특집논문 (Special Paper)
방송공학회논문지 제28권 제5호, 2023년 9월 (JBE Vol.28, No.5, September 2023)
https://doi.org/10.5909/JBE.2023.28.5.564
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

대화형 봇에 의한 대화 주제 정렬 과정에서의 인공지능 윤리 가이드라인에 관한 연구

방 준 성[a], 이 병 탁[a], 박 판 근[b]‡

A Study on Guidelines for AI Ethics in the Process of Aligning Conversation Topics by Conversational Bot

Junseong Bang[a], Byung-Tak Lee[a], and Pangun Park[b]‡

요 약

본 고에서는 대화형 봇이 사용자 그룹 간 대화에 참여할 때 준수해야 할 인공지능 윤리의 가이드라인 개발을 위한 고려사항에 대해 연구한다. 대화형 봇은 특정 사용자 그룹과의 제한된 상호작용 속에서 사용자 간 토픽 일치성, 친밀성 등에 영향을 받는다. 이에 따라 인공지능 윤리 기준을 상황에 맞게 적용할 필요가 있다. 구체적으로 본 연구에서는 대화형 봇의 대화 주제 정렬 과정에서 발생할 수 있는 윤리 기준의 가변성을 지적하고, 이를 효과적으로 처리할 수 있는 윤리적 응답 생성 프레임워크를 제안한다. 또한 제안한 프레임워크에서 활용하는 사용자 간 토픽 일치성 및 친밀도가 정량적인 수치로 반영할 수 있음을 실험을 통해 확인하였다.

Abstract

In this study, we investigate the considerations for developing guidelines on artificial intelligence (AI) ethics that conversational bots should adhere to when participating in conversations among user groups. Conversational bots are influenced by user-to-user topic alignment, intimacy, and something else within limited interactions with specific user groups. Consequently, it is necessary to apply ethical rules for AI appropriately in these situations. Specifically, this research highlights the variability of ethical rules that may arise during the conversation topic alignment process of conversational bots and proposes an ethical response generation framework to effectively address them. Furthermore, through experiments, we confirm that the proposed framework can quantitatively reflect user-to-user topic alignment and intimacy.

Keyword : AI Ethics, Conversational Bot, Group Chatting, Conversation Topic Alignment, Conversation Friendship

a) 주식회사 와이매틱스(Ymatics Corporation)
b) 충남대학교(Chungnam National University)
‡ Corresponding Author : 박판근(Pangun Park)
E-mail: : pgpark@cnu.ac.kr
Tel: +82-42-821-6862
ORCID: https://orcid.org/0000-0003-3744-4476
· Manuscript August 3, 2023; Revised September 4, 2023; Accepted September 18, 2023.

Ⅰ. 서 론

인공지능(AI: Artificial Intelligence)은 인간의 학습, 지각, 추론 등의 능력을 인공적으로 구현한 컴퓨터 시스템이다[1]. 디지털 전환(Digital Transformation)과 함께 AI 기술이 고도화되며 AI는 제조, 의료, 교육, 금융 등 산업 전반의 다양한 분야에 그 활용 가능성을 보여주고 있다. 대화형 인터페이스(Conversational Interface)는 텍스트나 음성을 기반으로 사용자와 시스템 사이의 자연스러운 상호작용을 가능하게 하는 기술로, 챗봇(Chatbot) 또는 음성봇(Voice-bot)의 형태로 사용자에게 정보를 제공하거나 사용자의 명령을 수행한다. AI 기술이 활용된 대화형 인터페이스의 발전은 서비스를 지능화하고 접근성을 높여 사용자의 경험을 개선한다[2].

대화형 봇(Conversational Bot)은 서비스를 제공하기 위해 사용자와 대화를 잇게 되는데 그 과정에서 윤리적 고려사항을 따라야 한다[3]. 대화형 AI 봇의 활용이 증가하는 만큼, 개인정보나 기밀 유출, 편향된 응답, 개인과 사회에 대한 영향 등 다양한 문제점이 제기되고 있다[4]. 대화형 봇을 설계할 때 사용자의 권리와 안전을 보장하기 위해 AI 윤리(AI Ethics)에 관한 연구와 논의를 확장하여 지속적으로 진행할 필요가 있다[5]. 대화형 봇은 개인 사용자와 1:1 대화뿐만 아니라 그룹 채팅 환경에서 다수의 사용자와 동시에 상호작용 할 수 있다. 이때, 대화형 봇은 그룹 채팅 과정에서 각 사용자의 요구를 처리하기도 하지만, 그룹의 대화 과정을 중재하며 공동의 목표를 달성할 수 있도록 대화를 돕기도 한다[6]. 다수의 사용자와의 상호작용은 대화형 봇의 설계와 구현에 있어서 복잡성을 증가시킨다[7]. 그룹 대화에 참여하는 대화형 봇의 윤리적 기준과 가이드라인 개발이 더 어려울 수 있음을 짐작할 수 있다.

메타버스는 가상과 현실이 융합된 공간에서 사용자가 다른 사용자 혹은 객체와 상호작용하며 경제·사회·문화 활동을 할 수 있는 디지털 세계이다. 메타버스 환경에서 대화형 봇은 사용자와 자연스러운 대화를 나누며 가상 환경 내에서 사용자를 안내하거나 필요한 정보를 제공할 수 있다. 대화형 봇은 메타버스에서 사용자의 선호도, 취향, 행동 등을 학습하여 개인별로 맞춤화된 서비스를 제공할 수 있다. 예를 들어, 대화형 봇은 사용자에서 적합한 가상현실 콘텐츠를 추천할 수 있다. 가상 상점이나 온라인 쇼핑몰에서 대화형 봇이 고객들에게 제품 정보를 제공하고 구매 도움을 줄 수 있다. 대화형 봇은 아바타의 형태로 사용자에게 가시화되어 상호작용에 참여할 수 있다. 메타버스와 같은 가상세계에서 대화형 봇은 그 활용 가능성이 높으며 그에 따라 대화 윤리의 기준을 따르는 AI 아바타 개발이 요구된다.

본 고에서는 사용자 그룹 대화에 참여하는 챗봇 혹은 음성봇의 대화 주제 정렬 과정에서의 AI 윤리 가이드라인 개발에 관한 연구를 다룬다. II장에서는 대화형 AI 봇에 대한 대화 윤리 기준 개발의 필요성을 언급한다. III장에서는 그룹 대화에서의 대화형 AI 봇의 대화 주제 정렬에 대해 살펴본다. IV장에서는 그룹 대화를 위한 윤리적 응답 생성 프레임워크 개발 방법을 제안하고 실험 결과를 제시한다. V장에서는 연구 결과를 요약하며 본 논문을 마무리한다.

Ⅱ. 대화형 AI 봇에 대한 대화 윤리 기준 개발의 필요성

1. 대화형 AI 봇의 역할과 기능

대화형 봇은 자연어로 대화하는 기술을 중심으로 정보 제공, 사용자 질의에 대한 응답, 작업 자동화 등의 역할을 수행한다. AI 기술의 발전은 대화형 봇이 사용자와 더욱 복잡하고 다양한 대화를 할 수 있게 했다[8]. 대화형 봇과 사용자 사이의 상호작용은 서비스 성공 여부를 결정하는 주요 인자이다. 그 상호작용의 과정을 통해 사용자의 의도를 정확히 이해하고 적절한 응답과 조치를 할 수 있어야 한다[9]. 사용자 경험 향상을 위해서는 대화형 AI 봇이 다양한 대화 스타일과 대화 유형을 지원해야 한다[10]. 상호작용에는 감정 표현이 있을 수 있는데 이는 사용자의 대화 경험에 큰 영향을 미친다[11].

