KOMIT : 행동 변화 유도를 위한 한국어 동기면담 심리상담 데이터 셋

유동제‡, 민경은‡, 나예찬‡, 여진영*

‡ 유동제·민경은·나예찬은 본 연구에 동등하게 기여함. * 교신저자: 여진영

연세대학교

pass120@yonsei.ac.kr, roddmsl0208@gmail.com , ahdfyd@yonsei.ac.kr , jinyeo@yonsei.ac.kr

KOMIT: Korean Motivational Interviewing Dataset for Evoking the Behavior Change

Dongje Yoo*, Kyeongeun Min*, Yechan La*, Jinyoung Yeo

Yonsei University

요 약

비대면 심리상담 수요가 증가하면서 상담 품질을 일정 수준 이상으로 유지하는 한국어 심리상담 데이터셋의 필요성이 커지고 있다. 동기면담(Motivational Interviewing, MI)은 내담자가 스스로 변화의 필요성을 수용하고 변화를 계획하도록 돕는 협력적 대화 방식이지만, 지금까지 공개된 MI 데이터셋은 단순 반영(reflection) 중심으로 구성되어 행동 변화를 촉진하는 전략적 발화가 부족하다는 한계가 있다. 본 연구는 MI 4 단계(Engaging - Focusing - Evoking - Planning)를 따르는 멀티턴 상담 세션을 대규모로 합성하는 파이프라인을 제안한다. 정신건강 커뮤니티 상담 사연을 기반으로 내담자 프로파일을 추출하고, 신념 기반 상태 전이 및 행동 선택을 수행한 뒤, LLM 이 내담자/상담사 발화를 번갈아 생성한다. 또한 RAG를 통한 MI 상담 데이터 예시와 COT 추론 구조를 주입하여서 상담의 전문성을 더하였다. 총 1,000 개 세션을 생성·평가한 뒤, 6 가지 MI 품질 지표 기준으로 필터링하여 812 개의 고품질 세션을 확보하였다. 유효성 평가에서 기존 데이터셋 대비 유도력·효과성 지표가 크게 개선되었고, 모델 학습 후 시뮬레이션 비교에서도 전반적 점수가 향상되어 심리상담 활용의 효과성을 보이고 있다.

Ⅰ. 서 론

심리상담의 비대면 전환이 가속화되면서 AI 심리상담 시스템에 대한 수요와 연구가 활발해졌으나, 실제 상담 현장에서 요구되는 전략적 유연성은 여전히 부족하다[4]. 특히 한국어 공개 상담 데이터는 제한적이며, 기존 합성 기반 한국어 동기면담 데이터셋(KMI[3])은 반영 중심 발화 비중이 매우 높아 변화 유도(Evoking) 단계에서 필요한 발화 패턴을 학습하기 어렵다. 본 연구의 목표는 (1) 행동 변화를 유도하는 동기면담 구조를 반영한 고품질 한국어 멀티턴 데이터셋을 구축하고, (2) 구축한 데이터와 데이터 기반 MI 챗봇을 개발·평가하는 것이다. 해당 동기면담 데이터셋(KOMIT)과 프롬프트,코드는 깃헙[1]에 공개한다.

Ⅱ. 본론

1) 이론적 배경: 동기면담(Motivational Interviewing)은 협력(Partnership)·수용(Acceptance)·공감(Compassion)·유도(Evocation)를 핵심 정신으로 하며, OARS(Open Questions, Affirmation, Reflection, Summarizing)를 통해 내담자의 양가감정과 자신감을 다룬다[6]. MI 세션의 대화는 Engaging - Focusing - Evoking - Planning 의 4 단계의 순서를 따르며, 단계 전환을 위해서는 맥락에 맞는 질문·강화·요약이 조합되어야 한다[6]. 특히 Evoking과 Planning은 변화 동기와 계획을 직접 다루는 단계로 라포 형성에 초점을 맞추는 이전 단계와는 다른 전략 분포가 요구된다. 또한 동기면담에서 내담자의 상태 단계는 Precontemplation, Contemplation, Preparation, Termination 단계를 거친다[5]. 본 연구진은 상담사와 내담자의 발화를 설계할 때 이러한 이론적 배경을 고려하였다.

2) 문제 정의: 본 연구진은 KMI를 LLM(Hyperclovax-0.5B[2])으로 학습하고 내담자 시뮬레이터를 구축해 시뮬레이션을 진행했다. 그 결과, 행동 변화를 촉진하는 것을 목표하는 MI의 취지와는 다르게 모델이 공감적 발화(반영) 전략에 치중하여 발화하는 현상이 발견되었다. 이는 데이터의 구성 자체가 반영 위주로 구성이 되어 있기 때문이라고 분석된다. 이에 본 연구진은 반영뿐 아니라 행동 변화를 촉진하는 전략도 균형 있게 수반한 한국어 MI 데이터 셋을 합성하고자 한다.

