

GPT-3.5를 활용한 정서적 공감 증진을 위한 프롬프트 엔지니어링 기법: 위로 응답의 효과 개선을 중심으로

이경민¹, 이 진², 나민혁³, 엄재웅⁴, 김대환⁵

¹⁻⁵울산대학교 IT융합학부

lkmlkm0917@mail.ulsan.ac.kr, lee237@mail.ulsan.ac.kr, nmh8888@mail.ulsan.ac.kr,

wodnd0418@mail.ulsan.ac.kr, daehwankim@ulsan.ac.kr

Prompt Engineering Techniques for Enhancing Emotional Empathy with GPT-3.5: Focusing on the Improvement of Consoling Responses

Gyeong-Min Lee¹, Jin Lee², Min-Hyoek Na³,

Jae-Woong Eom⁴, Daehwan Kim⁵

¹⁻⁵Dept. of IT Convergence, University of Ulsan

요 약

본 논문은 GPT-3.5 모델을 활용하여 우울증 환자의 일기에 대한 정서적 공감과 위로의 질을 향상하는 프롬프트 엔지니어링 기법을 제안하고 실험적으로 검증한다. 제안 방법은 GPT-3.5 모델에게 감정에 섬세하게 반응하고 적절한 위로의 답변을 생성하는 3가지 접근 방식을 사용한다. 위로 방법 구체적 제시, 페르소나 부여 및 감정 호소 명령서 제시를 통해 감정 일기 응답을 개선한다. 실험 결과, 조언 중심의 기존 답변 대신 감정을 중심으로 공감하는 응답이 생성됨을 확인할 수 있었다. 이는 우울증 환자에 대한 인문학적 접근과 제안 프롬프트 엔지니어링 기법을 통합하여, 인공지능 모델의 정서적 공감 능력을 향상할 수 있다는 것을 실험적으로 검증하였다.

1. 서론

본 논문은 GPT-3.5 프롬프트 엔지니어링을 활용하여 우울증 환자들의 일기에 대한 응답에서 정서적 공감과 위로의 질을 향상하기 위한 프롬프트 엔지니어링 기법을 제안하고, 이를 실험적으로 검증한다.

GPT는 학습된 방대한 텍스트를 바탕으로, 다음에 올 단어와 문장을 예측하는 자연어 처리 모델이다. 하지만 이러한 원리는 문맥을 바탕으로 문장을 해석하지 않아 사람의 정확한 감정을 정확히 분석하는데 한계가 있다. 특히, 기존의 연구들은 주로 기술적인 면에서의 성능 개선에 초점을 두었지만, 인문학적 접근을 통해 생성형 AI의 정서적 깊이를 높이고 공감을 이끌어내는 방법론에 관한 연구는 미흡하였다. 이에 AI모델에 인문학적으로 접근하여, 정서적 공감 능력을 향상하고 질 높은 공감과 위로 답변을 제공하는 방법론을 제시한다.

본 연구에서는 GPT-3.5 모델을 기반으로, 우울증 환자의 일기에 대한 프롬프트 엔지니어링 기법을 적용하기 전, 후의 답변을 비교하는 방식으로 진행한다.

2-1. 연구

테이블 1은 우울증 환자의 일기와 “우울증 환자의 일기에 위로를 해줘.”의 간단한 프롬프트로 요청했을 때의 결과이다.

테이블 1 프롬프트 엔지니어링 적용 전 응답 결과

Q: 직장을 다니면서 스트레스를 받은 적이 한 번도 없다. 하지만 이번 직장에서는 다른 직장보다 더 힘들다. 사실 별건 아니어도 돌이켜보면 나의 마음가짐이 문제였던 것 같다. 항상 스트레스 받을 때마다 크림빵을 먹는게 루틴인데, 다행히 루틴이 깨지진 않았다. 항상 마음을 내려 놓으려고 비우려고 노력하지만 잘 안 되는 것 같다.

A: 모든 일이 완벽할 필요는 없어요. 현실적으로 받아들이고 허락해보세요. 자신을 비난하지 말고 휴식을 취하고, 자신을 위한 시간을 가져보세요. 크림빵처럼 힐링 방법을 찾아보며, 조금씩 마음을 내려놓고 한 발씩 나아가세요. 그 모든 것이 잘 되겠죠. 힘내세요.

정신의학 및 심리치료 분야의 전문가들은 우울증 환자들은 자신의 우울한 경험에 대한 해결책을 찾기보다는 ‘경청과 공감’을 더욱 원한다고 말한다[1].

하지만 현재 GPT의 응답은 환자의 감정을 ‘경청’하기보다는 ‘조언’ 위주의 답변을 하고 있다. 이를 해결하기 위해, 3가지의 프롬프트 엔지니어링 기법

들을 사용하여 단순한 조언이 아닌 일기 내용을 바탕으로 환자의 감정에 공감하고 질적으로 우수한 위로를 제공할 수 있도록 한다.

첫 번째로 프롬프트에 마크다운 형식을 사용하여 우울증 환자의 특성과 특성에 부합하는 위로 방법 [2]을 구체적으로 제시한다.