대화형 봇의 대화 설계는 사용자가 이용하는 봇에 대한 긍정적 혹은 부정적 인식뿐 아니라 사용자의 감정 변화나 의사결정에 영향을 미친다[12]. 대화형 봇의 대화는 싱글-턴(Single-turn) 대화와 멀티-턴(Milti-turn) 대화로 구분할 수 있다. 싱글-턴 대화는 단일 문장의 질의와 그에 대한 단일 응답으로 이루어지는 대화 형태를 말한다. “오늘 서울 날씨가 어때?”라는 사용자 질의에 “서울은 오늘 맑은 날씨입니

다.”라고 응답하는 AI와의 대화를 예로 들 수 있다. 멀티-턴 대화는 사용자와 AI 사이에 여러 차례의 질의-응답 혹은 메시지 교환 과정을 통해 이루어지는 대화 형태를 말한다. 멀티-턴 대화에서는 대화 과정에서의 정보를 저장하고 대화 맥락 이해를 위해 정보를 적절히 활용할 수 있어야 한다. 멀티-턴 대화를 실현하기 위해 시나리오 기반 챗봇 시스템에서는 대화 상태 추적(Dialogue State Tracking) 기술에 대한 연구를 하고 있으며 대규모 언어 모델(Large Language Model: LLM) 기반의 AI 챗봇 시스템에서는 사전 학습 모델(Generative Pre-trained Transformer: GPT) 기반으로 맥락에 맞게 대화를 잇게 하기 위한 연구가 이루어지고 있다.

2. 대화형 AI 봇의 윤리적 책임

대화형 봇은 사용자와 대화를 진행하는 과정에서 개인정보를 처리하고 저장할 수 있다. 이때 개인정보의 보호와 데이터의 적절한 이용에 대한 책임이 있다. 대화형 봇은 사용자 질의를 처리하는 과정에서 사용자의 이름, 연락처, 주거, 직장, 취향 등의 민감한 데이터를 수집할 수 있다. 개인정보의 무단 수집, 저장, 공유는 사용자의 권리를 침해하는 행위이며 법적인 제재가 따를 수 있다. 사용자뿐 아니라 운영자가 인지하지 못하는 상태에서 개인정보가 유출될 수 있는 여지가 있다. 대화형 봇이 사용자에게 요구하지 않은 상황에서 사용자의 실수로 개인정보가 노출되거나 공유될 수 있기도 하다. 이러한 부분도 고려하여, 대화형 봇의 개발과 운영 과정에서 개인정보 보호와 관련된 법률 및 지침을 준수할 수 있어야 한다.

대화형 봇은 편견 없는 정보 제공과 공정한 대화를 유지해야 한다. 편향은 사용자의 불만을 초래할 뿐만 아니라, 사회적인 문제나 논란을 일으킬 수 있다. 특정 인종, 성별 또는 문화에 대한 편견이 포함된 데이터로 학습된 대화형 봇은 사용자에게 편향된 정보를 제공할 수 있다. 최근 몇 년 동안 챗봇의 편향에 의한 여러 차례 사고가 발생했다. 그룹 대화에서는 다양한 문화적, 사회적 배경의 사용자가 참여할 수 있다. 그로 인해 사용자별 편견 판단의 기준과 범위가 다를 수 있는데, 대화형 봇이 이러한 상황에서 대화를 중재하는 시스템적 기능에 대한 연구는 아직 부족하다.

대화형 봇의 편향적 응답은 개인과 사회에 영향을 줄 수 있다. 개인적 측면에서, 챗봇은 사용자에게 즉각적인 응답을 제공하여 정보 검색이나 문제 해결을 더욱 빠르게 할 수 있게 해주지만, 제공되는 정보의 정확성에 대한 우려도 있다. 사회적 측면에서 특정 집단이나 문화에 대한 사회적 편견이 강화되는 결과를 초래할 수도 있다. 대화형 봇의 동작과 표현은 시스템 설계자와 개발자에 의한 알고리즘과 데이터에 영향을 많이 받게 된다. 대화형 봇은 그 구현과 운영 과정에서 윤리적 영향이 고려되어야 한다. 지속적인 모니터링과 평가를 통해 대화형 봇 시스템의 윤리적 운영이 보장되어야 한다.

3. 대화형 AI 봇의 대화 윤리 개발을 위한 고려사항

대화형 봇과 인간의 대화에서 윤리에 대한 고려는 중요하다. 첫째, 대화형 봇은 사용자에게 서비스 제공에 있어서 개인정보를 보호하고 그 이용에 대해 적절한 기준을 따라야 한다. 사용자 동의 없는 개인정보 수집, 저장, 공유를 해서는 안되며, 사용자나 운영자의 인지가 부족한 상황에서 개인정보 유출이 발생할 가능성이 있는 경우에 이를 판별하여 알려줄 수 있어야 한다. 둘째, 대화형 봇은 신뢰성 있는 정보를 제공하여 사용자가 잘못된 의사결정을 하지 않도록 도와야 한다. 이를 위해 정보의 출처를 확인하여 사용자에게 제공할 필요가 있다. 셋째, 대화형 봇은 사용자에게 편향된 정보를 제공하지 않아야 한다. 공정하고 중립적인 태도로 사용자에게 응답해야 한다. 넷째, 대화형 봇은 사용자가 챗봇을 사람으로 오해하는 상황을 만들면 안된다. 인간 사용자는 대화형 봇이 전달하는 정보가 정확하지 않을 가능성에 대해 인지할 수 있어야 하며, 인간 사용자의 다른 사용자에 대한 대화 방식과 대화형 봇에 대한 대화 방식이 달라 대화형 봇과의 의사소통이 원활하지 않을 수 있기 때문이다. 다섯째, 대화형 봇은 사용자의 감정과 그의 사회적, 문화적 배경을 고려하여 대화에 임해야 한다. 이외에도 대화형 봇의 대화 설계에 있어서 다양한 윤리 고려사항을 찾아 시스템 구현이 가능하도록 연구할 필요가 있다. AI 시스템은 사회적, 문화적 배경을 갖는 다양한 사용자들에게 영향을 미치며 그 사용자들은 다른 방식으로 편향성을 경험하게 되기 때문이다. AI 시스템은 인간의 의사결정에 의해

추가적인 처리를 진행하기도 한다. 다양한 문화적 맥락에서 비롯되는 편향성의 다양한 형태를 이해하기 위해서 사회·문화적인 연구도 필요하다[13].

Ⅲ. 그룹 대화에서의 대화형 AI 봇의 대화 주제 정렬

대화 주제 정렬(Conversation Topic Alignment)은 대화형 봇과 사용자와의 대화 과정에서 주제의 일관성과 연속성을 유지하게 하여, 대화형 봇의 작업의 효율성을 높인다. 대화 주제 정렬은 대화형 봇이 사용자의 질의나 관심사를 더 잘 이해할 수 있도록 하여 응답의 품질을 높일 수 있게 한다. 사용자는 원하는 정보나 서비스를 더욱 효과적으로 제공받을 수 있게 된다. 멀티-턴 대화에서 대화 주제 정렬은 대화형 봇과 사용자 사이에 자연스러운 대화 흐름을 만들어 사용자 경험의 만족도를 높이는 데에 기여한다. 1:1 혹은 그룹 대화에서의 주제 정렬 과정은 대화 윤리를 고려해야 하는데, 그룹 대화의 경우에 다른 배경과 상황에 있는 사용자들 사이에서 대화 윤리를 고려하며 대화를 이끌어나가는 것은 쉬운 일이 아니다.

