



## 청소년의 또래집단 규범이 비윤리적 행동 중단의도, 인공지능 챗봇에 대한 친밀감과 상호작용 만족도에 미치는 영향

메시지 프레임의 조절효과

**박남기** 연세대학교 언론홍보영상학부 교수  
**피연진** 연세대학교 일반대학원 언론홍보영상학과 박사과정 학생  
**이휘란** 연세대학교 일반대학원 언론홍보영상학과 석사과정 학생  
**이승희** 연세대학교 일반대학원 언론홍보영상학과 박사과정 학생  
**신제인** 연세대학교 일반대학원 언론홍보영상학과 석사과정 학생

## Effects of Adolescents' Peer Group Norms on Intention to Stop Unethical Behaviors, Closeness to and Satisfaction with AI Chatbots\*

The Moderating Role of Message Framing

**Namkee Park\*\***

(Professor, Department of Communication, Yonsei University)

**Yunjin Pi**

(Doctoral Student, Department of Communication, Yonsei University)

**Hwiran Lee**

(Master's Student, Department of Communication, Yonsei University)

**Seunghye Lee**

(Doctoral Student, Department of Communication, Yonsei University)

**Jane Shin**

(Master's Student, Department of Communication, Yonsei University)

Given that the interactions between humans and artificial intelligence (AI), particularly the use of AI chatbots by adolescents, are increasing, it is necessary to discuss related ethical issues. For instance, adolescents are highly likely to be exposed to socially undesirable language use online, and AI chatbots can easily learn users' language use. Thus, adolescents' use of AI chatbots may lead to ethically problematic issues. However, empirical studies that have investigated ethical issues in the

\* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of the Republic of Korea(본 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임)(NRF-2021S1A5A2A0107008612).

\*\* npark@yonsei.ac.kr

context of adolescents and AI chatbot interactions are limited. The present study examined how the interactions between adolescents and AI chatbots were conducted when they were in unethical situations, and how the persuasive effects of AI chatbots worked in such situations. Specifically, the present study examined how adolescents' peer group norms and message framing influenced their intention to stop unethical behaviors, their sense of closeness to AI chatbots, and their satisfaction with AI chatbots. A 3 (peer group norms toward an unethical behavior: agree vs. disagree vs. no specific opinion) x 2 (message framing: positive vs. negative) factorial design was utilized in two separate online experiments with South Korean adolescents aged 13-18. The scenarios presented to the participants involved unethical situations such as online course auto-enrollment or unauthorized absenteeism, thereby simulating real-life experiences that adolescents may frequently encounter. Results showed that closeness to an AI chatbot was highest when the AI chatbot had an opposite opinion to peer group norms. However, intention to stop unethical behaviors and satisfaction with the AI chatbot based on peer group norms did not present significant differences. In terms of message framing, negative framing in the AI chatbot's communication led to higher levels of both chatbot closeness and interaction satisfaction with the AI chatbot. In addition, the interaction effect between peer group norms toward the unethical behaviors and AI chatbot message framing was not significant. The importance of the present study lies in its empirical examination of how AI chatbots can positively affect adolescents' moral reasoning and attitude development from an ethical standpoint. By demonstrating the nuanced ways in which AI chatbots can shape the ethical landscapes of young minds, the study contributes to understanding of adolescents-AI chatbot interaction. It also suggests the potential of AI chatbots in ethical education, providing vital clues on how AI chatbots can effectively enhance ethical values and attitudes among adolescents. The present study also offers practical implications in the area of educational AI chatbot, offering possible ways to employ AI chatbots for promoting ethical education.

**Keywords:** AI Chatbot, AI Ethics, Peer Group Norms, Message Framing

## 국문초록

본 연구는 청소년의 또래집단 규범과 인공지능 챗봇 메시지 프레임이 비윤리적 행동 중단의도, 인공지능 챗봇에 대한 친밀감, 챗봇과의 상호작용에 대한 만족도에 미치는 영향을 분석함으로써 비윤리적 상황에서 인공지능 챗봇의 효과성을 탐구했다. 이를 위해 3(비윤리적 행동에 대한 또래집단 규범: 찬성 vs. 반대 vs. 특별한 의견 없음) x 2(메시지 프레임: 긍정 vs. 부정)의 개체 간 요인설계를 바탕으로 만 13-18세 청소년을 대상으로 두 가지 온라인 실험을 수행하였다. 각 실험의 참여자는 여섯 개의 실험조건 중 하나에 무작위로 배치되었다. 분석 결과, 비윤리적 행동에 대한 또래집단 규범이 반대일 경우, 챗봇에 대한 친밀감이 높게 나타났으나 비윤리적 행동 중단의도와 챗봇과의 상호작용에 대한 만족도는 유의미한 차이를 보이지 않았다. 메시지 프레임의 경우, 부정 프레임 메시지에서 챗봇에 대한 친밀감과 상호작용에 대한 만족도가 높았다. 한편, 비윤리적 행동에 대한 또래집단 규범과 인공지능 챗봇 메시지 프레임의 상호작용 효과

는 유의미하지 않은 것으로 나타났다. 본 연구는 윤리적 관점에서 청소년의 도덕적 의사결정과 태도 형성에 인공지능 챗봇이 긍정적인 영향을 미칠 수 있음을 실증적으로 살펴보았다는 점에서 의의가 있다. 또한, 윤리적 가치와 태도 교육에 있어 인공지능 챗봇의 활용 가능성을 시사한다.

핵심어 : 인공지능 챗봇, 인공지능 윤리, 또래집단 규범, 메시지 프레이밍

# 1. 서론

인공지능(artificial intelligence: AI) 챗봇이 청소년에게 미치는 영향에 대한 논의는 아직 충분하지 않지만, 비윤리적 사용에 대한 우려가 지속적으로 제기되고 있다. 실제로 우리나라에서 발생한 ‘이루다 사태’와 같이, 최근 해외 청소년들 사이에서는 대화형 챗봇을 조롱하거나 정신적으로 학대하는 행위가 크게 유행하고 있는 것으로 전해졌다(Sung, 2023). 또한 비전문적인 AI 챗봇이 우울증을 앓는 청소년들에게 해로울 수 있으며, 심지어 자살 충동을 유발할 수 있다는 조사 결과는 청소년들의 무분별한 인공지능 챗봇 사용에 대한 심각한 경고로 여겨진다(Norton & Crowson, 2024).

물론 인공지능 챗봇이 청소년의 학습과 성장에 기여할 수 있다는 근거는 분명 존재한다. 특히 대규모 언어 모델(Large Language Model: LLM)을 기반으로 하는 인공지능 챗봇은 언어 학습, 정신건강 및 심리 상담 분야에서 활발히 사용되고 있으며, 이러한 인공지능 챗봇의 효과는 다수의 선행연구를 통해 입증되었다(Grove, 2021; Han, 2021; Lin & Chang, 2020; Zhang, Shan, Lee, Che, & Kim, 2023). 인공지능 챗봇의 장점 중 하나는 청소년들의 관심사에 맞추어 그들의 눈높이에서 대화할 수 있다는 것이다(이지예, 2021). 본 연구는 인공지능 챗봇이 청소년들의 올바른 윤리 의식 확립에도 긍정적인 영향을 미칠 수 있다고 가정하고, 인공지능 챗봇의 긍정적 활용 가능성과 효과적인 메시지 전략을 살펴보고자 한다.

구체적으로 본 연구는 집단 규범에 따른 설득효과와 메시지 프레이밍의 조절효과를 살펴보고자 한다. 최근 인공지능 챗봇이 청소년에게 무방비하게 노출되면서 청소년기 이용자들의 태도 형성에 부정적인 영향을 미칠 가능성에 대한 우려가 이어지고 있다(김예슬, 2024). 더욱이 윤리적 판단을 요구하는 상황에서 긍정 메시지 프레이밍과 부정 메시지 프레이밍의 효과가 상이하게 나타날 수 있다는 결과가 보고됨에 따라(Erlandsson, Nilsson, & Västfjäll, 2018) 비윤리적 상황에서 청소년을 사회적으로 바람직한 방향으로 설득할 수 있는 인공지능 커뮤니케이션 전략을 모색할 필요가 있다. 이를 위해 본 연구는 연구 참여자들에게 또래집단이 비윤리적 상황에 대해 어떠한 입장(찬성 vs. 반대 vs. 의견없음)을 취하고 있는지에 대한 시나리오를 제시하고, 인공지능 챗봇이 전달하는 메시지의 프레이밍(긍정 vs. 부정)이 설득 효과에 어떠한 차이를 가져오는지 살펴보고자 한다.

본 연구는 지금까지 제한적이었던 인공지능-청소년 상호작용에 관한 학술적 논의를 확장시킬 것으로 기대된다. 인공지능의 윤리적 이슈는 지속적으로 논의되어 왔음에도 불구하고, 인공지능과 청소년의 상호작용과 관련된 연구는 상대적으로 제한적이었다. 기존 연구들(Ahn, Zhang,

Park, & Lee, 2020; Lin & Chang, 2020; Zhang et al., 2023)은 주로 인공지능 챗봇을 활용한 청소년의 학습 효과와 정신건강에 미치는 영향에 초점을 맞추었다. 이와 달리, 본 연구는 청소년이 비윤리적 상황에 직면했을 때 어떤 결정을 내리는 것이 올바른지를 판단하는 과정에서 인공지능 챗봇의 활용 가능성을 탐색한다. 본 연구가 청소년의 또래집단 규범(peer group norm)과 챗봇 메시지 프레이밍 유형을 통해 제시하는 효과적인 이용자 상호작용 전략은 향후 인공지능 챗봇을 설계하는 과정에서 참고할 수 있는 중요한 지침을 제공할 수 있을 것으로 기대된다.

## 2. 이론적 논의

### 1) 청소년과 인공지능의 상호작용: 윤리적 이슈

인공지능의 학습능력이 급속히 발전함에 따라 인간-인공지능의 상호작용이 크게 증가하고 있다. 인공지능 기술의 활용은 성인뿐만 아니라 청소년에게까지 확대되어 상담과 학습을 위한 다양한 인공지능 챗봇이 등장하고 있다. 예컨대, 호주에서 개발된 챗봇 ‘애쉬(Ash)’는 청소년의 정신건강에 초점(Grove, 2021)을 맞춰 학교에서 스트레스, 부적응, 우울증 등과 같은 심리적 문제를 겪는 청소년들과 자연스럽게 소통하며 자살이나 약물, 알코올 의존과 같은 위험 상황을 예방하는 역할을 수행한다(이지예, 2021). 국내에서는 여성가족부와 한국청소년상담복지개발원이 협력하여 ‘솔로봇(솔루션 챗봇)’ 서비스를 도입했으며, 이를 통해 청소년들은 대인관계, 진로, 학업, 외모 등의 고민을 온라인상에서 캐릭터와의 교감을 통해 자율적으로 해결할 수 있게 되었다(구무서, 2022). 이처럼 인공지능 챗봇은 청소년의 성장과 정신건강에도 중요한 영향을 미치고 있다.

한편, 인공지능은 이용자와의 상호작용을 통해 딥러닝을 수행하므로 윤리적인 학습을 위해서는 인공지능을 이용하는 사용자의 윤리 의식이 중요하다. 힐, 포드, 그리고 파레라스(Hill, Ford, & Farreras, 2015)는 사람들은 인간과의 상호작용보다 인간-인공지능 챗봇 간의 상호작용에서 비윤리적 언어를 더 많이 사용하는 경향이 있다고 밝혔다. 예를 들어, 인공지능 챗봇과의 대화에서는 대인관계에서처럼 인상관리에 신경을 쓰지 않아 비윤리적 언어 사용이 증가할 수 있다(박남기, 2020). 한편, 인공지능 챗봇 이용자들은 비윤리적인 언어 사용 시, 챗봇이 호응하기보다는 계도하거나 경고하는 등 올바른 방향으로 이끌어줄 것을 기대하기도 한다(박남기, 2020). 이러한 이용자의 기대는 인공지능을 단순한 기계가 아닌 사회적 행위자로 인식하는 “컴퓨터는 사회적 행위자(Computers Are Social Actors: CASA)” 패러다임(Reeves & Nass,

1996)의 주장과도 연결된다. 즉, 인공지능 챗봇은 윤리적 기준을 갖춘 도덕적 행위자로 인식되는 경향이 있다는 것이다(박남기, 2020).

청소년은 또래집단으로부터 영향을 많이 받으며, 온라인 환경에서 욕설과 같은 부정적인 언어 표현에 쉽게 노출될 수 있다. 실제로 양명희와 강희숙(2011)은 청소년의 94.6%가 온라인에서 학습이나 모방을 통해 욕설을 사용하고 있다고 보고했다. 인공지능이 이용자의 언어 데이터를 수집하고 이를 학습한다는 특성(Kooli, 2023)을 고려할 때, 청소년들이 부정적이거나 욕설을 포함한 언어를 사용할 경우 이를 학습할 가능성이 높다. 따라서 청소년들은 인공지능 챗봇을 활용함에 있어 자신의 언어 선택과 행동에 대한 윤리적인 의식을 가질 필요가 있다.

최근 청소년을 대상으로 한 인공지능 챗봇의 활용이 증가함에 따라 청소년 이용자의 윤리적 역할이 더욱 중요해지고 있다. 장 등(Zhang et al., 2023)은 청소년들은 성인에 비해 챗봇에 대한 흥미, 신기성 등 이용 동기가 강해 챗봇을 통한 학습효과가 더 크게 나타나는 경향이 있다고 밝혔다. 또한 영어 학습을 위한 인공지능 챗봇 사용은 학생들의 영어 학습에 대한 태도를 긍정적으로 변화시키기도 했다(Han, 2021). 린과 창(Lin & Chang, 2020)은 작문 교육에 챗봇을 사용할 때 학습 효과가 높아지는 것을 확인했으며, 참여자 중 75%가 챗봇과의 상호작용이 작문의 정확성과 질을 평가하는 데 도움이 되었다고 응답했다.