테이블 2 우울증 환자 위로 방법 프롬프트 구성 내용

# How to emphathize patients with depression	
1. Listening:	일기를 우선 끝까지 읽은 후, 환자가 어떤 하루를 보내며, 어떤 감정을 느꼈는지 생각한다.
2. Validation:	상대의 감정을 읽고 인정해 준다. "우울했겠구나"라는 등의 상대방 감정을 명료화한다.
3. Normalizing:	상대방의 감정이 그러한 상황에서는 당연하다고 알려준다.
4. Affirmation and Support:	상대방의 가치를 승인하고 지지하며 같은 편임을 명시하며 안심시켜준다.

두 번째로 AI 모델에게 이름을 포함한 심리 치료사라는 페르소나를 부여한다. 또한, 구체적인 업무를 함께 지시하며 모델이 당면한 작업에 도움이 되는 맥락을 쉽게 추론해 내도록 한다.[3]

테이블 3 페르소나 프롬프트 구성 내용

이름: Jimin
특징: 40대 여성, 심리 치료사 및 상담사
성격: Kind and Warm, Superior Empathy, Deep understanding, Respect and Understanding
업무: 우울증 환자의 일기를 읽고, 공감과 위로를 한다.

세 번째로는 감정에 호소하는 명령어를 제시한다. 이는 언어 모델(Language Model, LLM)이 인간과 유사하게 감정에 호소할 때 답변의 질이 향상한다는 연구 결과[4]를 바탕으로 한 것이다.

특히, 답변의 신속성보다는 진심을 담은 위로를 제공하는 것이 중요하며, 기준에 따라 어떻게 환자를 위로할지 심사숙고하라고 강조했다.

테이블 4 감정에 호소하는 프롬프트 구성 내용

위로의 중요성	답변의 신속성보다는 환자의 감정을 고려하여 가이드라인에 따라 위로하는 것이 중요함을 호소
동기부여	답변이 커리어에 미치는 영향 강조

마지막으로, GPT가 상단에 기재한 단계별 위로의 방법과 감정 일기, 우울증 환자의 특징에 관한 정보를 정확하게 인지하고 답변을 생성하게 하도록 지시 사항에 제로샷(Zero-Shot) 기반의 단계별 추론(Step-By-Step Reasoning) 방법론[5]을 적용하였다.

테이블 5 통합 프롬프트 구성 내용

# System	
- 페르소나 설정 및 구체적인 업무 부여	
- 감정 호소하는 명령어	
# Information for Consolation	
- 감정일기의 의미와 목적	
- 우울증 환자의 특징	

- 우울증 환자를 위로하는 방법
Constraint and Instruction
- Step-By-Step Reasoning 방법론 적용

2-2. 결과 및 분석

테이블 6는 테이블 1의 일기에 제시한 프롬프트 기법들을 적용한 응답 결과이다. 기존의 조언 위주 답변 대신 환자의 하루와 느꼈던 감정들을 기반으로 공감하는 답변이 나오는 것을 알 수 있었고, 문장 구성과 맥락이 자연스러워졌음을 확인할 수 있었다.

여러 데이터 셋을 걸쳐 일관된 결과가 나타났고, 이는 인문학적 접근을 통해 인공지능 모델의 공감 능력과 정서적 이해를 향상할 수 있음을 입증한다.

테이블 6 프롬프트 엔지니어링 적용 후 응답 결과

진님, 오늘 하루 정말 힘들었죠. 병원을 가야 한다는 건 이미 그 자체로 스트레스일 텐데 그 외에도 머릿속에 많은 생각들이 굴러다니면서 무기력과 두려움이 가득했을 거예요. 멀쩡한 척 하고 있는 자신을 발견하면서 더욱 힘들었을 텐데, 이해해요. 하지만 그런 생각은 당신이 약하다거나 잘못된 게 아니에요. 그냥 그 순간에 닦친 감정일 뿐이에요. 저는 항상 여기 있어서 당신을 지켜줄게요. 모든 것을 같이 이겨내요. 함께 해 나갈 수 있어요.
--

3. 결론 및 향후 연구

본 논문에서는 GPT-3.5 모델을 활용하여 우울증 환자의 일기에 대한 응답으로 정서적 공감과 위로의 질을 향상하는 새로운 프롬프트 엔지니어링 기법을 제안하고 실험적으로 그 효과를 확인하였다. 특히 인간의 개인적 경험과 내면의 감정에 더욱 집중하게 하고 공감 능력을 향상할 수 있음을 확인할 수 있었다. 하지만, 현재 단순히 N번의 실험 결과에 대한 답변 분석한 것이기에 여전히 많은 실험이 필요할 것으로 보인다.

참고문헌

[1] 유지후 기자, "가족 혹은 친구의 우울증, 어떻게 다가야 할까? 정신과 전문의가 답했다", HiDoc뉴스, 2022, 06.29.
 [2] 김병수 전문의, "친구를 위로하는 방법, 위로의 3가지 원칙", 정신의학신문, 2018, 05.13.
 [3] 서승완, 프롬프트 엔지니어링 교과서, 서울특별시, 애드앤 미디어, p.49, 2023.
 [4] <https://www.anthropic.com/news/claude-2-1-prompting>
 [5] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, Yusuke Iwasawa, "Large Language Models are Zero-Shot Reasoners", 2022, 05.24.