1. 대화형 봇과 사용자의 1:1 대화에서의 주제 정렬

대화형 봇은 제공하려는 서비스에 따라 그 대화 주제의 범위가 한정되어 있을 수 있다. 기존의 상용 챗봇은 그 개발 및 구축 비용 절감과 응답 정확도 요구사항 만족을 위해 제공하려는 서비스에 따라 대화 주제의 범위가 설정되어 있으며, 대화를 통해 작업-중심의(Task-oriented) 서비스를 제공한다. 최근에 대규모 언어 모델(LLM: Large Language Model)을 이용하는 챗봇의 경우는 그 방대한 학습데이터에 의해 대화 주제의 범위가 넓을 뿐 아니라, 모델에 의해 생성되는 대화의 표현도 다양하다. 기존의 챗봇이 시나리오 기반으로 동작할 경우에 사용자가 대화 주제를 벗어나면 대화 범위를 넘었음을 알리게 되지만, 광범위한 대화 주제에 대해 응답이 가능한 LLM 기반의 대화형 봇은 대화 과정에 따라 그 응답 내용이 달라질 수 있다.

대화형 봇과 사용자 사이의 1:1 대화의 경우에 그림 1에서와 같이 대화형 봇은 대화의 시작점에서 사용자별로 다른 요구에 따라 대화의 중심을 이동하며 대화 주제의 일관성을 유지하려 한다. 대화형 봇과 사용자 사이의 공통 대화 주제의 범위 내에서 그들은 메시지를 교환하며 대화를 이어나갈 가능성이 높다. 대화형 봇은 컴퓨팅 시스템이기 때문에, 각 사용자의 관심사에 따라 다른 대화를 시도할 수 있다. 사용자별 개인정보나 대화 패턴 등에 대한 정보를 이용하는 것이 가능한 경우에 대화 주제 정렬을 위한 대화-턴이 줄어들 수 있다. 이 경우에 대화 윤리를 구현하기 위해 금칙어 등을 정의하여 문장 생성에서 제외하거나 사전 체크리스트를 설정하고 대화 문장을 평가하여 해당 문장을 삭제 또는 그 문장에 대한 대체 표현을 하는 방법이 있을 수 있다.

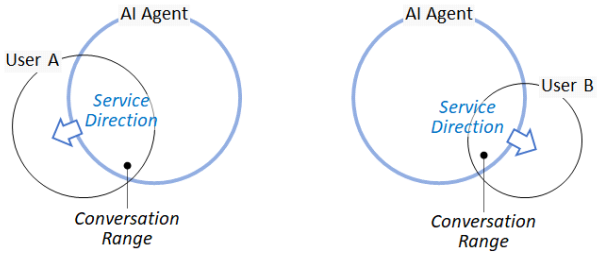

그림 1. 개별 사용자에 대한 대화 주제 정렬
Fig. 1. Alignment of conversation topic for each individual user

2. 대화형 봇과 사용자들의 그룹 대화에서의 주제 정렬

그룹 대화(Group Conversation)는 대화형 봇이 한 명 이상의 사용자와 대화하는 형태이다. 여러 명의 사용자가 대화형 봇과 동시에 상호작용하며 대화를 공유하고 작업을 진행할 수 있다. 그룹 대화는 팀 프로젝트, 회의 일정 결정 등을 위한 그룹 회의에 유용하다. 그룹 대화는 1:1의 개인 대화와는 몇 가지 차이점이 있다. 첫째, 대화형 봇과의 그룹 대화에서는 여러 명의 사용자가 대화에 참여하게 되어 다수의 사용자들이 동시에 질의를 하며 대화를 진행하게 된다. 그렇기 때문에, 개별 사용자들의 요구를 들어주면서도 대화 주제 정렬을 통해 대화 흐름을 유지시킬 필요가 있다. 둘째, 그룹 대화에서 대화형 봇은 각 사용자의 질의와 그 응답을 맥락에 따라 기억하고 이해할 수 있어야 한다. 각

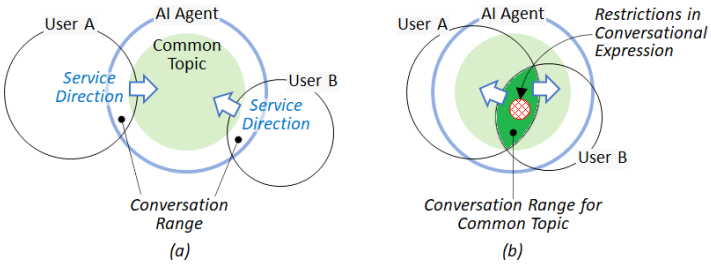

그림 2. 여러사용자들에 대한 대화 주제 정렬
Fig. 2. Alignment of conversation topic for multiple users

사용자의 이전 발언이나 질의뿐 아니라, 다른 사용자와의 대화도 이해하며 적절한 응답을 할 수 있어야 한다. 셋째, 대화형 봇은 사용자들 사이의 질의-응답과 상태를 동기화하며 대화의 흐름을 조율해야 한다. 넷째, 그룹 대화에서는 집단의 의사결정을 결과로 가질 수 있는데, 대화에 참여한 대화 주체 모두가 서로에게 영향을 미칠 수 있기 때문에 대화형 봇은 중립적으로 의사결정의 진행을 도울 수 있어야 한다.

그룹 대화는 1:1 대화와 비교하여 대화 모델 복잡성이 높다. 그렇기 때문에 이 경우에는 기존과 같은 방법으로 금칙어나 사전 체크리스트를 통해 대화 윤리를 유지하는 방법을 적용하는데에 한계가 있을 수 있다. 대화형 봇과 사용자들 사이의 그룹 대화의 경우에 그림 2(a)에서와 같이 대화형 봇은 사용자들의 공통 관심사를 찾아 그들을 대화 주제의 범위로 안내한다. 1:1의 개인 대화에서 대화형 봇이 사용자의 관심사에 따라 대화 주제의 중심을 이동하는 것과 다를 수 있다. 그림 2(a)에서 두 사용자는 일반적으로 공통의 주제를 설정해두고 대화를 잇는다. 그리고, 그들의 대화에서 이용하는 단어들은 그 공통의 주제에 관련된 것들일 가능성이 크다. 그룹 대화에서 대화형 봇은 사용자들이 공통 주제에 대해 메시지를 교환하여 대화의 일관성을 유지하도록 대화 주제를 정렬한다. 그림 2(b)에서와 같이 대화-턴 수가 늘어날수록 두 사용자는 공통 주제의 범위에서 주로 대화하는 주제의 영역이 나타나게 된다.

그룹 대화에서 대화형 봇이 사용자들과 대화를 할 때 윤리적 기준을 충족하면서 대화를 이어나갈 필요가 있다. 대화형 봇은 그룹 대화의 과정에서 개별 사용자들에게 편향적인 응답이나 불쾌감을 주는 표현을 하지 않는 것이 좋다. 예를 들어, 미국인 사용자와 중국인 사용자가 그룹 대화에 참여해 있을 때 대화형 봇이 어느 한 사용자와의 대화에 집중하여 공개된 대화창에 다른 사용자가 불편할 수 있는 표현이나 정보의 전달을 할 수 있다. 예를 들어, “OO인들은 OO해서 싫어”라는 표현은 대화에 참여해 있는 누군가를 불편하게 만들 수 있다. 인간 사용자들로만 구성된 그룹 대화에서 타인에게 공개되면 안되거나 타인이 불편할 수 있는 내용이 있는 경우에 개인 대화를 새로 열어 이용하는 것과 비슷한 맥락이다.

사회적, 문화적인 차이로 발생하는 대화의 불편함까지도 대화형 봇이 발견하여 대응하기 위해 그림 2(b)에서와 같이 대화 표현의 제한된 영역(Restrictions in Conversational Expression)을 설정하여 대화 응답을 생성할 수 있다. 해당 영역에서는 서비스 정책에 따라 가까운 사이의 대화에서도 특정 단어의 이용이나 표현이 금지되거나 대체될 수 있다. 각 사용자의 메시지에 포함된 비속어, 금지어 등은 시스템에서 자동 차단될 수 있다. 1:1 대화에 대한 부분이기는 하지만, 이에 대한 노력은 이미 진행되고 있다. 그룹 대화에 참여한 각 사용자가 불편해하는 대화 표현을 찾아 이를 최소화하여 그룹 대화 내에서 긍정적 대화 분위기를 형성할 수 있도록 윤리적 기준을 설정하고 이에 따라 대화 응답이 생성될 수 있게 하는 방법이 필요하다.