학습 환경뿐만 아니라 청소년의 정신건강 측면에서도 인공지능 챗봇의 영향에 관한 연구가 활발히 진행되고 있다(Ahn et al., 2020; Mariamo, Temcheff, Léger, Senecal, & Lau, 2021; Wang et al., 2022). 예컨대, 비윤리적 행동과 관련하여 청소년들은 성, 약물, 음주와 같은 잠재적으로 위험한 행동에 관한 정보를 얻기 위해 인공지능 챗봇을 활용하는 경향이 있으며, 이들은 챗봇이 전달하는 정보를 검색 엔진이 제공하는 정보보다 긍정적으로 평가했다(Crutzen, Peters, Portugal, Fisser, & Grolleman, 2011).

현재 인간-인공지능 간의 상호작용, 특히 청소년들의 인공지능 챗봇 활용이 증가하고 있는 상황에서, 이와 관련한 윤리적 이슈에 대한 논의 또한 필요하다. 청소년은 온라인 환경에서 부정적인 언어표현에 쉽게 노출될 수 있으며, 인공지능 챗봇은 이러한 사용자의 언어를 학습하므로 청소년의 인공지능 챗봇 사용은 윤리적 문제를 야기할 수 있다. 그러나 현재까지 청소년과 인공지능 챗봇의 윤리적 이슈에 관한 경험적 연구는 상대적으로 부족한 실정이다. 이에 본 연구는 청소년들이 비윤리적인 상황에 처했을 때 인공지능 챗봇과의 상호작용이 어떻게 이루어지는지, 그리고 그러한 상황에서 챗봇의 설득효과가 어떻게 작용하는지에 대해 탐구하고자 한다.

## 2) 또래집단 규범

청소년은 사회적 상호작용에 대한 욕구 충족을 위해 비슷한 나이의 친구들을 찾아 또래집단을 형성한다. 청소년기에는 친구와의 관계가 매우 중요하므로 청소년들은 또래집단에서의 수용 여부에 민감하게 반응한다(Steinberg & Morris, 2001). 이로 인해 청소년들은 자신을 둘러싼 또래집단에 동조하거나 동화되는 경향이 높다(Cairns & Cairns, 1994). 사회규범이론(social norm theory)에 따르면 또래집단은 청소년들에게 가장 중요한 준거집단이며, 또래들의 행동은 청소년들의 행동 기준이 된다(Berkowitz, 2005). 특히 또래집단을 준거집단으로 인식하는 청소년들의 경우, 주어진 상황에서 또래들이 어떤 행동을 하는지를 인지하는 것만으로도 윤리적 판단이나 행동에 영향을 받을 가능성이 높다. 이는 집단 규범(group norm)의 개념으로 설명될 수 있다. 집단 규범은 특정 집단을 특징짓고 다른 집단들과의 차별성을 나타내는 태도 및 행동 규칙 양식을 의미한다(Hogg & Reid, 2006). 지금까지 청소년의 또래집단 규범에 관한 연구는 집단 규범과 청소년의 행동의도 및 실제 행동 사이의 관계를 확인하고자 했다. 예를 들어, 청소년의 건강행동 관련 연구(Lally, Bartle, & Wardle, 2011)는 또래의 식습관 규범, 청소년의 건강행동 의도와 실제 건강행동이 모두 정적인 관계를 갖고 있음을 밝혔다.

청소년의 사회적으로 바람직하지 않은 행동과 또래집단 규범에 관한 연구도 다수 진행되었다. 청소년의 약물남용(Santor, Messervey, & Kusumakar, 2000), 온라인에서의 모욕적인 대화(Sasson & Mesch, 2014), 온라인 성적 행동(Baumgartner, Valkenburg, & Peter, 2011)과 또래집단 규범의 관계를 밝힌 연구가 존재한다. 가령, 또래집단이 위험한 행동에 참여한다고 인식한 청소년은 위험한 행동에 참여하는 경향이 높았다(Brown, Clasen, & Eicher, 1986). 특히 청소년의 온라인에서의 문제 행동은 오프라인 문제 행동과 유사한 방식으로 또래집단의 승인에 영향을 받았다(Baumgartner et al., 2011). 또한 또래집단의 음주관련 소셜미디어 콘텐츠에 노출된 청소년은 음주를 선호하는 또래집단 규범을 바탕으로 음주 행동을 시작할 가능성이 높다고 밝혀졌다(Nesi, Rothenberg, Hussong, & Jackson, 2017).

또래집단 규범이 청소년의 비윤리적 행동 중단의도(intention to stop)에 직접적인 영향을 미치는지에 관한 연구는 제한적이긴 하지만 관련 선행연구는 집단 규범의 영향력을 확인했다. 예를 들어, 김태경과 윤성준(2022)은 청소년의 음주에 대한 또래규범과 음주 절제의도 사이의 관련성을 살펴보고, 그 결과 부정적 또래규범이 음주 행위를 금지 혹은 절제하려는 의도를 갖게 한다고 밝혔다. 보건복지부의 청소년 흡연 예방 공익광고인 <노담 캠페인>의 설득 효과를 살펴본 이다은과 김현석(2023)의 연구에 따르면, 금연 광고를 많이 시청할수록 또래규범이 점화되어 비흡연 의도에 긍정적인 영향을 미쳤고, 또래집단 소속감의 조절효과 또한 확인되었다. 이는

청소년들이 또래집단의 영향으로 인해 또래집단에서 승인된 행위에 자신을 맞추며 기존 행위를 변화시킬 수 있음을 보여주었다.

이와 유사하게 또래집단 규범에 따라 인공지능 챗봇에 대한 친밀감(intimacy)과 상호작용에 대한 만족감(chatbot satisfaction)도 달라질 수 있다. 사람들은 자신과 유사한 특성을 가진 사람이나 객체를 더 긍정적으로 평가하는 경향이 있다(Byrne, 1997). 또한, 사람들은 집단 규범을 공유하여 집단 내 차이를 최소화할 때 안정감을 얻으며(Blader & Tyler, 2009), 이에 동조하는 대상을 내집단으로 인식하고 편애하는 경향을 보인다(Hertel & Kerr, 2001). 특히, 청소년은 비슷한 나이의 또래집단으로부터 승인을 받기 위해 자신이 속한 집단의 규범을 따르려는 동조 성향을 보인다. 이 같은 또래집단 규범에 대한 인식은 사람이 아닌 대상의 호감도를 좌우할 수 있으며, 결과적으로 인공지능 챗봇에 대한 친밀감과 상호작용에 대한 만족감에도 영향을 미칠 수 있다.

CASA 패러다임(Reeves & Nass, 1996)에 따르면, 사람들은 기계에도 사회적 현존감을 느끼고 같은 인간과 상호작용하듯 기계와 상호작용한다. 이때 기계가 자신과 유사한 특성을 보일 때 더 긍정적인 태도를 갖는다(Reeves & Nass, 1996). 이와 관련하여 인간이 아닌 대상에게도 인간의 특징이나 가치를 부여하는 의인화(anthropomorphism) 정도가 인공지능에 대한 친밀감과 만족도에 영향을 미칠 수 있다(Xie, Zhu, Zhou, & Liang, 2023). 예를 들어, 챗봇의 의인화는 이용자와 챗봇 간의 자아일치성을 증가시켜 이용자의 태도와 행동에 긍정적인 영향을 미칠 수 있다(Zogaj, Mähner, Yang, & Tscheulin, 2023). 이는 이용자가 자신과 유사하다고 느끼는 인공지능 챗봇과의 상호작용에서 더 큰 친밀감과 만족감을 느낄 수 있음을 의미한다. 이러한 효과는 챗봇이 또래집단 규범과 일치하는 의견을 제시할 때 더 크게 나타날 수 있다.

결국 인공지능 챗봇이 비윤리적 행동을 반대하거나 제지하는 정도가 또래집단의 규범과 일치하는 것으로 지각된다면, 챗봇을 내집단으로 인식함으로써 챗봇에 대한 친밀감과 만족도가 높아질 수 있다는 추론이 가능하다. 서지민, 이지연, 그리고 박명자(2011)가 청소년의 동조성향은 특정 대상에 대한 긍정적인 태도와 연관이 있다고 밝힌 점은 이러한 추론을 뒷받침 할 수 있다. 또한, 청소년의 집단 규범과 관련하여 또래집단의 설득효과를 경험적으로 탐구한 연구도 있다. 초등학교 고학년을 대상으로 진행된 권보경(2022)의 연구에서는 설득 메시지의 발신자가 또래인 경우 행동의도에 더 큰 영향을 미치는 것으로 보고되었다.

이상의 선행연구를 종합해 볼 때, 또래집단 규범은 청소년의 행동 의도와 실제 행동에 유의미한 영향을 미칠 수 있음을 알 수 있다. 이와 함께, 인공지능 챗봇과의 상호작용 과정에서 또래집단 규범이 챗봇과의 상호작용을 통한 친밀감과 만족도에 긍정적인 영향을 미칠 것으로 추측할



수 있다. 따라서 비윤리적 행동에 대한 또래집단 규범이 긍정적인지 혹은 부정적인지에 따라 비윤리적 행동 중단도, 인공지능 챗봇에 대한 친밀감, 챗봇과의 상호작용에 대한 만족도에 차이가 나타날 것으로 예측할 수 있다.

**연구기설 1.** 인공지능 챗봇이 의견을 제시한 비윤리적 행동에 대한 또래집단의 규범 조건(찬성/반대/특별한 의견 없음)에 따라 청소년들의 비윤리적 행동 중단도에 차이가 있을 것이다.

**연구문제 1.** 인공지능 챗봇이 의견을 제시한 비윤리적 행동에 대한 또래집단 규범(찬성/반대/특별한 의견 없음)에 따라 청소년들의 a) 챗봇에 대한 친밀감, b) 챗봇과의 상호작용에 대한 만족도에 차이가 있는가?

### 3) 메시지 프레이밍의 조절효과

메시지 프레이밍은 긍정 또는 부정적으로 설득 메시지를 제시하는 방식이다(Kahneman & Tversky, 1979). 메시지가 설득하는 내용을 이용자가 선택할 경우 얻을 수 있는 긍정적인 측면을 강조하는 메시지를 이익(gain) 혹은 긍정(positive) 프레이밍, 이를 선택하지 않을 경우 경험할 수 있는 부정적인 측면을 강조하는 메시지를 손실(loss) 혹은 부정(negative) 프레이밍이라 한다(Kühberger, 1998; Levin, Schneider, & Gaeth, 1998). 메시지 프레이밍은 동일한 상황 혹은 의사결정에 관한 내용이라 하더라도 선택에 따른 이익과 손실 중 어떤 측면을 강조하느냐에 따라 설득효과가 달라져 수용자의 행동에 미치는 영향에 차이를 보일 수 있다(안혜영·이승조, 2020; Levin et al., 1998).

선행연구는 긍정/부정 메시지 프레이밍 효과의 차이를 검증해왔다. 하나의 메시지 프레이밍이 모든 상황에서 절대적으로 더 높은 효과를 보인다는 연구 결과는 드물며, 연구주제 혹은 개인적 특성 등에 따라 상이한 결과를 보였다(김성재·유명길, 2010). 오키프와 쟈슨(O'Keefe & Jensen, 2009)은 메시지 프레이밍의 효과는 다양한 외부 요인을 고려해야 할 필요가 있다고 강조했다. 메시지 프레이밍의 효과는 메시지 주제, 개인적 요인(성별, 관여도, 교육 수준 등), 콘텐츠 유형과 같은 다양한 변수에 따라 긍정 프레임의 효과가 더 높기도, 부정 프레임의 효과가 더 높기도 하는 등 이용자의 행동의도에 상이한 영향을 미친다(서소상·최유진, 2018; 오이름·이현진·김활빈, 2019; O'Keefe & Jensen, 2009; Rothman & Salovey, 1997).

인공지능 챗봇과 관련된 친밀감이나 상호작용 만족도에 관한 메시지 프레이밍 연구는 제한적이지만, 이용자의 태도 맥락에서 프레이밍의 효과를 유추해볼 수 있다. 에이젠(Ajzen, 1991)

에 따르면, 태도는 어떠한 대상에 대한 개인의 긍정적 혹은 부정적 평가 및 감정을 포함하는 심리적 변수로 정의되며, 메시지의 제시 방식에 따라 이용자는 특정 대상이나 상황에 대한 긍정적 혹은 부정적 태도를 형성할 수 있다. 오아름 등(2019)은 청소년 대상 자살 예방 캠페인 관련 메시지의 경우, 부정 프레이밍이 적용되었을 때 메시지에 대한 태도가 더 호의적이었으며 메시지 전파 의도가 높아졌다고 밝혔다. 반면, 이병관과 윤태웅(2012)은 공익광고에 대해 부정 프레이밍보다 긍정 프레이밍 메시지에 대한 태도가 더 호의적이었다고 밝혔다. 한편, 얼란드선 등(Erlandsson et al., 2018)은 윤리적인 판단을 요구하는 행동인 기부 결정은 부정 메시지 프레이밍이 더 효과적이었으나, 기부광고와 자선 단체에 대한 호감도는 긍정 메시지 프레이밍이 더 효과적이라고 밝혔다. 이처럼 선행연구는 기부와 같은 윤리적 판단을 요구하는 상황에서는 긍정 메시지 프레이밍과 부정 메시지 프레이밍의 효과에 차이가 있음을 밝혔다.