3) 데이터 생성 파이프라인: (그림 1)과 같이, (a) 실제 상담 사연을 정신건강 커뮤니티[3]로부터 수집해 저장한 뒤, (b) LLM(gpt-5-mini)이 사연 요약·목표·신념을 포함한 프로필을 추출한다. (c) 이 프로필 기반의 내담자 역할의 LLM은 사전 정의된 목표와 신념이 다루어진 정도에 따라 내담자 상태 전이가 결정되고 내담자 상태에 해당하는 전략[5]을 선택하여 발화한다. (d) LLM(gpt-5)이 내담자 발화와 MI 전략이 포함된 상담사 발화를 교대로 생성하여 발화 20 턴 세션을 만든다. 이때 상담사 LLM은 적응적 전략 선택과 깊은 공감 유도를 위해 내담자의 상황,감정 파악과 상담 단계 파악,전략 선택을 답변 생성 전에 하도록 한다. 이때 상담 단계와 전략은 MI 공식 문서[6]의 설명을 기반으로 한다. 응답할 때는 <think></think><answer></answer> 형식으로 응답하게 한다.

전략 반복을 방지하기 위해 동일 전략 3 회 연속 사용을 금지하고, RAG를 통해 실제 MI 상담 데이터(AnnoMI[2])를 참조해 전문성을 더하였다.

4) 데이터 정제 및 분할: 생성된 1,000 개 세션을 MI 품질 지표 6개(협력감, 수용성, 공감력, 유도력, 유사성, 효과성)로 LLM을 이용해 평가하고, 평균 4.0 점 이하 또는 개별 2.0 점 이하 세션을 제거하여 812개 세션을 확보하였다. 이후 8:2 로 학습/평가 데이터를 분할하였다.

5) 평가: 데이터의 품질 평가를 위해 본 연구 데이터(KOMIT)와 KMI를 MI의 6가지 평가 지표에 대하여 LLM을 활용해 5점 리커르트 척도로 평가를 하였다. 평가 결과(표1), KOMIT은 KMI 대비 협력감(+0.26), 수용성(+0.30), 공감력(+0.35), 유도력(+0.85), 유사성(+0.31), 효과성(+1.32)로 행동 변화와 관련된 지표인 유도력과 효과성에서의 우세가 두드러진다.

1) https://github.com/dongje/komit
2) naver-hyperclovax/HyperCLOVAX-SEED-Text-Instruct-0.5B
3) https://www.mindcafe.co.kr/

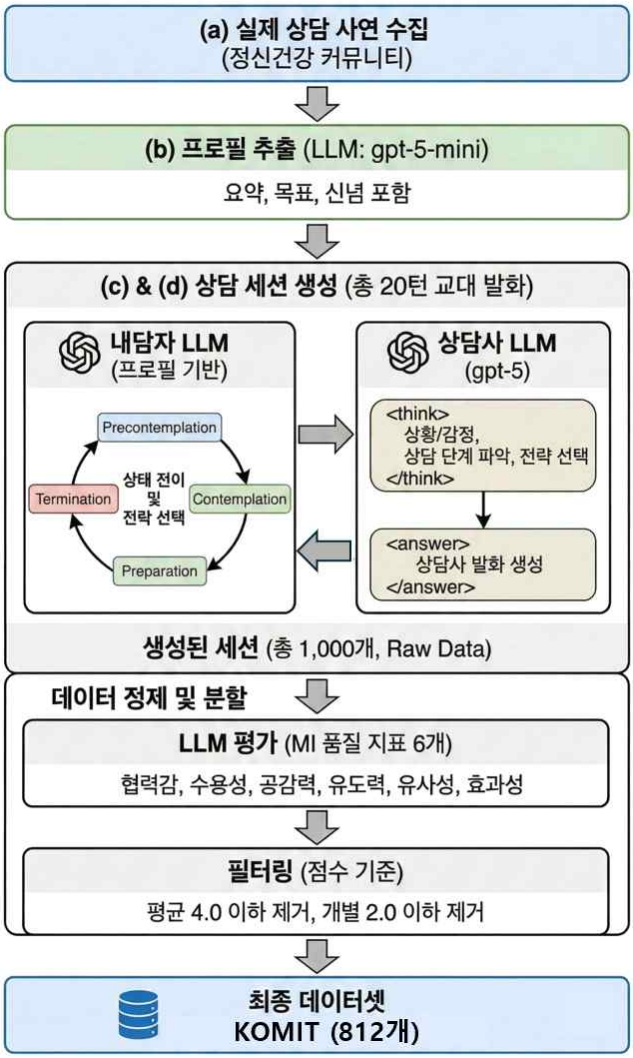

그림 1 KOMIT 데이터 생성 및 정제 파이프라인

Metric	KMI	KOMIT	Improvement
협력감(Partnership)	3.89	**4.15**	**+0.26**
수용성(Acceptance)	4.60	**4.90**	**+0.30**
공감력(Compassion)	3.65	**4.00**	**+0.35**
유도력(Evocation)	3.10	**3.95**	**+0.85**
유사성(Similarity)	3.39	**3.70**	**+0.31**
효과성(Effectiveness)	2.43	**3.75**	**+1.32**