Ⅳ. 그룹 대화를 위한 윤리적 응답 생성 프레임워크 개발 방법 및 실험

1. 윤리적 응답 생성 프레임워크 개발 방법

대화형 봇의 그룹 대화를 위한 윤리적 응답 생성 프레임

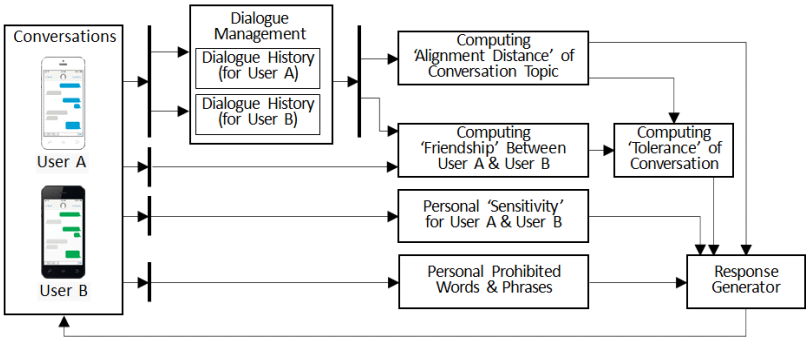

그림 3. 그룹 대화에서의 윤리적 응답 생성을 위한 프레임워크 개념도
Fig. 3. Conceptual diagram of framework for bias inspector using LLM

워크는 챗봇의 윤리적 대화를 위한 기능과 특성을 효과적으로 구현하기 위한 방법의 일 예를 보여줄 수 있다. 개발자는 이러한 프레임워크를 통해 챗봇의 윤리적 대화 설계를 효과적으로 진행할 수 있다.

대화형 봇, 특히 일반적인 AI 챗봇 시스템은 다이어로그관리(Dialogue Management) 기능을 갖는다. 하나의 대화는 몇 개의 질의-응답 셋으로 구성된 다이어로그를 갖는다. 대화 관리를 위해서 시스템은 다이어로그 히스토리(Dialogue History)를 기록한다. 그룹 대화에서는 사용자별 다이어로그 히스토리를 기록하여야 한다. 챗봇이 대화 주제 정렬을 위해서 공통 주제로부터 사용자의 메시지가 얼마나 관련되어 있는지 계산해야 한다. 이를 '정렬 거리'(Alignment Distance)라고 정의한다. 그림 3에서와 같이 다이어로그 히스토리를 참조하여 정렬 거리를 계산하게 된다. 챗봇은 정렬 거리가 줄어들도록 사용자들과 상호작용하며 그들을 공통 대화 주제로 이끈다. 공통 대화 주제로 안내하면서, 사용자 정보를 바탕으로 대화 표현의 '허용'(Tolerance) 값을 계산한다. 대화 응답을 생성할 때 그 허용치에 따라 공손하고 정중한 표현을 생성한다. 사용자들 사이의 관계 혹은 대화 내용을 통해 그들의 친밀성(Friendship) 정보를 알 수 있다면 그 허용치 값을 계산할 때 반영한다. 사용자들이 가까운 친구나 선후배 사이라면 격식이 없는 표현을 서로 주고받을 수 있기 때문이다. 동일한 대화 응답 표현에 대해서도 사용자별로 그 반응 정도가 다를 수 있다. 사용자별 '민감도'(Sensitivity) 정보를 얻을 수 있다면, 대화 응답 생성시에 대화 표현의 허용치 값과 함께 참조하여 챗봇이 대화 응답을 생성할 수 있다. 각 사용자가 개별적을 설정한 금지어가 있으면 반영하여 표현을 수정한다. 그러나, 대화형 봇과 인간 사용자와의 대화 과정에서 이용할 수 있는 단어나 표현의 범위에 대한 것은 공정하게 모두에게 적용되어야 하는 사회 윤리에 대한 것과 구별되어야 함에 유의해야 한다.

2. 실험 및 결과 분석

위에서 제시한 대화형 봇의 윤리적 대화 보조를 위해서는 주제 정렬 정도, 사용자 친밀성 정도의 수치화가 객관적으로 요구된다. 챗봇이 대화 주제의 연관성을 점수화하고 이들 사이의 관계를 분석하는 것이 가능한지 확인하기 위하여, LLM을 이용하여 대화 주제 정렬 거리(특히, 토픽 일치성) 및 사용자 친밀성을 수치적으로 평가하였다. 토픽 일치성에 대한 평가가 중요한 이유는 대화 윤리를 지키기 위해 문장 내 일부 단어의 삭제, 문장의 변형 등이 있을 수 있는데 이러한 과정에서도 AI 챗봇이 목적성 있는 대화를 할 수 있도록 하는 것이 필요하기 때문이다. 사용자 친밀성에 대한 평가도 진행하였는데, 이는 대화 윤리를 지키면서 대상별 유연한 표현의 범위를 설정하는데 활용할 수 있기 때문이다.

2.1. 토픽 일치성 평가 실험

토픽 일치성 평가를 위한 실험 문장 데이터세트는 GPT-4를 활용하여 생성하였으며, "토픽 일치성이 매우 큰 경우", "보통인 경우", "희박한 경우"로 구분하여 아래의 입력 프롬프트를 사용하여 준비하였다. 언어 의존성 여부를 확인하기 위하여, 한국어로 생성한 문장에 대응되는 영어 문장을 구어체 성격에 맞도록 번역하여 실험하였다(표 1).

표 1. 토픽 일치성 평가를 위한 대화 데이터세트(한국어, 영어)
Table 1. Conversation dataset for topic consistency evaluation(Korean, English)