한편, 청소년의 또래집단 규범은 해석수준이론(construal level theory)에서 제시하는 심리적 거리 차원 중 사회적 거리(social distance)와 관련이 있다. 사회적 거리는 '나와 타인에 대한 유사성'의 거리를 의미하며, 이는 사람들이 자신과 가까운 집단(내집단, in-group)에 더 큰 호감을 표시하고 내집단 구성원의 행동을 구체적으로 생각하며, 해당 집단의 행동에 더 큰 영향을 받도록 한다(박현정, 2014). 청소년은 또래집단을 사회적 거리감이 가까운 집단으로 인지하며, 또래집단의 대다수가 어떠한 생각을 갖고 있는가에 따라 행동 의도를 결정한다(Akerlof, 1997; Yanovitzky, Stewart, & Lederman, 2006). 사회적 거리와 메시지 프레이밍과 관련된 선행연구에 따르면, 사회적 거리가 먼 상황일수록 긍정 프레이밍이, 가까운 상황의 경우 부정 프레이밍이 효과적인 것으로 나타났다(Nan, 2007). 본 연구의 맥락에서는 청소년들이 또래집단과의 사회적 거리를 가깝게 느껴 또래집단의 다수 의견에 대해 부정 프레이밍 메시지가 더 효과적일 수 있을 것으로 예측된다. 그러나 지금까지 청소년의 또래집단 규범과 메시지 프레이밍 효과에 관한 연구는 아직 충분하지 않으므로 예상과 다른 결과가 도출될 가능성도 존재한다.

이에 본 연구는 비윤리적 상황에서 청소년의 또래집단 규범과 인공지능 챗봇 메시지의 프레이밍이 상호작용하였을 때의 설득효과가 어떻게 달라지는지 살펴보고자 한다. 이러한 논의를 통해 메시지 프레이밍 방식에 따라 또래집단 규범의 효과는 차이를 보일 것이라는 가설을 도출하였다.

## **연구기설 2. 비윤리적 행동에 대한 인공지능 챗봇의 메시지 프레이밍(긍정/부정)에 따라 청소년들**

의 a) 비윤리적 행동 중단의도, b) 챗봇에 대한 친밀감, c) 챗봇과의 상호작용에 대한 만족도에 차이가 있을 것이다.

**연구기설 3.** 비윤리적 행동에 대한 또래집단 규범 조건과 인공지능 챗봇 메시지 프레이밍은 청소년들의 a) 비윤리적 행동 중단의도, b) 챗봇에 대한 친밀감, c) 챗봇과의 상호작용에 대한 만족도에 상호작용 효과를 미칠 것이다.

### 3. 연구방법

본 연구는 3(또래집단 규범: 비윤리적 행동에 대한 긍정 의견 vs. 부정 의견 vs. 특별한 의견 없음) X 2(인공지능 챗봇 메시지 프레이밍(message frame): 긍정 vs. 부정)로 설계한 피험자 간 실험설계를 사용하고 연구결과의 일반화를 위해 두 개의 독립된 실험을 진행하였다. 이 두 실험은 청소년이 경험할 수 있는 서로 다른 비윤리적 행동 시나리오를 다루는 것을 제외하고는 동일한 방식으로 진행되었다. 각 실험은 모두 독립적으로 진행되어 각기 다른 참여자가 두 개의 실험에 각각 배치되었다. 실험 1에서는 온라인 강의 자동수강 프로그램 사용과 관련된 시나리오를 사용했으며, 실험 2에서는 무단결석과 관련된 시나리오를 사용하였다.

2023년 12월 18일부터 29일까지 총 12일간 온라인 설문조사 전문업체인 한국리서치에 의뢰하여 만 13-18세의 청소년을 대상으로 온라인 실험과 설문조사를 실시하였다. 한국리서치는 청소년기 자녀를 둔 성인 패널 총 10,453명에게 이메일, 카카오톡, 전화 등을 통해 자녀의 온라인 실험 참여에 대한 동의를 구했다. 부모의 동의를 받은 청소년 패널 1,282명에게 설문링크를 전달해 설문참여 동의 및 설문을 진행하였다. 이 중 설문 미응답자 656명, 설문참여 비동의자 32명, 설문 중단자 235명, 조작점검을 통과하지 못한 응답자 48명, 불성실 응답자 5명을 제외한 실험 1 참여자 총 152명, 실험 2 참여자 총 154명이 최종 분석에 사용되었다.

### 4. 실험 1

#### 1) 참여자

실험 1에서는 조작점검을 통과하지 못한 응답자 23명과 불성실한 응답자 2명을 제외한 총 152명의 응답이 최종 분석에 사용되었다. 본 연구는 성별(여성 51.3%,  $n = 78$ )을 균등하게 할당하여 표집하였으며, 분석에 포함된 참여자의 평균 나이는 15.8세( $SD = 1.65$ )였다. 참여자들의 인구통계학적 정보는 아래 <Table 1>과 같다.

실험 참여자는 또래집단 규범과 인공지능 챗봇의 메시지 프레임에 따라 무작위로 배치되었다. 각 조건 별 참여자 수와 인구통계학적 정보는 다음과 같다. 비윤리적 행동에 대한 또래집단의 찬성 의견 - 챗봇의 긍정 메시지 프레임(25명) 조건에서 여성 13명, 평균 나이는 15.9세( $SD = 1.79$ )였고, 또래집단의 반대 의견 - 챗봇의 긍정 메시지 프레임(25명) 조건에서는 여성이 11명, 평균 나이는 16세( $SD = 1.46$ ), 또래집단의 특별한 의견 없음 - 챗봇의 긍정 메시지 프레임(25명) 조건의 경우 여성이 13명, 평균 나이는 15.8세( $SD = 1.72$ )였다. 또래집단의 찬성 의견 - 챗봇의 부정 메시지 프레임(25명) 조건에서는 여성이 14명, 평균 나이는 15.8세( $SD = 1.69$ ), 또래집단의 반대 의견 - 챗봇의 부정 메시지 프레임(25명) 조건은 여성 13명, 평균 나이는 15.6세( $SD = 1.75$ )였으며, 또래집단의 특별한 의견 없음 - 챗봇의 부정 메시지 프레임(27명) 조건에서는 여성 14명, 평균 나이 15.7세( $SD = 1.63$ )로 구성되었다.

Table 1. Socio-demographic Distribution of the Sample - Experiment 1

Variable		<i>n</i>	%
Sex	Male	74	48.7
	Female	78	51.3
Age	13	15	9.9
	14	22	14.5
	15	34	22.4
	16	22	14.5
	17	25	16.4
	18	34	22.4
Education	Middle School	71	46.7
	High School	81	53.3
Family status	Below 2 people	6	3.9
	3 people	44	28.9
	4 people	82	53.9
	Above 5 people	20	13.2
Monthly allowance (₩)	Below 50,000	42	27.6
	50,000 ~ 100,000	54	35.5
	100,000 ~ 200,000	37	24.3
	200,000 ~ 300,000	14	9.2
	300,000 ~ 400,000	3	2.0
	Above 500,000	2	1.3

## 2) 실험설계 및 절차

실험 절차는 다음과 같다. 첫째, 청소년기 자녀를 둔 성인 패널에게 연구 내용을 설명하고 자녀의 온라인 실험·설문조사 참여 동의를 구했다. 둘째, 부모의 동의를 받은 청소년 참여자들은 실험·설문조사 시작 전 간단한 연구 설명을 듣고, 참여자가 원할 때 언제든지 설문을 중단할 수 있다는 안내를 받았다. 셋째, 실험 참여자들에게 인공지능 챗봇에 대해 간략히 소개했다. 넷째, 참

여자들은 또래집단 규범과 관련한 가상의 시나리오 메시지 중 하나를 무작위로 받고, 그 상황에 자신이 직접 있는 것처럼 상상하며 자극물을 경험했다(부록 I-1 참조). 이후 참여자들은 비윤리적 행동에 대한 적절성을 판단하기 위해 인공지능 챗봇 메시지 프레임 유형 중 하나에 무작위로 배치되었다(부록 I-2 참조). 마지막으로, 참여자들은 또래집단 규범과 인공지능 챗봇 메시지 프레임에 관한 조작점검 문항과 실험 자극물과 관련한 설문에 응답했다.

### 3) 자극물

실험 자극물은 청소년들이 일상적으로 경험할 수 있는 비윤리적 행동을 중심으로 한 가상의 시나리오와 그에 대한 인공지능 챗봇의 응답으로 구성되었다. 구체적으로 비윤리적 행동은 온라인 강의 자동수강 프로그램 사용이라는 상황으로 설정되었다.

본 연구는 호그와 레이드(Hogg & Reid, 2006)가 정의한 바와 같이, 또래집단 규범을 비슷한 연령대에 있는 구성원들로 이루어진 집단을 특징짓고 다른 집단들과의 차별성을 나타내는 태도 및 행동 규칙 양식으로 정의하였다. 이를 바탕으로 실험에서 사용된 시나리오는 세 가지 수준의 또래집단 규범(비윤리적 행동에 대한 찬성 의견 vs. 반대 의견 vs. 특별한 의견 없음)으로 조작되었다. 찬성 의견 조건에서는 대부분의 반 친구들이 비윤리적 행동에 찬성하는 입장을 취하고 있다는 내용을 담았다. 반대 의견 조건에서는 이를 반대하는 것으로, 특별한 의견이 없는 조건에서는 비윤리적 행동에 대해 특별한 의견이 없다는 내용으로 조작했다.

인공지능 챗봇의 메시지 프레임은 자극물에서 나타난 비윤리적 상황에 대한 적절성을 판단하기 위한 목적으로 챗봇 메시지 형태로 제작되었다. 긍정 프레임에서는 비윤리적 행동을 따르지 않았을 경우 발생하는 긍정적 결과를 강조하고, 부정 프레임에서는 비윤리적 행동을 수행했을 때의 부정적 결과를 강조하였다(Kühberger, 1998). 긍정 프레임 메시지는 온라인 강의 자동수강 프로그램을 사용하지 않을 시에 참여자가 얻을 수 있는 개인적 및 사회적 이점에 대한 정보를 제공함으로써 긍정적인 감정을 유도하도록 조작되었다. 반대로 부정 프레임 메시지는 해당 프로그램 사용의 부정적 결과, 즉 참여자와 공동체가 겪을 수 있는 부정적 영향을 강조하여 부정적 감정을 유발하도록 설정되었다.

### 4) 측정변수

#### (1) 조작점검 문항

각 조건 별로 가상의 시나리오와 인공지능 챗봇의 메시지를 읽게 한 후, 또래집단 규범 유형을 묻

는 2문항, 메시지 프레임 유형을 묻는 2문항에 응답하도록 했다. 또래집단 규범 유형은 박, 정, 그리고 이(Park, Jung, & Lee, 2011)의 연구에서 사용된 문항 중 일부를 발췌해 본 연구에 맞게 수정했다. “시나리오에서 언급된 대다수의 친구들은 비윤리적 행동(온라인 강의 자동수강[실험 1], 무단결석[실험 2])을 지지(vs. 반대 vs. 특별한 의견 없음)하고 있다”, “시나리오에서 언급된 대다수의 친구들은 비윤리적 행동(온라인 강의 자동 수강[실험 1], 무단결석[실험 2])에 긍정(vs. 부정 vs. 특별한 의견 없음)적인 태도를 보인다”의 2개의 문항이 포함되었으며, 각 문항은 6점 리커트 척도(1 = 매우 그렇지 않다, 6 = 매우 그렇다)로 측정했다(Cronbach's  $\alpha = .81$ ).

메시지 프레임의 조작점검은 남민정과 조민호(2012)의 연구와 케사리오, 그랜트, 그리고 히긴스(Cesario, Grant, & Higgins, 2004)의 연구를 참고해 2개의 문항을 사용했다. “이득 강조 - 손실 강조”, “긍정적 결과 강조 - 부정적 결과 강조”의 2개의 문항을 6점 의미분별척도로 측정했다(Cronbach's  $\alpha = .98$ ).

## (2) 비윤리적 행동 중단의도

인공지능 챗봇이 제시한 메시지에 대한 설득효과를 확인하고자 종속변수로 비윤리적 행동 중단 의도와 챗봇에 대한 인식을 측정했다. 비윤리적 행동 중단의도는 에이젠과 피쉬바인(Ajzen & Fishbein, 1975)의 연구를 바탕으로 한 최어진과 박남기(2023)의 측정항목을 참고해 본 연구에 맞게 수정했다. “나는 비윤리적 행동(온라인 강의 자동수강[실험 1], 무단결석[실험 2])을 중단할 것이다”, “나는 비윤리적 행동(온라인 강의 자동수강[실험 1], 무단결석[실험 2])을 계속할 것이다(역문항)”, “나는 비윤리적 행동(온라인 강의 자동수강[실험 1], 무단결석[실험 2])을 중단하기로 결정했다”, “나는 앞으로 비윤리적 행동(온라인 강의 자동수강[실험 1], 무단결석[실험 2])을 하지 않을 것이다”의 4개의 문항이 포함되었다. 각 문항은 7점 리커트 척도(1 = 매우 그렇지 않다, 7 = 매우 그렇다)로 측정했다( $M = 4.95$ ,  $SD = 1.53$ , Cronbach's  $\alpha = .94$ ).

## (3) 인공지능 챗봇에 대한 친밀감

인공지능 챗봇에 대한 친밀감은 이, 이, 그리고 사(Lee, Lee, & Sah, 2020)의 척도를 활용했다. “챗봇과 감정적으로 가까워진 느낌이 들었다”, “챗봇과 대화를 나누고 싶었다”, “챗봇과 친구가 된 것 같은 느낌이 들었다” 등 3개의 문항으로 측정했다( $M = 3.73$ ,  $SD = 1.49$ , Cronbach's  $\alpha = .88$ ).