표 1 데이터셋 품질 비교 평가 결과

또한 데이터 학습의 효용성 평가를 위해 LLM(Hyperclovax-0.5B)에 KMI와 KOMIT을 각각 2에폭으로 학습하여 내담자 시뮬레이터와 상담을 하도록 시뮬레이션했다. 시뮬레이션에 사용된 내담자 사연은 held-out 케이스 63개이다. 20턴까지 시뮬레이션을 한 후 LLM을 통하여 6가지 MI지표에 대해서 5점 리커르트로 평가를 하였다. 그 결과(표2), KOMIT 학습 모델은 KMI 대비 협력감(+0.73), 수용성(+0.71), 공감력(+0.33), 유도력(+1.46), 유사성(+0.75), 효과성(+1.54)로 유도력과 효과성 뿐만 아닌 모든 지표에서 베이스라인보다 우세를 보임으로써 상담 효용성을 보이고 있다.

Metric	KMI	KOMIT	Improvement
협력감(Partnership)	2.97	**3.7**	**+0.73**
수용성(Acceptance)	3.78	**4.49**	**+0.71**
공감력(Compassion)	2.97	**3.3**	**+0.33**
유도력(Evocation)	2.13	**3.59**	**+1.46**
유사성(Similarity)	2.44	**3.19**	**+0.75**
효과성(Effectiveness)	1.67	**3.21**	**+1.54**

표 2 Hyperclova(0.5B) 학습 모델의 상담 시뮬레이션 평가

6) 한계 : 본 연구의 품질 평가는 LLM 기반 지표에 의존하므로, 평가자 편향과 실제 임상가 판단 간 괴리가 존재할 수 있다. 또한 합성 데이터 기반 학습은 실제 내담자 다양성과 위기 상황(자·타해 위험 등)을 완전히 포괄하지 못한다. 따라서 KOMIT 데이터와 챗봇은 임상 대체가 아닌 연구·보조 도구로 활용되어야 하며, 실제 적용 시 인간 전문가 감독과 안전 가드레일이 필요하다.

Ⅲ. 결론

본 연구는 MI 4 단계를 따르는 한국어 멀티턴 상담 데이터셋을 대규모로 합성하기 위해, 프로파일 기반 내담자 시뮬레이션과 상담사 MI 전략 생성이 결합된 파이프라인을 제안하였다. 정제 기준을 통해 812 개의 고품질 세션을 확보했으며, 유효성 및 시뮬레이션 평가에서 기존 한국어 동기면담 심리상담 데이터셋(KMI) 대비 유도력·효과성 등 핵심 지표 개선을 확인했다. 보완을 위한 향후 연구로는 (1) 내담자 상태 분포를 하나에 편중되지 않게 제어하는 시뮬레이션 설계, (2) 사람/전문가 기반 평가를 통한 외적 타당도 확보가 필요하다.

ACKNOWLEDGMENT

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 한국형 ARPA-H 프로젝트 지원에 의하여 이루어진 것임(과제고유번호 : RS-2024-00512374).

참 고 문 헌

[1] Miller, W. R., & Rollnick, S. (2012).*Motivational interviewing: Helping people change*. Guilford press.

[2] Wu, Z., Balloccu, S., Kumar, V., Helaoui, R., Reiter, E., Recupero, D. R., & Riboni, D. (2022, May). Anno-mi: A dataset of expert-annotated counselling dialogues. In*ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*(pp. 6177-6181). IEEE.

[3] Kim, H., Lee, S., Cho, Y., Ryu, E., Jo, Y., Seong, S., & Cho, S. (2025, April). KMI: A Dataset of Korean Motivational Interviewing Dialogues for Psychotherapy. In*Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*(pp. 10803-10828).

[4] Cho, Y. M., Rai, S., Ungar, L., Sedoc, J., & Guntuku, S. (2023, December). An integrative survey on mental health conversational agents to bridge computer science and medical perspectives. In*Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*(pp. 11346-11369).

[5] Yang, Y., Achananuparp, P., Huang, H. Y., Jiang, J., Lim, N. G., Ern, C. T. S., ... & Lim, E. P. (2025, July). Consistent client simulation for motivational interviewing-based counseling. In*Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*(pp. 20959-20998).

[6] Motivational Interviewing Network of Trainers. (n.d.). Motivational Interviewing Network of Trainers (MINT). Retrieved January 1, 2026, from https://motivationalinterviewing.org/