Dataset for topic consistency evaluation	
Korean Conversation	English Conversation
# 대화 그룹 A(토픽 일치성이 매우 큰 경우) 철수: 최근에 넷플릭스에서 본 '마인드헌터'라는 드라마 알아? 길동: 아, 나도 그거 보고 있어. 정말 흥미롭더라. 철수: 그렇지. 특히 그 심리적인 부분이 너무 재밌어. 길동: 맞아, 그런 심리적인 내용을 잘 풀어낸 드라마는 드문 것 같아. 철수: 나도 그래. 그리고 배우들의 연기도 훌륭하더라고. 길동: 아, 정말 그렇지. 특히 주연 배우들의 연기가 뛰어났어. 철수: 그럼 다음 시즌이 기대되지 않아? 길동: 그래. 다음 시즌이 빨리 나왔으면 좋겠어. 철수: 그럼 다음에 다른 재미있는 드라마 추천해줘. 길동: 네, 그럴게. 너도 나한테 추천해줘.	# Conversation Group A(with very high topic consistency) Cheolsu: You know that Netflix drama "Mindhunter" I've been watching recently? Gildong: Oh, I'm watching that too. It's really interesting. Cheolsu: Yeah, especially the psychological aspects are so fascinating. Gildong: Totally, it's rare to see a drama that delves into psychology like that. Cheolsu: I agree. And the acting is just superb. Gildong: Yeah, absolutely. Especially the lead actors. Cheolsu: So you're looking forward to the next season too, huh? Gildong: Yeah, can't wait for it. Cheolsu: Recommend me some other good dramas next time. Gildong: Will do. You do the same.
# 대화 그룹 B(토픽 일치성이 보통인 경우) 철수: 최근에 넷플릭스에서 본 '마인드헌터'라는 드라마 알아? 길동: 아, 나는 요즘 도서관에서 소설책을 빌려서 읽고 있어. 철수: 그래? 어떤 책 읽고 있어? 길동: '해리 포터'라는 책을 읽고 있어. 철수: 나도 '해리 포터' 좋아하는데, 그건 아마도 영화로 봤을거야. 길동: 그래? 나는 책으로 읽는게 더 좋더라. 철수: 나도 책으로 읽어볼까 해. 길동: 그래, 그럼 어떻게 생각했는지 나한테 알려줘. 철수: 그럴게. 너도 '마인드헌터' 볼 생각 있으면 봐봐. 길동: 응, 내가 시간 날 때 한번 봐볼게.	# Conversation Group B(with moderate topic consistency) Cheolsu: You know that Netflix drama "Mindhunter" I've been watching recently? Gildong: Ah, I've been reading novels from the library these days. Cheolsu: Really? What are you reading? Gildong: I'm reading "Harry Potter." Cheolsu: I like "Harry Potter" too, but I probably watched it as a movie. Gildong: Really? I prefer reading the book. Cheolsu: Maybe I'll give the book a try then. Gildong: Do that, and let me know what you think. Cheolsu: Sure, and if you get a chance, check out "Mindhunter." Gildong: Yeah, I'll take a look when I have some time.
# 대화 그룹 C(토픽일치성이 희박한 경우) 철수: 최근에 넷플릭스에서 본 '마인드헌터'라는 드라마 알아? 길동: 아, 나는 온라인 게임에만 관심 있어. 철수: 그래? 그 드라마 심리묘사가 좋아서 흥미진진한데. 길동: 슈팅 게임이 더 재미있지. 철수: 범죄 스릴러 드라마라 보는 내내 집중하게 되더라. 길동: 내 손으로 직접 해야 재미있지. 철수: 주인공들 연기가 대단해서 금방 몰입하게 돼. 길동: 요즘 온라인 게임도 몰입감이 엄청 뛰어나. 철수: 꼭 한 번 그 드라마 봐봐. 재미있을 거야. 길동: 나는 게임이나 할란다.	# Conversation Group C(with sparse topic consistency) Cheolsu: You know that Netflix drama "Mindhunter" I've been watching recently? Gildong: Nah, I'm only interested in online gaming. Cheolsu: Well, the psychological depth in that drama is really captivating. Gildong: I find shooting games more exciting. Cheolsu: It's a crime thriller, keeps you on the edge of your seat the whole time. Gildong: I need to do it myself to find it fun. Cheolsu: The acting is so great, you get absorbed into it easily. Gildong: Online games are really immersive these days too. Cheolsu: You should really watch it at least once. You'll enjoy it. Gildong: Nah, I'm just gonna game.

(사용자)

"철수" 와 "길동" 이 일상생활에 관해 대화를 하고 있습니다.

(A) 2명의 토픽 일치성이 매우 큰 경우,

(B) 2명의 토픽 일치성이 보통인 경우,

(C) 2명의 토픽 일치성이 거의 없는 경우,

각각의 경우에 대해서 10회 분량으로 대화를 주고받는 데이터세트를 생성하시오.

토픽 일치성을 평가하기 위하여 LLM을 사용하였고, 평가 수치의 타당성과 재현성 검증을 위하여 동일 데이터세트에 대해서 10회 반복 실험을 수행하고, 두 개의 LLM 모

델(GPT-4, Claude2) 결과를 비교하여 표와 그래프로 정리하였다(Appendix I, 표 2, 그림 4).

주어진 10개의 문장 데이터세트에 대해서 GPT-4와 Claude2는 비교적 재현성 있는 평가 수치를 제시함을 그래프와 표준편차를 통해서 관찰하였다(그림 4). 토픽 일치성은 주관적인 개념으로 절대적인 수치는 전문가마다 다르겠지만, 고도화된 LLM을 통한 반복적인 실험과 통계적인 처리, 적절한 프롬프트 엔지니어링으로 계량화 가능함을 확인하였다.

2.2. 사용자 친밀성 평가 실험

사용자 친밀성 평가를 위한 실험 문장 데이터세트는 GPT-4를 활용하여 생성하였으며, “친밀성이 매우 큰 경우”, “보통인 경우”, “희박한 경우”로 구분하여 준비하였다. 언어 의존성 여부를 확인하기 위하여, 한국어로 생성한 문장에 대응되는 영어 문장을 구어체 성격에 맞도록 번역하여 실험하였다(표 3).

표 2. GPT-4, Claude2를 이용한 토픽 일치성 평가 결과(10회 반복실험, 100점 척도)
Table 2. Evaluation result of topic consistency using GPT-4 and Claude2(10-times repetition, 100-point scale)

Dataset	LLM	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6	Exp.7	Exp.8	Exp.9	Exp.10	Avg.	Stdev.
# Conversation Group A(with very high topic consistency)	GPT-4 (Korean)	100	100	100	100	100	100	100	100	100	100	100	0.0
	GPT-4 (English)	100	100	100	100	100	100	100	100	100	100	100	0.0
	Claude2 (Korean)	100	100	100	100	100	100	100	100	100	100	100	0.0
	Claude2 (English)	90	90	90	90	90	90	90	90	90	90	90	0.0
# Conversation Group B(with moderate topic consistency)	GPT-4 (Korean)	70	75	60	70	70	75	70	70	70	70	70	3.7
	GPT-4 (English)	75	75	70	70	70	75	75	70	70	70	72	2.3
	Claude2 (Korean)	50	50	60	30	50	30	30	50	50	50	45	9.8
	Claude2 (English)	60	50	50	50	60	40	50	60	50	50	52	5.7
# Conversation Group C(with sparse topic consistency)	GPT-4 (Korean)	20	50	30	20	40	20	30	20	30	30	29	9.0
	GPT-4 (English)	40	50	40	50	40	50	50	30	30	40	42	7.1
	Claude2 (Korean)	0	0	30	0	0	0	0	0	0	0	3	8.6
	Claude2 (English)	20	10	10	20	30	10	10	30	10	10	16	7.6

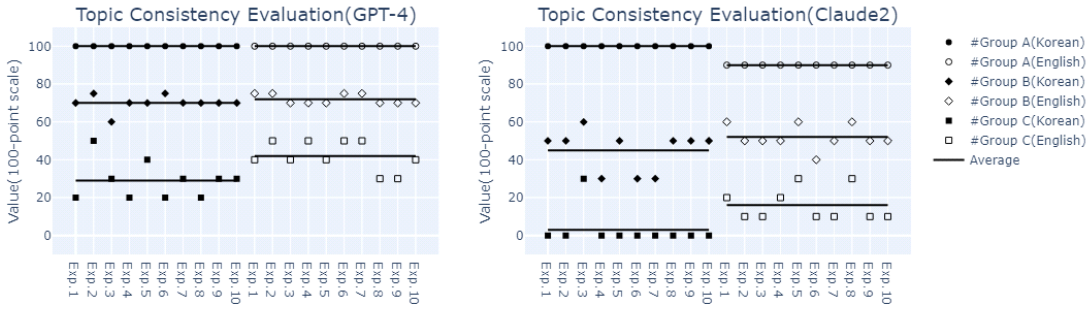

그림 4. GPT-4, Claude2를 이용한 토픽 일치성 평가 결과 그래프(10회 반복실험, 100점 척도)
Fig. 4. Evaluation result graph of topic consistency using GPT-4 and Claude2(10-times repetition, 100-point scale)

표 3. 친밀성 평가를 위한 대화 데이터세트(한국어, 영어)
Table 3. Conversation dataset for friendship evaluation(Korean, English)