#### (4) 인공지능 챗봇과의 상호작용에 대한 만족도

인공지능 챗봇과의 상호작용에 대한 만족도는 사이, 리우, 그리고 추안(Tsai, Liu, & Chuan, 2021)의 연구에서 사용된 척도를 활용했다. “인공지능 챗봇과의 상호작용이 즐거웠다”, “인공지능 챗봇과의 상호작용이 만족스러웠다”, “인공지능 챗봇과의 상호작용이 답답했다(역문항)”, “인공지능 챗봇 때문에 머리가 아팠다(역문항)”, “인공지능 챗봇과의 상호작용이 매우 어색했다(역문항)” 등 5개의 문항으로 측정했다( $M = 4.54$ ,  $SD = 1.23$ , Cronbach's  $\alpha = .90$ ).

#### (5) 통제변수

통제변수에는 선행연구들을 바탕으로 나이, 성별 등 인구통계학적 변수들이 포함되었고, 참여자의 반발 성향(reactance propensity)과 또래 동조 성향(peer conformity) 등 개인 성향 변수를 추가해 살펴보았다. 청소년들은 자신이 특정 행동을 수행하도록 요구받을 경우, 개인의 자율성이 제한되었다고 인식하여 반발심과 같은 심리적 저항감이 나타나는 경향이 있으므로, 높은 반발 성향을 가진 청소년들은 부정적인 태도와 행동의도를 보일 것으로 예측된다. 또한, 이들은 또래집단의 승인을 얻기 위해 그들의 규범을 따르는 경향이 있으므로, 또래 동조 성향 수준에 따라 태도 형성 및 행동 의도가 달라질 것으로 예상되어 두 변인을 공변량으로 추가하여 살펴보았다.

반발 성향은 홍과 페다(Hong & Faedda, 1996)의 척도를 본 연구에 맞게 수정하여 사용했으며, “나는 누군가의 충고를 받으면 간섭받는 것 같이 생각된다”, “나는 독립적이고 자유로운 선택을 하지 못할 경우 좌절감을 느낀다”, “누군가가 내가 이미 알고 있는 명백한 사실을 지적하면 짜증이 난다” 등 총 9개의 문항이 포함되었다( $M = 3.09$ ,  $SD = .80$ , Cronbach's  $\alpha = .88$ ).

또래 동조 성향은 번트(Berndt, 1979)가 개발하고 이수봉(2017)이 재수정한 척도를 본 연구에 맞게 수정하여 사용했으며, “속마음으로는 원치 않을 때도 친구들의 기분을 맞추기 위해 어떤 행동을 하는 때가 있다”, “가까운 친구가 어디에 가자고 하거나 무엇을 하자고 할 때 별로 내키지 않아도 따를 때가 있다”, “나는 옷이나 신발 등을 구입할 때 주변 친구들의 의견을 쉽게 따르는 편이다”를 포함, 총 10개의 문항으로 측정하였다( $M = 2.21$ ,  $SD = .44$ , Cronbach's  $\alpha = .76$ ).

### 5) 연구결과

측정변수들의 평균, 표준편차, 상관관계의 결과는 <Table 2>와 같다.

Table 2. Descriptive Statistics and Correlation among Measured Variables

	1	2	3	4	5	6	7	M	SD
1. PGN	1							-	-
2. MF	.02	1						-	-
3. IS	-.01	.16	1					4.95	1.53
4. CN	-.05	.15	.01	1				3.73	1.49
5. CS	-.03	.17*	.01	.58**	1			4.54	1.23
6. RP	-.10	-.14	-.23**	.11	.04	1		3.09	0.80
7. PC	-.13	-.05	-.12	.35**	.09	.43**	1	2.21	0.44

Note. PGN = Peer Group Norm, MF = Message Frame, IS = Intention to Stop, CN = Closeness, CS = Chatbot Satisfaction, RP = Reactance Propensity, PC = Peer Conformity. M and SD represent mean and standard deviation, respectively. \* $p < .05$ , \*\* $p < .01$ .

### (1) 연구가설 1 검증: 비윤리적 행동에 대한 또래집단 규범의 영향

온라인 강의 자동수강이라는 비윤리적 행동에 대한 또래집단 규범이 청소년들의 비윤리적 행동 중단의도, 챗봇에 대한 친밀감, 챗봇과의 상호작용에 대한 만족감에 미치는 영향을 살펴보고, 공변량인 반발 성향, 또래 동조 성향, 나이, 그리고 성별의 영향을 살펴보기 위해 이원공분산분석(two-way ANCOVA)을 실시하였다. 이원공분산분석을 실시하기에 앞서 분산의 동질성 가정을 확인했으며 모든 조건에서 분산의 동질성 가정이 만족하는 것으로 나타났다(비윤리적 행동 중단 의도:  $F(5, 146) = 1.68, p = .143$ , 친밀감:  $F(5, 146) = 1.54, p = .180$ , 만족도:  $F(5, 146) = 1.30, p = .266$ ).

〈연구가설 1〉과 관련하여, 또래집단 규범에 따른 비윤리적 행동 중단의도의 평균과 표준편차는 〈Table 3〉에 제시되었다. 〈Table 3〉의 결과를 통해, 비윤리적 행동에 대해 또래집단 규범이 반대하는 조건은, 찬성 조건이나 특별한 의견이 없는 조건보다 비윤리적 행동 중단의도가 높은 것이 확인되었다. 그러나 〈Table 4〉에서 볼 수 있듯이, 비윤리적 행동 중단의도에 또래집단 규범이 미치는 영향은 통계적으로 유의미하지 않은 것으로 나타나 〈연구가설 1〉은 기각되었다.

Table 3. Means and Standard Deviations of Intention to Stop based on Peer Group Norm and Message Frame

	Agree	Disagree	No opinion	Total
Positive frame	4.53 (1.71)	5.14 (1.40)	4.47 (1.80)	4.71 (1.65)
Negative frame	5.04 (1.54)	5.5 (1.12)	5.03 (1.40)	5.19 (1.37)
Total	4.78 (1.63)	5.32 (1.27)	4.76 (1.61)	4.95 (1.53)

Note. The numbers in the parentheses indicate standard deviation.



Table 4. Results of Analysis of Variance on Intention to Stop with Peer Group Norm and Message Frame Treatments

Source	df	F	p	partial $\eta^2$
Peer group norm	2	2.51	.085	.034
Message frame	1	2.28	.133	.016
Peer group norm x Message frame	2	0.02	.979	.000
Reactance propensity	1	5.60*	.019	.038
Peer Conformity	1	0.38	.538	.003
Age	1	1.63	.204	.011
Sex	1	0.10	.752	.001
Error	142	(2.20)		

Note. The numbers in the parentheses indicate MSE. Adjusted  $R^2 = .06$ . \* $p < .05$ .

〈연구문제 1-a〉와 관련하여, 또래집단 규범에 따른 인공지능 챗봇에 대한 친밀감의 평균과 표준편차는 아래의 〈Table 5〉와 같다. 〈Table 5〉에 제시된 바와 같이 또래집단 규범 조건에 따라 챗봇에 대한 친밀감은 유의미한 차이를 보이지 않는 것으로 확인되었다. 또한, 〈Table 6〉에 나타난 분석 결과를 보면, 챗봇에 대한 친밀감을 또래집단 규범 조건에 따라 구분했을 때, 집단 간에 통계적으로 유의미한 차이가 없는 것으로 나타나 〈연구문제 1-a〉는 차이가 없는 것으로 나타났다.

Table 5. Means and Standard Deviations of Closeness based on Peer Group Norm and Message Frame

	Agree	Disagree	No opinion	Total
Positive frame	3.6 (1.46)	3.65 (1.41)	3.25 (1.55)	3.50 (1.46)
Negative frame	4.04 (1.77)	3.83 (1.36)	3.96 (1.38)	3.94 (1.49)
Total	3.82 (1.62)	3.74 (1.38)	3.62 (1.49)	3.73 (1.49)

Note. The numbers in the parentheses indicate standard deviation.

Table 6. Results of Analysis of Variance on Closeness with Peer Group Norm and Message Frame Treatments

Source	df	F	p	partial $\eta^2$
Peer group norm	2	0.03	.969	.000
Message frame	1	4.66*	.033	.032
Peer group norm x Message frame	2	0.68	.510	.009
Reactance propensity	1	0.17	.680	.001
Peer conformity	1	18.89***	.000	.117
Age	1	0.48	.489	.003
Sex	1	6.05*	.015	.041
Error	142	(1.89)		

Note. The numbers in the parentheses indicate MSE. Adjusted  $R^2 = .15$ . \* $p < .05$ , \*\*\* $p < .001$ .

마지막으로, <연구문제 1-b>와 관련하여 인공지능 챗봇과의 상호작용에 대한 만족도의 평균과 표준편차는 <Table 7>에서 볼 수 있다. <Table 7>을 살펴보면, 또래집단 규범 조건에 따라 집단 간 챗봇에 대한 만족도는 유의미한 차이가 없었다. 챗봇과의 상호작용에 대한 만족도에 영향을 미치는 다른 변수들에 관한 분석 결과는 <Table 8>에서 확인할 수 있는 것과 같이 또래 집단 규범 조건에 따른 챗봇과의 상호작용에 대한 만족도의 집단 간 차이는 통계적으로 유의미하지 않았다. 따라서 <연구문제 1-b>는 차이가 없는 것으로 확인되었다.

Table 7. Means and Standard Deviations of Chatbot Satisfaction based on Peer Group Norm and Message Frame

	Agree	Disagree	No opinion	Total
Positive frame	4.54 (1.15)	4.52 (1.36)	3.93 (1.44)	4.33 (1.34)
Negative frame	4.45 (1.27)	4.96 (0.82)	4.83 (1.09)	4.75 (1.08)
Total	4.50 (1.20)	4.74 (1.13)	4.40 (1.34)	4.54 (1.23)

Note. The numbers in the parentheses indicate standard deviation.

Table 8. Results of Analysis of Variance on Chatbot Satisfaction with Peer Group Norm and Message Frame Treatments

Source	df	F	p	partial $\eta^2$
Peer group norm	2	0.95	.391	.013
Message frame	1	6.14*	.014	.042
Peer group norm x Message frame	2	2.38	.097	.032
Reactance propensity	1	0.09	.762	.001
Peer conformity	1	1.96	.164	.014
Age	1	3.86	.052	.026
Sex	1	6.71*	.011	.045
Error	142	(1.38)		

Note. The numbers in the parentheses indicate MSE. Adjusted  $R^2 = .09$ . \* $p < .05$ .

## (2) 연구가설 2 검증: 인공지능 챗봇 메시지 프레임의 영향

인공지능 챗봇이 전달하는 메시지 프레임 유형에 따라 청소년들의 비윤리적 행동 중단의도를 살펴본 <연구가설 2-a>의 평균과 표준편차는 <Table 3>과 같다. <Table 3>에 나타난 것처럼 부정적 메시지 프레임의 경우, 긍정적 메시지 프레임보다 비윤리적인 행동 중단의도가 높았다. 이 원공분산분석을 실시한 결과, 메시지 프레임의 차이가 비윤리적 행동 중단의도에 미치는 영향은 통계적으로 유의미하지 않아, <연구가설 2-a>는 기각되었다.

<연구가설 2-b>는 인공지능 챗봇의 메시지 프레임 유형에 따른 청소년들의 챗봇에 대한 친밀

감 차이를 살펴보았다. 부정적인 메시지가 긍정적인 메시지보다 청소년들의 챗봇에 대한 친밀감을 높였다(〈Table 5〉 참조). 〈Table 6〉의 분석 결과에 따르면, 메시지 프레임에 따라 청소년들이 챗봇에 대해 느끼는 친밀감에 유의미한 차이가 있었다( $F(1, 142) = 4.66, p < .05$ , 부분  $\eta^2 = .032$ ). 즉, 청소년들은 메시지의 내용이 긍정 또는 부정인지에 따라 챗봇에 대해 느끼는 친밀감을 다르게 인식한다는 것이다. 사후분석 결과, 부정 프레임 메시지를 받은 경우, 긍정 프레임 메시지보다 챗봇에 대한 친밀감이 높은 것으로 나타났다(difference of  $M = 0.44, p = .067$ ).

마지막으로 〈연구가설 2-c〉와 관련하여, 메시지 프레임 유형에 따른 챗봇과의 상호작용에 대한 만족도의 평균과 표준편차는 〈Table 7〉과 같으며, 부정적인 메시지 프레임이 긍정적인 측면을 강조한 경우보다 챗봇과의 상호작용에 대한 만족도가 높았다. 〈Table 8〉의 분석 결과는 메시지 프레임이 챗봇과의 상호작용에 대한 만족도에 유의미한 영향을 미치는 것을 보여준다( $F(1, 142) = 6.14, p < .05$ , 부분  $\eta^2 = .042$ ). 청소년들의 챗봇과의 상호작용에 대한 만족도는 챗봇이 제공하는 메시지 프레임 유형에 따라 다르게 나타났다. 사후분석 결과, 청소년들은 인공지능 챗봇의 메시지가 부정 프레임인 경우, 긍정 프레임보다 챗봇과의 상호작용에 대한 만족도가 높은 것으로 나타났다(difference of  $M = 0.42, p = .035$ ).

### (3) 연구가설 3 검증: 비윤리적 행동에 대한 또래집단 규범과 인공지능 챗봇 메시지 프레임의 조절효과

나이, 성별, 또래 동조 성향, 반발 성향 등의 공변량을 통제한 후 살펴본 결과, 또래집단 규범과 인공지능 챗봇 메시지의 프레임 간의 상호작용 효과는 통계적으로 유의미하지 않은 것으로 나타나 〈연구가설 3〉은 기각되었다.

