) 100% appropriate (2) 75% appropriate (3) 50% appropriate (4) 25% appropriate (5) 0% appropriate
III: (1) 100% unambiguous (2) 75% unambiguous (3) 50% unambiguous (4) 25% unambiguous (5) 0% unambiguous

Table 2: Generative AI (ChatGPT) answers to the narrative of Case 1.

Narrative of a patient's family	Answer 1	Answer 2
<p>My father had been suffering from severe constipation for a long time. It took days from the time he was admitted to the hospital until he was diagnosed, and in the end, the disease was found to be in such an advanced stage, but my father trusted the doctor like a god and had been coming to the hospital for a long time, but he was betrayed. I'm sure my father must be holding a grudge against him now.</p>	<p>#Emotions Emotions felt from the patient's family narrative are "anger" and "sadness".</p> <p>#ReasonThe family of the patient is angry because their father was suffering from severe constipation. They express anger that the attending physician instructed them that over-the-counter laxatives were fine and did not understand their father's suffering and did not take appropriate action. They are also surprised and saddened that the attending physician did not know that the patient's wife was accompanying him. He also expresses disappointment and sadness that it took days to diagnose the disease and that it was discovered in an advanced state. He also feels disgust and sadness at the loss of trust in the doctor, as he did not receive proper medical care, despite the fact that his father had trusted the doctor and had visited him for a long time.</p>	<p>#Emotion: Anger Reason: The patient's daughter is angry that her father has been suffering from severe constipation for a long time and her doctor did not provide proper medical care, only recommending over-the-counter laxatives.</p> <p>She is also angry that the attending physician did not know that the patient's wife was accompanying her and did not provide an appropriate explanation, and regrets that she should have been more proactive.</p> <p>He also attributes his anger to the fact that it took so long from the time he was admitted to the hospital until he was diagnosed, and that his disease had progressed.</p> <p>My father trusted his doctor like a god, and the fact that his trust was betrayed suggests that he may also be angry with his doctor.</p>

Table 3: Rater ratings of generated AI (ChatGPT) emotional output results.

Question	1	2	3	4	5
100%	7(15.9)	6(13.6)	23(52.3)	22(55.0)	19(43.2)
75%	21(47.7)	22(50.0)	10(22.7)	6(13.6)	14(31.8)
50%	11(25.0)	11(25.0)	10(22.7)	10(22.7)	8(18.2)
25%	1(2.3)	1(2.3)	1(2.3)	5(11.4)	3(6.8)
0%	4(9.1)	4(9.1)	0(0.0)	1(2.3)	0(0.0)
Total number of responses	44(100.0)	44(100.0)	44(100.0)	44(100.0)	44(100.0)
<ol style="list-style-type: none"> 1. Is the sentiment consistent with the response regarding "feelings"? 2. Is the reason for the response regarding "emotion" appropriate? 3. Is the rationale vague in response to the "emotion" response? 4. In response to the "emotion" response, is the emotion expressed in the supporting reasons consistent with the "emotion"? 5. For the response regarding "emotion," is the emotion expressed in the supporting reasons appropriate? 					

Results

Table 3 shows the aggregate results of the emotion judgments of the ChatGPT by the generated AI; the total percentage of 100% agreement and 75% agreement ranged from 63.6% to 75.0% for questions 1 through 5, respectively. The reason statements evaluated in Question 2 were quotes from the narrative and interpreted feelings. All of the responses that were interpreted as emotions were inconclusive. Next, the context of the emotional discrepancy was indicated as inadequate or excessive. The most common emotion perceived as excessive was anger. There were also many responses that listed only one emotion.

Next, the following characteristics of the responses were noted. First, (a) In the case of simple structures such as a single emotion, the emotion was appropriately captured. Second, (b) when emotions were mixed in a complex way, the rating of agreement decreased. Also, (c) in the case where the patient died, mixed responses were generated without distinguishing between past emotions and emotions at the time of the narrative. Furthermore, in the case of (d) where the interest (expression of interest, desire, and values: hereafter, interest) changed during the course of the narrative, the respondent responded to the emotion by addressing only the first half of the interest and ignoring the interest after the change.