Dataset for friendship evaluation	
Korean Conversation	English Conversation
# 대화 그룹 A(친밀성이 매우 큰 경우) 철수: 길동아. 오늘 같이 헬스장 어때? 나 혼자 가기 싫어서 말이야. 길동: 왜 심심해~. 날씨 한번 끝내주네. 어디서 만나? 너네 집이면 좀 멀어서 말이야. 철수: 그럼 우리가 자주 만나던 공원에서. 저녁 6시, 어떠냐? 길동: 좋아~ 오래만에 운동이라 재미있을거야. 너 몸 상태는 좋아졌냐? 철수: 그럭저럭. 이 상태로도 너랑 경주하기에 충분해. 그럼 쯤 있다 보자. 길동: 짜식~ 자신만만하기는~. 나 준비하러 간다. 철수: 러닝화 꼭 챙기고 와~ 지난 번 처럼 까먹지 말고. 길동: 알았써~ 잔소리 하지마. 물도 챙기고 갈께. 오늘은 미친듯이 하자! 철수: 완전 동감! 그럼 좀 있다 보자! 길동: 오키!	# Conversation Group A(with very high friendship) Cheolsu: Hey Gil-dong, wanna hit the gym together today? I hate going alone, man. Gildong: Why so lonely~? The weather is awesome. Where should we meet? Your place is kinda far, you know. Cheolsu: How about that park we usually hang out at? 6 PM, sound good? Gildong: Sweet~ It's been a while since we worked out; it'll be fun. How's your body holding up? Cheolsu: Meh, good enough. I can still race you. See you in a bit then. Gildong: Cocky much~? I'll go get ready. Cheolsu: Don't forget your running shoes this time. Don't space out like last time. Gildong: Gotcha~ Stop nagging. I'll bring some water too. Let's go all out today! Cheolsu: Totally agree! See you in a bit! Gildong: Okey-dokey!
# 대화 그룹 B(친밀성이 보통인 경우) 철수: 안녕하세요, 길동씨. 같이 운동 좀 할 생각 없나요? 제가 혼자 가기가 좀 그렇네요. 길동: 네, 오늘 날씨도 좋으니 같이 할 수 있습니다. 어디서 만날까요? 저희 집 앞이 좀 멀 것 같은데요. 철수: 그럼 시립 공원에서 만나는 게 어떨까요? 저는 6시쯤 가면 좋을 것 같아요. 길동: 네, 괜찮습니다. 꽤 오랜만에 운동이라 좋을 것 같습니다. 철수: 네, 재미있겠습니다. 그럼 나중에 봅시다. 길동: 알겠습니다. 제가 준비하고 갈게요. 철수: 러닝화 챙기시고 오세요. 길동: 네, 알겠습니다. 물도 챙겨갈게요. 재미있게 운동합시다. 철수: 네, 나중에 봅시다. 길동: 알겠습니다.	# Conversation Group B(with moderate friendship) Cheolsu: Hello, Gil-dong. Would you like to exercise together? I find it a bit difficult to go alone. Gildong: Yes, the weather is nice today, so I can join you. Where shall we meet? My place is a bit far. Cheolsu: How about the municipal park? I'd like to go around 6. Gildong: Sure, that works. It's been a while since I exercised, so it'll be nice. Cheolsu: Yes, it'll be fun. See you later then. Gildong: Alright, I'll go get ready. Cheolsu: Please bring your running shoes. Gildong: Sure, I will. I'll also bring some water. Let's have a fun workout. Cheolsu: Yes, see you later. Gildong: Alright.
# 대화 그룹 C(친밀성이 희박한 경우) 철수: 실례합니다. 운동을 같이하실 분 계시나요? 혼자 가기가 별로라서요. 길동: 혹시 제가 같이 운동해도 될까요? 가까운 곳이면 가능합니다. 철수: 불편하지 않으시다면, 시립 공원에서 6시에 뵙는게 어떨까요? 길동: 네, 알겠습니다. 철수: 초면에 응해 주셔서 감사합니다. 길동: 저도 운동하고 싶었습니다. 철수: 네, 러닝화도 챙겨 오십시요. 길동: 알겠습니다. 그리고, 물도 챙겨 가겠습니다. 철수: 고맙습니다. 길동: 네.	# Conversation Group C(with sparse friendship) Cheolsu: Excuse me. Is anyone interested in exercising together? I'd rather not go alone. Gildong: May I join you for the exercise? If it's somewhere close, that would be doable. Cheolsu: If it's not too much trouble, could we meet at the municipal park at 6 PM? Gildong: Yes, understood. Cheolsu: Thank you for agreeing on such short notice. Gildong: I also wanted to exercise. Cheolsu: Please bring your running shoes. Gildong: Understood. I will also bring some water. Cheolsu: Thank you. Gildong: Yes.

(사용자)
"철수" 와 "길동" 이 운동을 주제로 대화를 하고 있습니다.
(A) 2명이 아주 친밀한 경우,
(B) 2명이 보통 친밀한 경우,
(C) 2명이 친밀감이 없는 경우,
각각의 경우에 대해서 10회 분량으로 대화를 주고받는 데이터세트를 생성하시오.

친밀성을 평가하기 위하여 LLM을 사용하였고, 평가 수치의 타당성과 재현성 검증을 위하여 동일 데이터세트에 대해서 10회 반복 실험을 수행하고, 두 개의 LLM 모델(GPT-4, Claude2) 결과를 비교하여 표와 그래프로 정리하였다(Appendix II, 표 4, 그림 5).

주어진 10개의 문장 데이터세트에 대해서 GPT-4와 Claude2는 비교적 재현성 있는 평가 수치를 제시함을 그래

표 4. GPT-4, Claude2를 이용한 친밀성 평가 결과(10회 반복실험, 100점 척도)
Table 4. Evaluation result of friendship using GPT-4 and Claude2(10-times repetition, 100-point scale)

Dataset	LLM	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6	Exp.7	Exp.8	Exp.9	Exp.10	Avg.	Stdev.
# Conversation Group A(with very high topic consistency)	GPT-4 (Korean)	90	90	90	90	90	90	90	90	90	90	90	0.0
	GPT-4 (English)	90	90	85	90	90	90	95	90	90	90	90	2.1
	Claude2 (Korean)	90	90	90	90	90	90	90	90	90	90	90	0.0
	Claude2 (English)	90	90	90	90	90	90	90	90	90	90	90	0.0
# Conversation Group B(with moderate topic consistency)	GPT-4 (Korean)	60	60	60	60	40	60	60	70	60	60	59	6.7
	GPT-4 (English)	70	70	65	70	70	70	70	70	70	70	69.5	1.4
	Claude2 (Korean)	50	30	50	30	50	50	50	50	30	30	42	9.3
	Claude2 (English)	50	50	50	70	50	60	50	50	50	50	53	6.1
# Conversation Group C(with sparse topic consistency)	GPT-4 (Korean)	30	30	20	30	20	30	30	40	30	30	29	5.1
	GPT-4 (English)	40	40	40	40	40	30	35	30	40	40	37.5	3.8
	Claude2 (Korean)	20	10	10	10	10	30	10	10	10	10	13	6.1
	Claude2 (English)	10	10	10	30	20	30	20	20	10	20	18	7.1

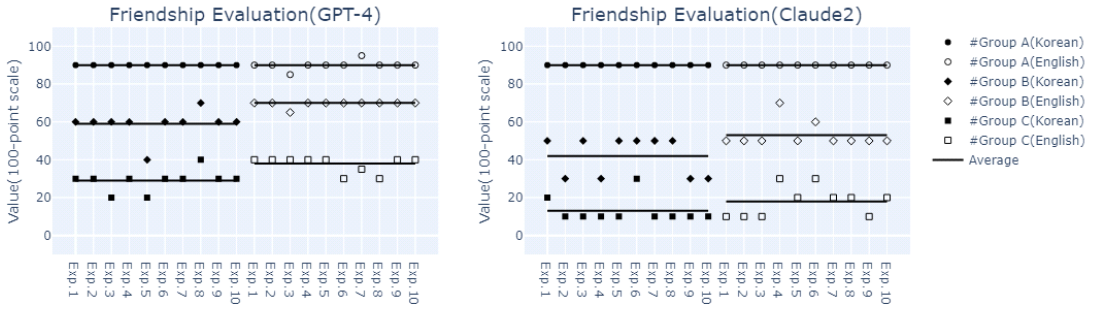

그림 5. GPT-4, Claude2를 이용한 친밀성 평가 결과 그래프(10회 반복실험, 100점 척도)
Fig. 5. Evaluation result graph of friendship using GPT-4 and Claude2(10-times repetition, 100-point scale)

프와 표준편차를 통해서 관찰하였다(그림 5). 사용자 친밀성은 주관적인 개념으로 절대적인 수치는 전문가마다 다르겠지만, 고도화된 LLM을 통한 반복적인 실험과 통계적인 처리, 적절한 프롬프트 엔지니어링으로 계량화 가능함을 확인하였다.