## 5. 실험 2

### 1) 실험설계 및 자극물

실험 2는 실험 1과 동일한 설계와 절차를 따르되, 결과의 일반화를 위해 실험 자극물의 주제를 무단결석 시나리오로 변경하여 구성했다. 즉, 시나리오의 주제만 실험 1과 다르게 하고, 내용 구성은 모두 동일하게 유지했다. 시나리오는 세 가지 수준의 또래집단 규범(비윤리적 행동에 대한 찬성 의견 vs. 반대 의견 vs. 특별한 의견 없음)으로 조작했다. 또래집단 규범이 비윤리적 행동에 대한 찬성 의견 조건에서는 대부분의 반 친구들이 무단결석에 찬성한다는 내용, 반대 의견 조

건에서는 반대한다는 내용, 특별한 의견이 없는 조건에서는 대부분의 반 친구들이 무단결석에 대해 특별한 의견이 없다는 내용으로 구성하였다(부록 II-1 참조).

인공지능 챗봇의 메시지 프레임 또한 실험 1과 동일하게 유지했으나 비윤리적 행동 주제에 만 차이를 두었다. 실험 2 자극물의 긍정 프레임 메시지에서는 무단결석을 하지 않을 경우 참여자가 얻을 수 있는 개인적 및 사회적 긍정적인 결과를 강조하여 참여자가 긍정적인 감정을 느낄 수 있도록 했다. 반면, 부정 프레임 메시지에서는 무단결석 시 참여자와 공동체가 경험할 수 있는 부정적 영향을 강조하는 내용으로 제시했다(부록 II-2 참조).

## 2) 참여자

실험 2에서는 부모의 동의를 받은 후 조작점검을 통과하지 못한 응답자 25명과 불성실한 응답자 3명을 제외한 총 154명의 응답이 분석에 사용되었다. 본 연구에서는 성별(여성 48.1%,  $n = 74$ )을 기준으로 할당표집 했으며, 참여자들의 평균 나이는 15.64세( $SD = 1.68$ )였다. 참여자들의 인구통계학적 정보는 아래 <Table 9>와 같다.

Table 9. Socio-demographic Distribution of the Sample - Experiment 2

Variable		<i>n</i>	%
Sex	Male	80	51.9
	Female	74	48.1
Age	13	18	11.7
	14	27	17.5
	15	32	20.8
	16	23	14.9
	17	23	14.9
	18	31	20.1
Education	Middle school	75	48.7
	High school	79	51.3
Family status	Below 2 people	1	.6
	3 people	44	28.6
	4 people	83	53.9
	Above 5 people	26	16.8
Monthly allowance (₩)	Below 50,000	41	26.6
	50,000 ~ 100,000	46	29.9
	100,000 ~ 200,000	40	26.0
	200,000 ~ 300,000	14	9.1
	300,000 ~ 400,000	9	5.8
	Above 500,000	4	2.6

실험 참여자는 또래집단 규범과 인공지능 챗봇의 메시지 프레임에 따라 무작위로 배치되었다. 비윤리적 행동에 대한 또래집단의 찬성 의견 - 챗봇의 긍정 메시지 프레임(25명) 조건에서 여성이 13명, 평균 나이는 15.6세( $SD = 1.58$ ), 또래집단의 반대 의견 - 챗봇의 긍정 메시지 프레임(25명) 조건에서는 여성이 12명, 평균 나이는 15.9세( $SD = 1.45$ ), 또래집단의 특별한 의견 없음 - 챗봇의 긍정 메시지 프레임(26명) 조건의 경우 여성이 11명, 평균 나이는 15.5세( $SD = 1.96$ )였다. 또래집단의 찬성 의견 - 챗봇의 부정 메시지 프레임(25명) 조건에서는 여성이 12명, 평균 나이는 15.7세( $SD = 1.59$ ), 또래집단의 반대 의견 - 챗봇의 부정 메시지 프레임(26명) 조건은 여성이 12명, 평균 나이는 15.5세( $SD = 1.88$ )였으며, 또래집단의 특별한 의견 없음 - 챗봇의 부정 메시지 프레임(27명) 조건에서는 여성이 14명, 평균 나이는 15.6세( $SD = 1.65$ )로 구성되었다.

### 3) 측정변수

#### (1) 조작점검 문항

실험 2의 조작점검 문항은 실험 1의 조작점검 문항과 같은 문항으로 구성되었다. 또래집단 규범 조작점검 문항들의 신뢰도는 Cronbach's  $\alpha = .80$ 이며, 메시지 프레임의 조작점검 문항들의 신뢰도는 Cronbach's  $\alpha = .97$ 이다.

#### (2) 측정변수

모든 측정변수들은 실험 1의 측정변수와 동일하다. 비윤리적 행동(무단결석) 중단의도의 신뢰도는 Cronbach's  $\alpha = .93$ 이며, 인공지능 챗봇에 대한 친밀감의 신뢰도는 Cronbach's  $\alpha = .87$ 이다. 인공지능 챗봇과의 상호작용에 대한 만족도의 신뢰도는 Cronbach's  $\alpha = .92$ 이며, 반발 성향의 신뢰도는 Cronbach's  $\alpha = .86$ 이다. 또래 동조 성향의 신뢰도는 Cronbach's  $\alpha = .75$ 로 나타났다.

### 5) 연구결과

측정변수들의 평균, 표준편차, 상관관계 결과는 <Table 10>과 같다.

Table 10. Descriptive Statistics and Correlation among Measured Variables

	1	2	3	4	5	6	7	M	SD
1. PGN	1							-	-
2. MF	.01	1						-	-
3. IS	.14	.02	1					5.73	1.20
4. CN	-.11	-.09	-.08	1				3.91	1.40
5. SC	-.06	-.06	.11	.65**	1			4.76	1.19
6. RP	-.06	-.12	-.24**	-.01	-.05	1		3.06	0.75
7. PC	-.01	.00	-.21*	.24**	.10	.35**	1	2.24	0.43

Note. PGN = Peer Group Norm, MF = Message Frame, IS = Intention to Stop, CN = Closeness, SC = Satisfaction with Chatbot, RP = Reactance Propensity, PC = Peer Conformity. M and SD represent mean and standard deviation, respectively. \* $p < .05$ , \*\* $p < .01$ .

### (1) 연구가설 1 검증: 비윤리적 행동에 대한 또래집단 규범의 영향

무단결석이라는 비윤리적 행동에 대한 또래집단 규범 조건에 따라 청소년들의 비윤리적 행동 중단의도, 챗봇에 대한 친밀감, 챗봇과의 상호작용에 대한 만족감에 미치는 영향을 각각 확인하고, 공변량인 반발 성향, 또래 동조 성향, 나이, 그리고 성별의 영향을 살펴보기 위해 이원공분산분석(two-way ANCOVA)을 실시하였다. 이원공분산분석을 실시하기에 앞서 등분산 가정을 확인했다. 비윤리적인 행동 중단의도에 대한 조건에서는 등분산 가정을 만족하지 못했으나( $F(5, 148) = 3.75, p < .05$ ), 챗봇에 대해 느끼는 친밀감 조건과 챗봇과의 상호작용에 대한 만족감의 조건에서는 등분산 가정을 만족하는 것으로 나타났다(친밀감:  $F(5, 148) = 1.71, p = .135$ , 만족도:  $F(5, 148) = 0.98, p = .430$ ).

또래집단 규범 조건에 따른 비윤리적 행동 중단의도의 평균과 표준편차는 <Table 11>과 같다. <연구가설 1>과 관련하여, 또래집단 규범이 인공지능 챗봇이 제공하는 비윤리적 행동에 대해 반대하거나 특별한 의견이 없는 경우, 찬성하는 조건보다 비윤리적 행동 중단의도가 높았다. 이때 등분산 가정은 충족되지 않았으나 그룹 간 표준편차의 차이가 크지 않아 이원공분산분석을 실시하였다. <Table 12>의 결과와 같이 또래집단 규범이 비윤리적 행동 중단의도에 유의미한 차이가 나타나지 않아 <연구가설 1>은 기각되었다.

Table 11. Means and Standard Deviations of Intention to Stop based on Peer Group Norm and Message Frame

	Agree	Disagree	No opinion	Total
Positive frame	5.44 (1.65)	5.93 (0.95)	5.72 (1.36)	5.70 (1.35)
Negative frame	5.35 (1.10)	5.96 (0.99)	5.93 (1.01)	5.75 (1.05)
Total	5.40 (1.39)	5.95 (0.96)	5.83 (1.18)	5.73 (1.20)

Note. The numbers in the parentheses indicate standard deviation.

Table 12. Results of Analysis of Variance on Intention to Stop with Peer Group Norm and Message Frame Treatments

Source	df	F	p	partial $\eta^2$
Peer group norm	2	2.29	.105	.031
Message frame	1	0.00	.967	.000
Peer group norm x Message frame	2	0.59	.556	.008
Reactance propensity	1	4.21*	.042	.028
Peer Conformity	1	3.06	.082	.021
Age	1	0.06	.812	.000
Sex	1	2.66	.105	.018
Error	144	(1.35)		

Note. The numbers in the parentheses indicate MSE. Adjusted  $R^2 = .72$ . \* $p < .05$ .

〈연구문제 1-a〉와 관련하여 또래집단 규범 조건에 따른 챗봇에 대해 느끼는 친밀감의 평균과 표준편차는 아래의 〈Table 13〉에 제시되었다. 결과에 따르면, 또래집단 규범이 특별한 의견 없음, 찬성, 반대 순으로 챗봇에 대한 친밀감이 높아지는 것으로 나타났다. 〈Table 14〉의 분석 결과에 따르면, 챗봇에 대한 친밀감은 또래집단 규범에 따라 유의미한 집단 간 차이를 보였다 ( $F(2, 144) = 3.38, p < .05$ , 부분  $\eta^2 = .045$ ). 이는 청소년에게 있어 또래집단 규범이 챗봇에 대한 친밀감에 영향을 미친다는 것을 의미한다. 사후분석 결과, 비윤리적 행동에 대해 반대하는 또래집단의 조건이 특별한 의견이 없는 조건보다 챗봇에 대한 친밀감이 더 높았으며(difference of  $M = 0.64, p = .054$ ), 비윤리적 행동에 대해 찬성하는 또래집단의 조건은 특별한 의견이 없는 경우보다 챗봇에 대한 친밀감이 더 높게 나타났다(difference of  $M = 0.38, p = .347$ ).

Table 13. Means and Standard Deviations of Closeness based on Peer Group Norm and Message Frame

	Agree	Disagree	No opinion	Total
Positive frame	3.88 (1.60)	4.29 (1.43)	3.96 (1.59)	4.04 (1.53)
Negative frame	4.04 (1.08)	4.14 (1.21)	3.21 (1.32)	3.79 (1.27)
Total	3.96 (1.35)	4.22 (1.31)	3.58 (1.50)	3.91 (1.40)

Note. The numbers in the parentheses indicate standard deviation.

Table 14. Results of Analysis of Variance on Closeness with Peer Group Norm and Message Frame Treatments

Source	df	F	p	partial $\eta^2$
Peer group norm	2	3.38*	.037	.045
Message frame	1	1.72	.192	.012
Peer group norm x Message frame	2	2.56	.080	.034
Reactance propensity	1	1.37	.243	.009
Peer Conformity	1	13.37***	.000	.085
Age	1	0.06	.451	.004
Sex	1	2.66	.105	.018
Error	144	(1.77)		

Note. The numbers in the parentheses indicate MSE. Adjusted  $R^2 = .10$ . \* $p < .05$ , \*\*\* $p < .001$ .

마지막으로 <연구문제 1-b>의 경우, 또래집단 규범 조건에 따른 챗봇과의 상호작용에 대한 만족도의 평균과 표준편차는 아래 <Table 15>와 같다. 또래집단 규범에 따라 집단을 구분했을 때, 집단 간 챗봇에 대한 만족도에는 유의미한 차이가 없었다. 이와 관련된 변수들을 포함한 분석 결과는 <Table 16>과 같다. <연구문제 1-b>의 챗봇과의 상호작용에 대한 만족도는 또래집단 규범 조건에 따른 집단 간 통계적으로 유의미한 차이가 없는 것으로 나타났다.

Table 15. Means and Standard Deviations of Chatbot Satisfaction based on Peer Group Norm and Message Frame

	Agree	Disagree	No opinion	Total
Positive frame	4.71(1.41)	4.91(1.14)	4.88(1.09)	4.83(1.21)
Negative frame	4.79(1.08)	5.00(1.23)	4.30(1.15)	4.69(1.18)
Total	4.75(1.25)	4.96(1.18)	4.58(1.15)	4.76(1.19)

Note. The numbers in the parentheses indicate standard deviation.

Table 16. Results of Analysis of Variance on Chatbot Satisfaction with Peer Group Norm and Message Frame Treatments

Source	df	F	p	partial $\eta^2$
Peer group norm	2	1.32	.272	.018
Message frame	1	0.73	.394	.005
Peer group norm x Message frame	2	2.08	.129	.028
Reactance propensity	1	1.27	.262	.009
Peer conformity	1	4.50	.036	.030
Age	1	0.56	.457	.004
Sex	1	1.57	.213	.011
Error	142	(1.39)		

Note. The numbers in the parentheses indicate MSE. Adjusted  $R^2 = .02$ . \* $p < .05$ .

## (2) 연구가설 2 검증: 인공지능 챗봇 메시지 프레임의 영향

분석 결과, 메시지 프레임 유형에 따른 비윤리적 행동 중단의도, 챗봇에 대한 친밀감, 챗봇과의 상호작용에 대한 만족도에서 통계적으로 유의미한 차이가 나타나지 않아 <연구가설 2>는 기각되었다.

## (3) 연구가설 3 검증: 비윤리적 행동에 대한 또래집단 규범과 인공지능 챗봇 메시지 프레임의 조절효과

나이, 성별, 또래 동조 성향, 반발 성향을 포함한 공변량을 통제한 분석에서 또래집단 규범과 메시지 프레임의 상호작용 효과는 통계적으로 유의미하지 않아 <연구가설 3>은 기각되었다.