Discussion

Table 3 shows that the total percentage of 100% agreement and 75% agreement ranged from 63.6% to 75.0%. 75% agreement was adopted because the three raters were two men and two women in their 20s with limited emotional and life experience and a medical mediator

who had gained the patients' trust and supported collaborative decision making during actual interviews with the patients. Since the hypothesis was that the results would "accurately capture negative emotions," the 100% agreement between the emotions responded to by the ChatGPT and the emotions responded to by the raters was low (15.9% and 13.6%), if 100% agreement is considered "accurate" in the hypothesis, the 100% agreement between the emotions (Question 1) and their rationale (Question 2) was low (15.9% and 13.6%). The results showed that the AI was unable to accurately assess negative emotions in the items that must be given the most weight in human emotion assessment. This indicates that emotion evaluation based on language alone is limited or impossible, considering that humans evaluate the emotions of others by synthesizing the situation, context, and nonverbal messages and matching them with their own interests. The definition of accurate should have been clarified in order to refine the evaluation.

As shown in Figure 3, "anger" was frequently over-rated. We considered that this was caused by grasping only the final emotion, "anger," and ignoring the primary emotion that caused the anger.

Next, for results (a)-(d), we considered that the emotions in the language of narration can be accurately taken, but not in the area of judging by context.

The following points are necessary to improve the agreement of ChatGPT's emotion judgments with human evaluations. For example, parameters such as environment, atmosphere, facial expressions, and tone of voice, which are quasi-linguistic and non-linguistic. It

is necessary to add these elements as linguistic information. The emotional evaluation of ChatGPT, a generative AI, was limited to the age of the evaluator and the number of evaluators. For more accurate evaluation, it is necessary to add parameters such as the age of the evaluators, the number of evaluators, and their expertise. It is also important to clarify the type of linguistic information.

Conclusion

The negative sentiment evaluation of the generative AI was only partially affirmed. The emotion evaluation of the ChatGPT of the generated AI based solely on linguistic information at the time of this study is limited. At present, it is difficult to accurately identify emotions in detail.

Conflicts of Interest

There are no corporate or other COI relationships that should be disclosed.

References

1. Japanese Ministry of Education, Culture, Sports, Science and Technology: Tentative Guidelines for the Use of Generative AI at the Primary and Secondary Education Level. <https://www.mext.go.jp/content/20230710-mxt_shuukyo02-000030823_003.pdf> (see 2023/10/01).
2. Ministry of Education, Culture, Sports, Science and Technology of Japan: Handling of teaching and learning aspects of generative AI in universities and technical colleges (Notification) <https://www.mext.go.jp/content/20230714-mxt_senmon01-000030762_1.pdf> (see 2023/10/01).
3. Yoshitaka Wada, Toshimi Nakanishi: Medical mediation: A narrative approach to conflict management. Seigne Ltd., 2011 Japanese
4. Takayuki Fukatsu: ChatGPT Work Art Revolution, Weekly Toyo Keizai, 2023/4/22, Toyo Keizai Shinposha, 42-43.
5. Robert Plutchik (2001) The Nature of Emotions. *American Scientist* 89: 344-350.
6. Kaarre J, Feldt R, Keeling LE, Dadoo S, Zsidai B, et al. (2023) Exploring the potential of ChatGPT as a supplementary tool for providing orthopaedic information. *Knee Surg Sports Traumatol Arthrosc* 31: 5190-5198. [[crossref](#)]
7. Robert C Benirschke, Joshua Wodskow, Kritika Prasai, Andrew Freeman, John M Lee, et al. (2023) Assessment of a large language model's utility in helping pathology professionals answer general knowledge pathology questions, *American Journal of Clinical Pathology* aqad106. [[crossref](#)]

Citation:

Nakanishi T, Tokuda A, Arao T, Sugiura Y, Hayasaka T (2023) How does the Language Prediction Model, the ChatGPT Evaluate Negative Emotions? *J Clin Res Med* Volume 6(3): 1-4.

J Clin Res Med, Volume 6(3): 4-4, 2023