본 실험에서는 1:1 대화를 가정하여 토픽 일치성과 사용자 친밀성의 정량적 수치화 가능성을 실험하였는데, 그룹 대화 상황에서도 유사한 절차로 실험 가능할 것으로 판단되며 추후 도전할 연구과제이다.

V. 결 론

본 논문에서는 대화형 봇이 사용자 그룹 간 대화에 참여할 때 금칙어 필터나 사전 체크리스트 방식이 한계가 있을 수 있음을 지적하였다. 그룹 대화에서 목적 중심 대화를 진행하며 대화 윤리를 준수할 수 있는 시스템 프레임워크 일예를 제시하였다. 그 프레임워크는 사용자 간 토픽 일치성, 친밀성을 반영할 수 있는 방안을 고려하고 있으며, LLM 모델(GPT-4, Claude2)을 기반으로 토픽 일치성 및 친밀

밀성을 정량적인 수치로 계산할 수 있음을 보였다. 대화 시스템의 토픽 일치성에 대한 평가는 대화 윤리를 지키기 위해 문장 내 일부 단어의 삭제, 문장의 변형 등이 있을 수 있는 상황에서 AI 챗봇이 목적성 있는 대화를 하도록 하는데 활용될 수 있으며 사용자 친밀성에 대한 평가는 대화 윤리를 지키면서 대상별 유연한 대화 표현의 범위를 설정하는데 활용될 수 있다. 본 연구는 시스템적 작동이 가능한 대화형 챗봇의 AI 윤리 가이드라인 개발에 기여할 수 있으며, 이를 위해서는 향후 실제 챗봇에 구현하여 다수 사용자 실증으로 윤리적 응답 생성 효과를 검증할 필요가 있다. 또한, 토픽 일치성과 친밀성 계산의 정확도를 높이기 위하여 고도화된 통계적 접근 방법을 도입해 볼 수 있다.

Appendix I. LLM을 이용한 토픽 일치성 평가 결과

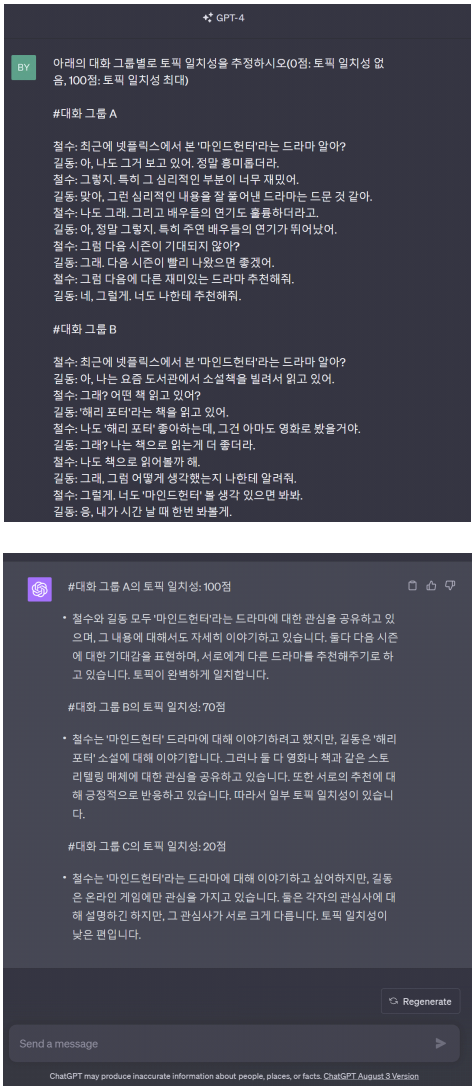

그림 6. GPT-4를 이용한 토픽 일치성 평가
Fig. 6. Evaluation of topic consistency using GPT-4

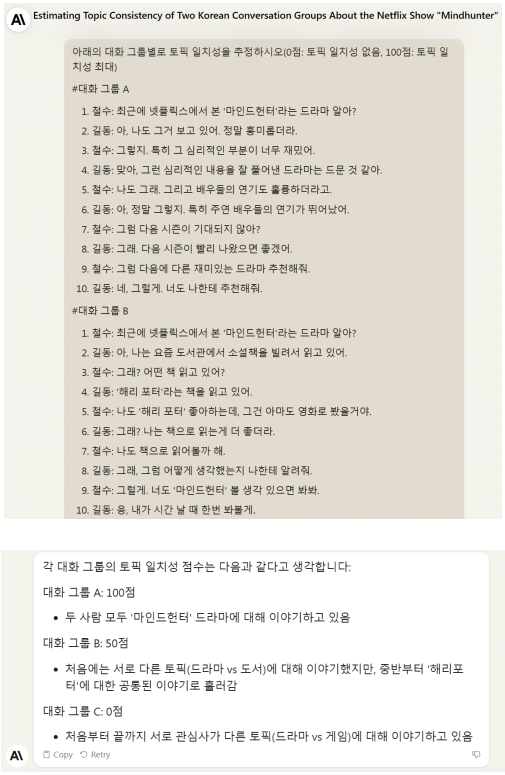

그림 7. Claude2를 이용한 토픽 일치성 평가
Fig. 7. Evaluation of topic consistency using Claude2