Table 17. Hypothesis Tests

Hypothesis	Rejection status		Hypothesis	Rejection status	
	Study 1	Study 2		Study 1	Study 2
H1	Not supported	Not supported	H3a	Not supported	Not supported
RQ1a	Not significant	Significant	H3b	Not supported	Not supported
RQ1b	Not significant	Not significant	H3c	Not Supported	Not supported
H2a	Not supported	Not supported			
H2b	Supported	Not supported			
H2c	Supported	Not supported			

## 6. 결론

### 1) 연구결과에 대한 해석 및 논의

본 연구는 청소년이 경험할 수 있는 비윤리적인 행동과 관련한 또래집단 규범과 인공지능 챗봇의 메시지 프레임이 비윤리적 행동 중단 의도, 인공지능 챗봇에 대한 친밀감, 챗봇과의 상호작용에 대한 만족도에 영향을 미치는지 분석하여 비윤리적인 상황에서 인공지능 챗봇의 효과성을 검증해 보고자 했다. 본 연구를 통해 도출된 결과와 그 함의는 다음과 같다.

첫째, 비윤리적인 행동과 관련한 또래집단 규범이 비윤리적 행동 중단 의도에 미치는 영향은 유의미하지 않았다. 이러한 결과가 도출된 이유는 두 가지 관점에서 설명될 수 있다. 먼저, 본 연구의 또래집단 규범은 사회규범 중 기술적 규범(descriptive norm)을 활용해 시나리오가 제시되었다. 그러나 선행연구(정일권, 2019; Pedersen et al., 2017)에서는 기술적 규범보다 명령적 규범(injunctive norm)의 효과성을 검증한 바 있다. 가령 정일권(2019)은 소셜미디어에서 기술적 규범 관찰 경험보다 명령적 규범 관찰 경험이 언어적 공격을 포함한 탈규범 제재 행동 동참에 효과적이라고 밝혔다. 그러나 본 연구는 명령적 규범에 대한 고려 없이 기술적 규범만을 다루었다는 점에서 효과가 나타나지 않은 것으로 추론해볼 수 있다. 향후 연구에서는 청소년의 비윤리적 행동 중단 의도에 있어 사회규범의 범위를 확장해 살펴볼 필요가 있다. 다음으로, 잘못된 행동에 대해 또래집단의 압력은 개인마다 다른 효과를 보인다. 선행연구에 따르면, 또래와 맺은 관계의 수준이 높은 경우, 또래집단 규범이 청소년 행동에 미치는 영향이 큰 반면, 낮은 경우에는 영향이 적은 것으로 나타났다(박현정·타쿠히코 데구치·토시카즈 요시다, 2007). 또한, 사회화 수준이 높은 청소년에게 또래집단 규범이 더 크게 인식되고 위법을 감수하는 경향이 높다는 연구결과도 존재한다(Clasen & Brown, 1985). 이 외에 인지된 압력, 순응 성향 등 여러 변수들과 또래집단 압력은 복잡하게 작용하지만, 본 연구에서는 영향을 미칠 수 있는 다양한 변수들

을 측정하지 않았기 때문에 또래집단 규범이 비윤리적 행동 중단의도에 유의미한 영향을 미치지 않았다고 추측할 수 있다. 향후 연구에서는 또래집단 규범의 효과에 영향을 미치는 개인적 특성 변수들을 고려할 필요가 있다. 한편, 본 연구에서 통계적으로 유의미하지 않았으나 또래집단 규범이 비윤리적 행동에 대해 반대하는 조건의 경우 비윤리적 행동 중단의도의 평균값이 두 개의 실험에서 모두 높은 것을 확인했다. 이는 비윤리적 행동을 또래집단이 반대하는 규범을 갖고 있다면 청소년이 동조할 가능성이 높다는 것을 시사한다. 향후 연구에서는 또래집단 규범이 청소년의 비윤리적 행동에 미치는 영향을 보다 구체적으로 탐구하기 위해 다양한 실험 조건에서 실험 참여자 수를 늘려 검토할 필요가 있다.

둘째, 비윤리적 행동에 대한 또래집단 규범이 인공지능 챗봇에 대한 친밀감과 상호작용에 대한 만족도에 미치는 영향을 분석한 결과, 실험 2에의 조건에서만 챗봇에 대한 친밀감에 유의한 영향을 미쳤다. 온라인 강의 자동수강 프로그램의 활용은 최근 대두되는 문제이기에 청소년들의 윤리적 가치나 사회규범 확립이 부족한 반면, 무단결석은 장기간에 걸쳐 논의되어 온 비윤리적인 행동이자 지양해야 할 행동으로 인식되어 결과에 차이가 나타났을 수 있다. 실험 2에서는 비윤리적 행동에 대한 또래집단의 규범이 반대일 때 인공지능 챗봇에 대한 친밀감이 가장 높았고, 이는 찬성 및 특별한 의견 없음 조건에 비해 높은 수준이었다. 이 같은 결과는 또래집단의 의견이 비윤리적 행동에 대해 찬성하거나 반대할 때, 청소년들은 자신의 행동 결정에 대해 더 많은 고민을 하게 되어 인공지능 챗봇의 조언에 더 의존하게 되었을 것이라는 추론이 가능하다. 특히, 인공지능 챗봇이 비윤리적 행동에 대해 반대하는 내용을 기반으로 한 조언을 제공했기 때문에 또래집단 규범의 반대는 인공지능 챗봇의 조언과 일치하여 청소년들이 정서적 안정감을 느끼고 챗봇에 대한 긍정적인 인식을 형성했을 것으로 추측된다. 반면, 또래집단 규범이 특별한 의견이 없는 경우에는 명확한 사회규범의 부족으로 인해 챗봇의 감정적 공감이나 지지가 상대적으로 필요하지 않아 친밀감이 낮았을 것으로 보인다. 이는 또래집단이 비윤리적 행동에 대해 명확한 의견을 갖고 있지 않았기 때문에 청소년들이 해당 행동에 대해 크게 고민하지 않았을 가능성이 크며, 결과적으로 인공지능 챗봇에게서 조언이나 감정적 지지를 기대하지 않았다는 것을 의미한다.

인공지능 챗봇과의 상호작용에 대한 만족도는 실험 1과 실험 2에서 모두 유의미한 결과가 나타나지 않았다. 인공지능 챗봇은 청소년들에게 정보 검색 혹은 간단한 질문에 답변하는 도구로 인식되었을 수 있다. 이 같은 인식은 또래집단 규범 수준과 관계없이 인공지능 챗봇과의 상호작용에 대한 만족도에 차이를 보이지 않았을 수 있다. 친밀감은 정서적 지원에 의해 영향을 받는 경향이 있지만(Parks & Floyd, 1996), 만족도는 상호작용의 전반적인 유용성, 용이성, 유희성 등에 더 크게 영향을 받는다(Muhammad, Jiang, Shubin, & Sandra, 2020). 따라서 또

래집단 규범은 인공지능 챗봇에 대한 친밀감에 영향을 미칠 수 있으나 챗봇과의 상호작용에 대한 만족도에는 유의미한 영향을 미치지 못한 것으로 해석할 수 있다. 향후 연구에서는 인공지능 챗봇의 역할과 기능을 확장하고 청소년 개인의 상황에 더욱 맞춤형된 조언을 제공함으로써 챗봇의 유용성을 강화하는 방향으로 연구를 진행해야 할 것이다. 나아가 이러한 맞춤형된 메시지가 인공지능 챗봇과의 상호작용에 대한 만족도에 어떤 영향을 미치는지 살펴볼 필요가 있다.

셋째, 인공지능 챗봇 메시지 프레임이 청소년들의 비윤리적 행동 중단의도와 인공지능 챗봇에 대한 친밀감과 만족도에 미치는 영향을 살펴본 결과, 메시지 프레임은 비윤리적 행동 중단의도에 유의미한 영향을 미치지 않았지만, 실험 1 조건에서는 인공지능 챗봇에 대한 친밀감과 상호작용 만족도에는 영향을 미치는 것으로 나타났다. 최근 대두되고 있는 이슈인 온라인 강의 자동수강 프로그램의 경우 청소년들에게 다소 낮은 주제이지만, 무단결석은 누구나 잘 알고 있는 비윤리적인 행동이다. 시나리오 주제에 대한 청소년들의 지식 수준의 차이가 메시지 수용 결과에 차이를 보였을 수 있다. 지식 수준이 낮은 경우, 다양한 방면에서 정보를 습득하기보다는 제시된 메시지로 정보를 구성하려는 경향이 높아 제시된 메시지의 영향력이 더 크다(Cowley & Mitchell, 2003). 해당 정보에 대한 주관적 지식이 높은 경우, 내적 정보를 활용하는 반면, 주관적 지식이 낮은 경우 외적 정보를 많이 활용한다(Chaiken & Maheswaran, 1994). 이에 본 연구에서도 주관적 지식이 상대적으로 낮은 온라인 강의 자동수강 프로그램의 시나리오에 대해 외적 정보인 인공지능 챗봇 메시지를 적극적으로 활용하여 챗봇에 대한 친밀감과 만족도에 유의미한 영향을 미쳤다고 볼 수 있다.

구체적으로, 실험 1에서는 인공지능 챗봇이 부정 프레임으로 메시지를 전달했을 때, 청소년들은 챗봇에 대한 친밀감과 상호작용에서의 만족도를 높게 평가했다. 이는 히긴스(Higgins, 1998)의 조절초점이론(regulatory focus theory)으로 설명 가능하다. 이 이론에 따르면, 조절 초점은 항상 초점과 예방 초점으로 구분되는데, 항상 초점은 자신의 상황을 발전시키는 데 초점을 맞춰 긍정적인 결과에 관심을 두는 반면, 예방 초점은 자신의 상황을 보호하고 방어하는 것에 집중해 부정적인 결과를 회피하려는 경향이 있다(강영애·최승담, 2016; 최진명, 2013). 인공지능 챗봇의 부정 프레임 메시지는 부정적 결과나 손실을 강조하므로 예방 초점을 활성화시켜, 청소년들의 손실 회피 동기를 강화할 수 있다. 이처럼 강화된 동기는 챗봇의 조언을 더 가치있고 중요하다고 인식하게끔 만들어, 결과적으로 챗봇에 대한 친밀감과 만족도를 증가시킬 수 있다. 하지만 본 연구에서는 청소년들의 조절 초점 성향을 직접 측정하지 않았으므로, 향후 연구에서는 이러한 성향을 고려하여 연관성을 명확히 규명할 필요가 있다.

챗봇에 대한 만족도와 관련하여 부정 프레임 메시지 전달했을 때 특별한 의견 없음 조

건과 반대 조건에서 긍정 프레이밍 메시지에 비해 만족도가 높은 것을 확인했다. 이는 또래집단 규범이 비윤리적 행동에 대해 반대하거나 특별한 의견이 없는 경우, 개인의 도덕적 기준이 인공지능 챗봇의 부정 프레이밍 메시지와 일치하여 챗봇과의 상호작용에 대한 만족도가 높아졌기 때문일 수 있다. 이 경우, 청소년들은 챗봇을 사회적으로 인지하고 높은 만족감을 경험했을 가능성이 있다. 반대로, 비윤리적 행동을 찬성하는 또래집단 규범을 경험한 경우에는 부정 프레이밍의 챗봇 메시지가 또래집단의 긍정적인 태도와 상충되었을 가능성이 있으며, 이러한 상충이 인지부조화를 일으켜 챗봇에 대한 만족감을 감소시켰을 것으로 추정된다. 이러한 결과는 챗봇의 메시지 프레이밍과 또래집단 규범 간의 상호작용이 청소년들의 챗봇에 대한 만족감에 영향을 미칠 수 있음을 시사하며, 챗봇 메시지의 내용과 방식이 청소년들의 인식과 감정에 중요한 역할을 한다는 점을 보여준다.

마지막으로, 비윤리적 행동에 대한 또래집단 규범 조건과 인공지능 챗봇 메시지 프레이밍의 상호작용효과를 분석한 결과, 비윤리적 행동 중단 의도, 챗봇에 대한 친밀감, 상호작용에 대한 만족도에서 모두 유의미한 상호작용 효과가 나타나지 않았다. 이는 인공지능 챗봇의 메시지 프레이밍이 청소년들의 도덕적 판단에 큰 영향을 미치지 못했기 때문이라고 볼 수 있다. 메시지 프레이밍이 청소년들의 도덕적 가치관이나 기준에 설득력이 없었을 가능성이 존재한다는 것이다. 가령 메시지 프레이밍이 또래집단 내에서 인식되지 않았거나 무시되었을 가능성이 있을 뿐만 아니라 자극물로 조작한 메시지 프레이밍이 또래집단의 문화나 가치와 맞지 않았을 수 있다. 즉, 메시지 프레이밍은 청소년들의 도덕적 가치관이나 기준에 설득력이 없었을 수 있다는 것이다. 또한, 특정 집단과 사회에는 보편적으로 여겨지는 명령적 규범과 기술적 규범이 존재하지만, 특정 개인은 규범에 대해 다르게 인식할 수 있다. 가령 청소년 개인의 신념이 외부의 영향보다 행동 결정에 더 큰 영향을 미친다는 것이다. 박과 스미스(Park & Smith, 2007)는 규범의 차원을 개인적 수준에서 인지되는 규범과 사회적 수준에서 인지되는 규범으로 구분해 연구할 필요가 있다고 주장했다. 따라서 향후 연구에서는 사회규범의 차원을 세부적으로 구분하고 그 영향력을 살펴볼 필요가 있다. 나아가 또래집단 규범과 메시지 프레이밍의 상호작용 효과가 유의미하게 나타나지 않은 결과는 다양한 요인들이 복합적으로 작용하였다고 볼 수 있다. 이 같은 결과를 심층적으로 이해하기 위해서는 향후 연구에서 청소년들의 도덕적 신념과 같은 특성 등을 고려할 필요가 있을 것이다.

## 2) 한계 및 의의

본 연구는 몇 가지 한계점을 갖고 있다. 첫째, 연구에서 사용된 비윤리적 행동 관련 시나리오와

관련하여 참여자들의 이전 경험을 조사하지 않았다는 점이다. 참여자들이 비윤리적 행동에 대해 개인적으로 경험하였거나 관여도가 높거나 심리적 거리감이 가까운 경우, 연구결과에 영향을 미칠 수 있다. 예컨대, 참여자가 비윤리적 행동과 관련한 시나리오를 개인적으로 친숙하게 느끼거나 이에 대한 강한 의견을 가지고 있다면, 그들의 반응과 태도는 크게 달라질 수 있다. 그러므로 향후 연구에서는 참여자의 이전 경험과 관련한 요소를 측정하여 연구결과의 정확성을 높일 필요가 있다.

둘째, 참여자들이 인공지능 챗봇과의 자유롭게 대화를 나누지 못했다는 점도 한계로 볼 수 있다. 양방향 소통이 가능한 인공지능 챗봇과의 상호작용은 참여자의 관심과 참여를 높일 수 있으며(Portela & Granell-Canut, 2017; Tsai et al., 2021), 연구결과에 긍정적인 영향을 미칠 수 있다. 그러나 본 연구에서는 참여자들에게 인공지능 챗봇의 응답이 사진 형태로만 제공되었기 때문에 참여자들이 느낀 인공지능 챗봇에 대한 친밀감과 상호작용에 대한 만족도는 실제보다 낮게 평가되었을 가능성이 있다. 따라서 향후 연구에서는 참여자들이 인공지능 챗봇과 보다 자유롭게 상호작용할 수 있도록 실험 자극물을 설계하고, 참여자의 참여도를 높이는 방법을 모색하는 것이 필요하다.

이러한 한계에도 불구하고 본 연구는 인공지능 챗봇의 역할과 영향력에 대한 중요한 시사점을 제공한다. 먼저, 현재 인공지능 윤리에 대한 논의가 지속적으로 진행되고 있는 가운데 청소년과 인공지능 챗봇과의 상호작용을 통해 윤리적 문제를 재조명했다는 점에서 의의를 가진다. 또한, 본 연구는 인공지능 챗봇과의 상호작용이 개인의 윤리 영역에까지 적용될 수 있는지에 대한 가능성을 탐구함으로써 인공지능 분야의 연구를 확장했다는 점에서 이론적인 함의를 가진다. 나아가 청소년을 대상으로 한 연구에서 인공지능 챗봇이 또래집단 규범과 같은 사회적 요인과 함께 청소년의 도덕적 의사결정과 태도 형성에 긍정적인 영향을 미칠 수 있다는 점이 실증적으로 확인되었다. 이는 인공지능 챗봇이 단순한 정보 제공이나 상담 도구를 넘어서 윤리적 가치와 태도를 교육하는 데 중요한 역할을 할 수 있음을 시사한다. 특히, 청소년과 인공지능 챗봇 간의 상호작용에 관한 기존 연구는 주로 학습 효과 또는 정신건강에만 초점을 맞추고 있어 인공지능 챗봇의 설득효과에 대한 실증적 연구가 부족했다는 점을 고려할 때, 본 연구는 청소년들이 경험할 수 있는 비윤리적 상황에서 또래집단 규범이라는 요인을 포함해 설득효과를 살펴보았다는 점에서 의의가 있다.

실무적인 관점에서 본 연구 결과는 청소년 윤리 교육 프로그램을 개발하는 실무자들에게 유용한 시사점을 제공한다. 최근 정규교육과정인 초등교육과정부터 고등교육과정까지 인공지능 활용 수업이 증가하고 있다. 청소년들이 인공지능을 접할 수 있는 기회가 확대됨에 따라 인공지

능 교육 프로그램의 효과에 관한 연구 또한 진행되고 있다(강윤경·장유나·배상영·홍세희, 2021; 이문형·김재웅, 2020). 그러나 인공지능이 청소년에게 윤리적인 관점에서 미칠 수 있는 영향에 관한 논의는 충분히 진행되지 않았다. 교육적 관점에서 청소년의 인공지능 활용이 증가하고 있으므로 본 연구의 결과는 인공지능 챗봇의 효과적인 활용 방안을 모색하는 데 도움이 될 수 있을 것이다. 예를 들어, 인공지능 챗봇이 청소년에게 비윤리적 행동에 대한 조언을 전달한다고 할 때 고려해야 할 요소는 다음과 같다. 먼저, 청소년이 직면할 수 있는 상황을 구체적으로 묘사해야 할 것이다. 실제 청소년이 경험할 수 있는 사례를 현실적인 예시와 함께 스토리텔링 형식으로 제시할 필요가 있다. 또한, 인공지능 챗봇 메시지에 감정적인 요소를 포함할 수도 있다. 가령 ‘힘들지만 숙제를 모두 완료하고 나면 앞으로 무슨 일이든 해낼 수 있다는 자신감과 긍정적인 기분이 지속될 거라고 장담해요.’ 혹은 ‘숙제를 조작하면 스스로 비양심적인 행동을 했다는 자책감과 부끄러움이 생기지 않을까요?’와 같은 인공지능 챗봇 메시지를 청소년들이 감정적으로 공감할 수 있도록 구성하는 방법이 있다. 추가로, ‘합시다’, ‘할까요’ 등의 청유형 어미로 문장을 마감하는 촉진적 메시지는 발화자의 확신과 존재감이 내포되어있는 만큼 공감대 형성을 용이하게 하여 듣는 사람으로 하여금 소구력을 크게 느끼도록 하는 것으로 나타났다(문찬, 2008; 이정현, 2010). 이와 같은 방식으로 인공지능 챗봇 메시지를 구성한다면 청소년들에게 사회적으로 바람직한 방향으로 효과적인 설득을 할 수 있을 것이다. 결론적으로 본 연구는 인공지능 챗봇이 청소년들에게 윤리적 가치 및 태도를 전달하고 윤리 교육에 기여할 수 있는 방법을 심도있게 탐구했다는 점에서 사회적 영향을 미칠 수 있는 중요한 이론적 및 실무적 함의를 지니고 있다고 하겠다.

## References

- Ahn, H., & Lee, S. (2020). Relative persuasive effects of gain/loss framing in messages to help poorly-fed children. *Korean Journal of Broadcasting and Telecommunication Studies*, 34(5), 78-107. [안혜영·이승조 (2020). 결식아동 돕기를 촉구하는 메시지에서 나타나는 이익/손실 프레임의 상대적 설득 효과. <한국방송학회>, 34권 5호, 78-107.]
- Ahn, Y., Zhang, Y., Park, Y., & Lee, J. (2020, April). *A chatbot solution to chat app problems: Envisioning a chatbot counseling system for teenage victims of online sexual exploitation*. Paper presented at the CHI Conference on Human Factors in Computing Systems, Honolulu, HI.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, 82(2), 261-277.
- Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica*, 65(5), 1005-1027.
- Baumgartner, S. E., Valkenburg, P. M., & Peter, J. (2011). The influence of descriptive and injunctive peer norms on adolescents' risky sexual online behavior. *Cyberpsychology, Behavior, and Social Networking*, 14(12), 753-758.
- Berkowitz, A. D. (2005). An overview of the social norms approach: Changing the culture of college drinking. In L. P. Lederman & L. P. Stewart (Eds.), *Challenging the culture of college drinking: A socially situated health communication campaign* (pp. 193-214). New York, NY: Hampton Press.
- Berndt, T. J. (1979). Developmental changes in conformity to peers and parents. *Developmental Psychology*, 15(6), 608-616.
- Blader, S. L., & Tyler, T. R. (2009). Testing and extending the group engagement model: Linkages between social identity, procedural justice, economic outcomes, and extrarole behavior. *Journal of Applied Psychology*, 94(2), 445-464.
- Brown, B. B., Clasen, D. R., & Eicher, S. A. (1986). Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Developmental Psychology*, 22(4), 521-530.
- Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, 14(3), 417-431.

- Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risk: Pathways of youth in our time*. New York, NY: Cambridge University Press.
- Cesario, J., Grant, H., & Higgins, E. T. (2004). Regulatory fit and persuasion: Transfer from “feeling right”. *Journal of Personality and Social Psychology*, 86(3), 388-404.
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460-473.
- Choi, E., & Park, N. (2023). The effect of AI customization on chatbot users’ intentions to stop using offensive language and to continue to use AI chatbot: The mediating role of psychological reactance. *Korean Journal of Journalism & Communication Studies*, 67(2), 5-45. [최어진·박남기 (2023). 인공지능 맞춤형 챗봇 이용자의 욕설 사용 중단의도와 챗봇 지속이용의도에 미치는 영향: 심리적 반발의 매개효과. <한국언론학보>, 67권 2호, 5-45.]
- Choi, J. (2013). The impacts of consumers’ regulatory focus on the adoption of means for goal achievement. *Journal of Marketing Studies*, 21(2), 45-58. [최진명 (2013). 조절초점 동기가 의사결정의 수단 선택에 미치는 영향. <마케팅논집>, 21권 2호, 34-58.]
- Clasen, D. R., & Brown, B. B. (1985). The multidimensionality of peer pressure in adolescence. *Journal of Youth and Adolescence*, 14(6), 451-468.
- Cowley, E., & Mitchell, A. A. (2003). The moderating effect of product knowledge on the learning and organization of product information. *Journal of Consumer Research*, 30(3), 443-454.
- Crutzen, R., Peters, G. J. Y., Portugal, S. D., Fisser, E. M., & Grolleman, J. J. (2011). An artificially intelligent chat agent that answers adolescents’ questions related to sex, drugs, and alcohol: An exploratory study. *Journal of Adolescent Health*, 48(5), 514-519.
- Erlandsson, A., Nilsson, A., & Västfjäll, D. (2018). Attitudes and donation behavior when reading positive and negative charity appeals. *Journal of Nonprofit & Public Sector Marketing*, 30(4), 444-474.
- Grove, C. (2021). Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in Psychiatry*, 11, 606041.
- Han, D. (2021). An analysis of Korean EFL learners’ experience on English classes using AI chatbot. *Robotics & AI Ethics*, 6(3), 1-9.
- Hertel, G., & Kerr, N. L. (2001). Priming in group favoritism: The impact of normative scripts in the minimal group paradigm. *Journal of Experimental Social Psychology*, 37(4), 316-324.



- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology*, 30, 1-46.
- Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245-250.
- Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, 16(1), 7-30.
- Hong, S. M., & Faedda, S. (1996). Refinement of the Hong psychological reactance scale. *Educational and Psychological Measurement*, 56(1), 173-182.
- Jeong, I. (2019). The influence of norms in social media on the recognition of offline norms. *Korean Journal of Journalism & Communication Studies*, 63(4), 123-158. [정일권 (2019). 소셜미디어에서의 직·간접적 탈규범 경험이 청소년의 사회규범 인식에 미치는 영향. <한국언론학보>, 63권 4호, 123-158.]
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Kang, Y., & Choi, S. (2016). The message framing effect on visitors' environmentally responsible behavior: Focus on regulatory focus. *The Tourism Sciences Society of Korea*, 40(5), 217-232. [강영애·최승담 (2016). 메시지 프레임이 환경적으로 책임 있는 방문객 행동에 미치는 영향: 조절초점을 중심으로. <관광학연구>, 40권 5호, 217-232.]
- Kang, Y., Jang, Y., Bae, S., & Hong, S. (2021). A meta-analysis of the effects of artificial intelligence education programs on adolescents. *Journal of Korean Association for Educational Information and Media*, 27(4), 1273-1294. [강윤경·장유나·배상영·홍세희 (2021). 인공지능 교육프로그램이 청소년에게 미치는 효과 연구에 대한 메타분석. <교육정보미디어연구>, 27권 4호, 1273-1294.]
- Kim, S., & Yu, M. (2010). Effects of message framing on the advertising effectiveness: Moderating role of regulatory focus and product type. *The Journal of Korea Content Association*, 10(8), 177-185. [김성재·유명길 (2010). 메시지 프레임이 광고효과에 미치는 영향: 조절초점과 제품유형의 조절효과. <한국콘텐츠학회논문지>, 10권 8호, 177-185.]
- Kim, T., & Yoon, S. (2022). A study on drinking temperance intention and drinking culture of Korean high school students based upon the Theory of Reasoned Action (TRA): Focused on the comparison of different school characteristics. *Journal of Cultural Industry Studies*, 22(1), 49-65. [김태경·윤성준