Appendix II. LLM을 이용한 친밀성 평가 결과

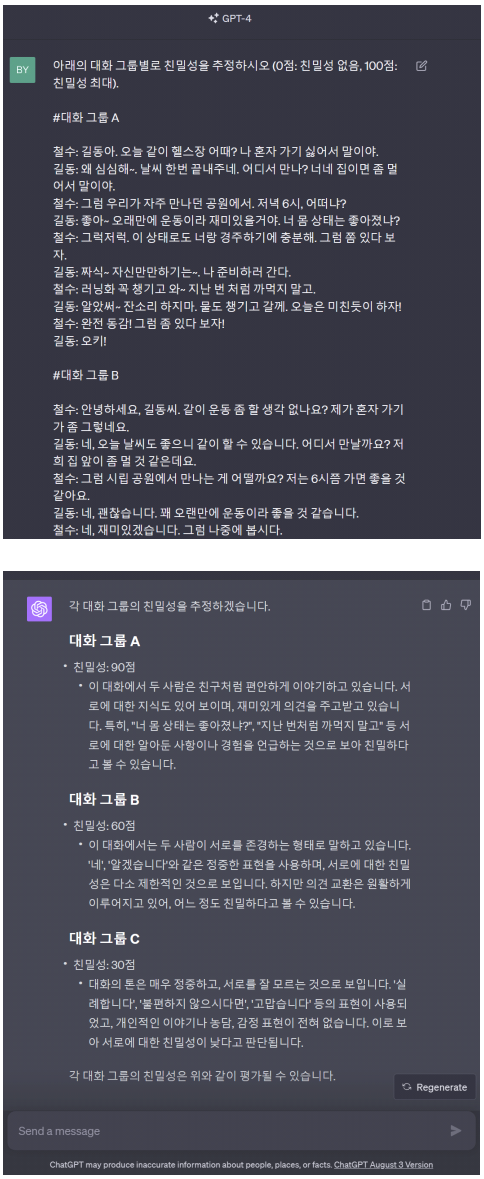

그림 8. GPT-4를 이용한 친밀성 평가
Fig. 8. Evaluation of friendship using GPT-4

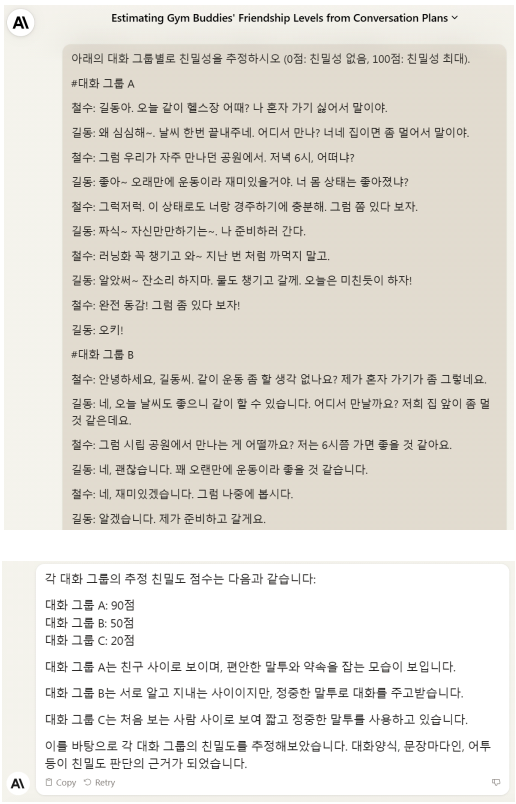

그림 9. Claude2를 이용한 친밀성 평가
Fig. 9. Evaluation of friendship using Claude2

참 고 문 헌 (References)

[1] J. Bang, “Artificial Intelligence Technology for Expanding Metaverse Services”, KICS: Information and Communications Magazine, Vol. 39, No. 2, pp. 64-73, Jan. 2022. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11032345 (accessed July 1, 2023)

[2] V.W. Zue and J.R. Glass, “Conversational Interfaces: Advances and Challenges”, Proceedings of the IEEE, vol. 88, no. 8, pp. 1166-1180, Aug. 2020.
doi: https://doi.org/10.1109/5.880078 (accessed July 1, 2023)

[3] E. Ruane, A. Birhane, A. Ventresque, “Conversational AI: Social and Ethical Considerations”, 27th AIAI Irish Conf. on Artificial Intelligence and Cognitive Science, Galway, Ireland, Dec. 5, 2019. https://api.semanticscholar.org/CorpusID:211838099 (accessed July 1, 2023)

[4] G. Murtarelli, A. Gregory, and S. Romenti, “A Conversation-based Perspective for Shaping Ethical Human‐Machine Interactions: The Particular Challenge of Chatbots”, Journal of Business Research, vol. 129, pp. 927-935, May 2021.
doi: https://doi.org/10.1016/j.jbusres.2020.09.018 (accessed July 1, 2023)

[5] C. Kooli, “Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions”, Sustainability, Vol. 15, No. 7, March 23, 2023.
doi: https://doi.org/10.3390/su15075614 (accessed July 1, 2023)

[6] N. Wagner, M. Kraus, T.Tonn, and W. Minker, “Comparing Moderation Strategies in Group Chats with Multi-User Chatbots”, Proceedings of 4th Conf. on Conversational User Interfaces (CUI), Article No. 35, July 2022.
doi: https://doi.org/10.1145/3543829.3544527 (accessed July 1, 2023)

[7] A. P. Chaves and M. A. Gerosa, “Single or Multiple Conversational Agents? An Interactional Coherence Comparison”, Proceedings of CHI Conf. on Human Factors in Computing Systems (CHI), Article No. 191, April 2018.
doi: https://doi.org/10.1145/3173574.3173765 (accessed July 1, 2023)

[8] C.-C. Lin, A. Y. Q. Huang, and S. J. H. Yang, “A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999‐2022)”, Sustainability, Vol. 15, No. 5, Feb. 22, 2023.
doi: https://doi.org/10.3390/su15054012 (accessed July 1, 2023)

[9] I. K. F. Haugeland, A. Følstad, C. Taylor, and C. A. Bjørkli, “Understanding The User Experience of Customer Service Chatbots: An Experimental Study of Chatbot Interaction Design”, International Journal of Human-Computer Studies, Vol. 161, May 2022.
doi: https://doi.org/10.1016/j.ijhcs.2022.102788 (accessed July 1, 2023)

[10] Q. N. Nguyen, A. Sidorova, and R. Torres, “User Interactions With Chatbot Interfaces vs. Menu-based Interfaces: An Empirical Study”, Computers in Human Behavior, Vol. 128, March 2022.
doi: https://doi.org/10.1016/j.chb.2021.107093 (accessed July 1, 2023)

[11] H. Yun, A. Ham, J. Kim, T. Kim, J. Kim, H. Lee, J. Park, and J. Jang, “Chatbot with Touch and Graphics: An Interaction of Users for Emotional Expression and Turn-taking”, Proceedings of 2nd Conf. on Conversational User Interfaces, Article No. 42, July 2020.
doi: https://doi.org/10.1145/3405755.3406147 (accessed July 1, 2023)

[12] G. R. S. Silva and E. D. Canedo, “Towards User-Centric Guidelines for Chatbot Conversational Design”, arXiv:2301.06474v1, Jan 16, 2023.
doi: https://doi.org/10.48550/arXiv.2301.06474 (accessed July 1, 2023)

[13] A. Caliskan, J. J. Bryson, A. Narayanan, “Semantics Derived Automatically From Language Corpora Contain Human-like Biases”, Science, Vol 356, No. 6334, pp. 183-186, Apr 14, 2017. https://www.science.org/doi/10.1126/science.aal4230 (accessed on Aug. 18, 2023)

저 자 소 개

방 준 성

- 2013년 : 광주과학기술원(GIST) 정보통신공학과 공학박사
- 2013년 ~ 현재 : 한국전자통신연구원(ETRI) 디지털융합연구소 책임연구원
- 2016년 ~ 현재 : 과학기술연합대학원대학교(UST) 인공지능학과 교수
- 2022년 ~ 현재 : 한양대학교 과학기술윤리법정책센터 기술전문위원
- 2023년 ~ 현재 : ㈜와이매틱스 대표이사
- ORCID : https://orcid.org/0000-0003-1446-7755
- 주관심분야 : Contextual Computing, AI Ethics, Conversational Bot, Computer Vision, XR

저 자 소 개

이 병 탁

- 2000년 : 한국과학기술원(KAIST) 전기전자공학 공학박사
- 1999년 ~ 2002년 : LG전자 정보통신 책임연구원
- 2003년 ~ 현재 : 한국전자통신연구원(ETRI) 호남권연구센터 책임연구원
- 2023년 ~ 현재 : ㈜와이매틱스 기술이사
- ORCID : https://orcid.org/0000-0003-1372-4561
- 주관심분야 : AI Ethics, multimodal LLM, AIoT, Digital Twin

박 판 근

- 2011년 : 스웨덴왕립공과대학(Royal Institute of Technology) 전자공학 공학박사
- 2011년 ~ 2013년 : University of California, Berkeley. 박사후연구원
- 2013년 ~ 2015년 : 한국전자통신연구원(ETRI) 선임연구원
- 2015년 ~ 2016년 : 경상대학교 조교수
- 2016년 ~ 현재 : 충남대학교 부교수
- ORCID : https://orcid.org/0000-0003-3744-4476
- 주관심분야 : Graph neural networks, Networked robots, Wireless network