- (2022). 합리적행위이론(TRA)을 적용한 한국 고등학생들의 음주절제의도와 음주문화에 관한 연구: 학교 특성 간 차이 비교를 중심으로. <문화산업연구>, 22권 1호, 49-65.]
- Kim, Y. (2024, March 25). *Obscene AI chatbots exposed to adolescents*. The Seoul Shinmun Daily. Retrieved 5/16/24 from <https://www.seoul.co.kr/news/society/2024/03/25/20240325001005> [김예슬 (2024, 3, 25). 음란 AI 챗봇, 청소년에 무방비 노출. <서울신문>.]
- Koo, M. (2022, October 20). *Ministry of Gender Equality and Family operates 'Youth self-counseling' online content 'Solrobot'*. Newsis. Retrieved 1/3/24 from [https://mobile.newsis.com/view.html?ar\\_id=NISX20221019\\_0002053968](https://mobile.newsis.com/view.html?ar_id=NISX20221019_0002053968) [구무서 (2022, 10, 20). 여가부, '청소년 자가상담' 온라인 콘텐츠 솔로봇 운영. <뉴스시스>.]
- Kooli, C. (2023). Artificial Intelligence Dissociative Identity Disorder (AIDIS): The dark side of ChatGPT. *QScience Connect*, 2023(2), 2.
- Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 75(1), 23-55.
- Kwon, B. (2022). Differences in the intention of elementary school students for eco-friendly behavior according to message endings and sources: Focused on zero waste. Unpublished doctoral dissertation, Seoul National University, Seoul, Korea. [권보경 (2022). <메시지 어미와 정보 전달자에 따른 초등학생의 친환경 행동 의도 차이: 제로웨이스트를 중심으로>. 서울대학교 대학원 박사학위 논문.]
- Lally, P., Bartle, N., & Wardle, J. (2011). Social norms and diet in adolescents. *Appetite*, 57(3), 623-627.
- Lee, B., & Yoon, T. (2012). A study on the framing effect in public service advertising a meta-analysis. *The Korean Journal of Advertising and Public Relations*, 14(2), 33-60. [이병관·윤태웅 (2012). 공익광고의 프레임 효과에 관한 연구 메타 분석. <한국광고홍보학보>, 14권 2호, 33-60.]
- Lee, D., & Kim, H. (2023). Effects of youth-targeted antismoking public service announcements: The moderating role of a sense of belonging to a peer group in persuasion and priming effects. *Journal of Communication Research*, 60(2), 5-46. [이다은·김현석 (2023). 청소년 흡연 예방 공익광고의 설득 및 점화 효과: 친구집단 소속감의 조절 역할. <언론정보연구>, 60권 2호, 5-46.]
- Lee, J. (2010). Message structure of persuasive storytelling in travel guide booklets of great Chungchung visit year. *Journal of the Korean Content Association*, 10(11), 380-390. [이정현 (2010). 관광안내서의 스토리텔링적 방문 설득 메시지 구조: 대충청 방문의 해 공식 안내서를 중심으로. <한국콘텐츠학회논문지>, 10권 11호, 380-390.]
- Lee, J. (2021, May 4). "If you're depressed, talk to 'Ash'"... 'Chatbot' preventing adolescent suicide. EBS

News. Retrieved 7/3/23 from <https://post.naver.com/viewer/postView.nhn?volumeNo=31420204&memberNo=10234477&vType=VERTICAL> [이지예 (2021, 5, 4). “우울하다면 ‘애쉬’에게 말을 거세요”...청소년 자살 막는 ‘챗봇’. <EBS뉴스>.]

Lee, M., & Kim, J. (2020). A study on the teaching of digital literacy in the AI age: Focused on Berkman Klein Center education. *Cartoon and Animation Studies*, 59, 533-555. [이문형·김재웅 (2020). AI시대의 청소년 디지털 리터러시 교육 연구: Berkman Klein Center 교육을 중심으로. <만화애니메이션 연구>, 59호, 533-555.]

Lee, S. (2017). The influence of adolescents' sensation seeking and peer conformity on cyberbullying: Mediation effect of anonymity. Unpublished doctoral dissertation, Daegu Haany University, Gyeongsan, Korea. [이수봉 (2017). <청소년의 감각추구 성향과 또래 동조성이 사이버불링에 미치는 영향: 익명성의 매개효과>. 대구한의대학교 대학원 박사학위 논문.]

Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human-Computer Interaction*, 36(10), 930-940.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149-188.

Lin, M. P.-C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot: A mixed-method classroom study. *Journal of Educational Technology & Society*, 23(1), 78-92.

Mariamo, A., Temcheff, C. E., Léger, P. M., Senecal, S., & Lau, M. A. (2021). Emotional reactions and likelihood of response to questions designed for a mental health chatbot among adolescents: Experimental study. *JMIR Human Factors*, 8(1), e24343.

Moon, C. (2008). Approach of storytelling skill to the product design education focused on marketing. *Journal of Product Research*, 26(4), 25-35. [문찬 (2008). 스토리텔링 기법을 적용한 마케팅 중심의 제품디자인 교육. <상품학연구>, 26권 4호, 25-35.]

Muhammad, A. Q., Jiang, Y. A., Shubin, Y. R., & Sandra, M. C. (2020). I, chat-bot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, 22(1), 101-113.

Nam, M., & Cho, M. (2012). The influence of restaurant customers' chronic regulatory focus, message framing, and situational involvement on visit intention. *Journal of Tourism Studies*, 24(4), 41-67. [남

- 민정·조민호 (2012). 외식고객의 성향적 조절초점, 메시지를 및 관여도가 레스토랑 방문의도에 미치는 영향. <관광연구논총>, 24권 4호, 41-67.]
- Nan, X. (2007). Social distance, framing, and judgment: A construal level perspective. *Human Communication Research*, 33(4), 489-514.
- Nesi, J., Rothenberg, W. A., Hussong, A. M., & Jackson, K. M. (2017). Friends' alcohol-related social networking site activity predicts escalations in adolescent drinking: Mediation by peer norms. *Journal of Adolescent Health*, 60(6), 641-647.
- Norton, J., & Crowson, I. (2024, January 6). *Father of Molly Russell warns parents to beware of character: AI chatbot that 'mocks' young people with depressive thoughts*. Daily Mail. Retrieved 2/1/24 from <https://www.dailymail.co.uk/news/article-12934029/Father-Molly-Russell-warns-parents-beware-Character-AI-chatbot-mocks-young-people-depressive-thoughts.html>
- O'Keefe, D. J., & Jensen, J. D. (2009). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease detection behaviors: A meta-analytic review. *Journal of Communication*, 59(2), 296-316.
- Oh, A., Lee, H., & Kim, H. (2019). The effectiveness of suicide prevention campaign for adolescents by message framing and content types. *Advertising Research*, 122, 35-70. [오아름·이현진·김활빈 (2019). 메시지 프레임과 콘텐츠 유형에 따른 청소년 자살 예방 캠페인 효과에 관한 연구. <광고연구>, 122권, 35-70.]
- Park, H. (2014). Communication message strategy to promote disease prevention behavior: The effects of temporal distance, social distance and gain-loss frame in the cervical cancer. *Korean Journal of Journalism & Communication Studies*, 58(3), 344-377. [박현정 (2014). 질병 예방행동 촉진을 위한 커뮤니케이션 메시지 전략 연구: 자궁경부암에서 심리적 거리, 해석수준, 이득-손실 프레임을 중심으로. <한국언론학보>, 58권 3호, 344-377.]
- Park, H. S., & Smith, S. W. (2007). Distinctiveness and influence of subjective norms, personal descriptive and injunctive norms, and societal descriptive and injunctive norms on behavioral intent: A case of two behaviors critical to organ donation. *Human Communication Research*, 33(2), 194-218.
- Park, H., Deguchi, T., & Yoshida, T., (2007). The effects of self-norms and consciousness peer group norms on individual behavior at class adaptation. *Korean Journal of Social and Personality Psychology*, 21(3), 13-25. [박현정·타쿠히코 데구치·토시카츠 요시다 (2007). 학급 적응감에 있어서 규범의식이 행동에 미치는 영향: 교사의 학생지도유형을 중심으로. <한국심리학회지: 사회 및 성격>, 21

권 3호, 13-25.]

- Park, N. (2020). Artificial intelligence and ethical issues. *Journal of Communication Research*, 57(3), 122-154. [박남기 (2020). 인공지능과 윤리적 이슈. <언론정보연구>, 57권 3호, 122-154.]
- Park, N., Jung, Y., & Lee, K. M. (2011). Intention to upload video content on the internet: The role of social norms and ego-involvement. *Computers in Human Behavior*, 27(5), 1996-2004.
- Parks, M. R., & Floyd, K. (1996). Meanings for closeness and intimacy in friendship. *Journal of Social and Personal Relationships*, 13(1), 85-107.
- Pedersen, E. R., Osilla, K. C., Miles, J. N., Tucker, J. S., Ewing, B. A., Shih, R. A., & D'Amico, E. J. (2017). The role of perceived injunctive alcohol norms in adolescent drinking behavior. *Addictive Behaviors*, 67, 1-7.
- Portela, M., & Granell-Canut, C. (2017, September). *A new friend in our smartphone?: Observing interactions with chatbots in the search of emotional engagement*. Paper presented at the XVIII International Conference on Human Computer Interaction, Cancun, Mexico.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Stanford, CA: CSLI Publications.
- Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin*, 121(1), 3-19.
- Santor, D. A., Messervey, D., & Kusumakar, V. (2000). Measuring peer pressure, popularity, and conformity in adolescent boys and girls: Predicting school performance, sexual attitudes, and substance abuse. *Journal of Youth and Adolescence*, 29(2), 163-182.
- Sasson, H., & Mesch, G. (2014). Parental mediation, peer norms and risky online behavior among adolescents. *Computers in Human Behavior*, 33, 32-38.
- Seo, J., Lee, J., & Park, M. (2011). High school students' buying attitudes toward school uniform brands according to clothing conformity. *The Research Journal of the Costume Culture*, 19(6), 1320-1333. [서지민·이지연·박명자 (2011). 중, 고등학생의 의복 동조 성향에 따른 교복 브랜드에 대한 태도와 구매행동. <복식문화연구>, 19권 6호, 1320-1333.]
- Shu, S., & Choi, Y. (2018). The effects of message framing and evidence type on obesity prevention for chinese adolescents. *Journal of the Korea Content Association*, 18(3), 626-635. [서소상·최유진 (2018). 메시지 프레임과 증거 유형에 따른 중국 청소년 비만 예방 메시지 효과. <한국콘텐츠학회논문지>, 18권 3호, 626-635.]

- Steinberg, L., & Morris, A. S. (2001). Adolescent development. *Journal of Cognitive Education and Psychology*, 2(1), 55-87.
- Sung, M. (2023, June 1). *While parents worry, teens are bullying Snapchat AI*. Tech Crunch. Retrieved 1/29/24 from <https://techcrunch.com/2023/05/31/people-keep-gaslighting-snapchat-my-ai/>
- Tsai, W. H. S., Liu, Y., & Chuan, C. H. (2021). How chatbots' social presence communication enhances consumer engagement: The mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing*, 15(3), 460-482.
- Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., & Piterova, A. (2022). An artificial intelligence chatbot for young people's sexual and reproductive health in India (SnehAI): Instrumental case study. *Journal of Medical Internet Research*, 24(1), e29969.
- Xie, Y., Zhu, K., Zhou, P., & Liang, C. (2023). How does anthropomorphism improve human-AI interaction satisfaction: A dual-path model. *Computer in Human Behavior*, 148, 107878.
- Yang, M., & Kang, H. (2011). A study on the actual conditions of abusive language used by juveniles and juveniles' attitude towards using abusive language. *Eomunhak*, 111, 57-87. [양명희·강희숙 (2011). 초·중·고 학생들의 욕설 사용 실태와 태도에 대한 연구. <어문학>, 111권, 57-87.]
- Yanovitzky, I., Stewart, L. P., & Lederman, L. C. (2006). Social distance, perceived drinking by peers, and alcohol use by college students. *Health Communication*, 19(1), 1-10.
- Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 28(11), 15223-15243.
- Zogaj, A., Mähner, P. M., Yang, L., & Tscheulin, D. K. (2023). It's a match! The effects of chatbot anthropomorphization and chatbot gender on consumer behavior. *Journal of Business Research*, 155, 113412.

최초 투고일 2024년 02월 02일

게재 확정일 2024년 05월 27일

논문 수정일 2024년 05월 29일

## 부록 1-1. 실험 1의 비윤리적 행동에 대한 또래집단 규범 조건 시나리오 자극물

### [비윤리적 행동에 대한 긍정 조건]

당신은 학교에서 'EBS의 온라인 강의를 수강하라는 숙제를 받았습시다. 그러나 당신에게 주어진 숙제는 이것만이 아닙니다. 제출 기한이 며칠 남지 않은 다른 과목의 숙제들도 아직 끝나치지 못한 상황입니다. 이미 부족한 잠 시간을 더 줄여 무리를 해야만 모든 숙제를 마칠 수 있을 것 같아 걱정입니다. 그러던 중, 당신은 우연히 뉴스 기사를 통해 매크로를 이용한 온라인 강의 자동수강 현상을 접하게 됐습시다. 학생들이 실제로 강의를 보지 않고도 매크로로 코드를 조작해 '수강 완료' 표시를 인증받는다는 것입니다.

이와 관련하여 반 친구들에게 묻자, 학생 30명 중 25명은 매크로를 통한 수강 조작을 당연하게 생각하고 있다고 답했습니다. 강의 매크로를 이용할지, 정정당당하게 숙제를 마칠지 고민이 이어집니다. 어떤 선택이 맞는지 해답을 얻고 싶은 당신은 인공지능 챗봇과 대화를 시도합니다.

### [비윤리적 행동에 대한 부정 조건]

당신은 학교에서 'EBS의 온라인 강의를 수강하라는 숙제를 받았습시다. 그러나 당신에게 주어진 숙제는 이것만이 아닙니다. 제출 기한이 며칠 남지 않은 다른 과목의 숙제들도 아직 끝나치지 못한 상황입니다. 이미 부족한 잠 시간을 더 줄여 무리를 해야만 모든 숙제를 마칠 수 있을 것 같아 걱정입니다. 그러던 중, 당신은 우연히 뉴스 기사를 통해 매크로를 이용한 온라인 강의 자동수강 현상을 접하게 됐습시다. 학생들이 실제로 강의를 보지 않고도 매크로로 코드를 조작해 '수강 완료' 표시를 인증받는다는 것입니다.

이와 관련하여 반 친구들에게 묻자, 학생 30명 중 25명은 매크로를 통한 수강 조작을 나쁘게 생각하고 있다고 답했습니다. 강의 매크로를 이용할지, 정정당당하게 숙제를 마칠지 고민이 이어집니다. 어떤 선택이 맞는지 해답을 얻고 싶은 당신은 인공지능 챗봇과 대화를 시도합니다.

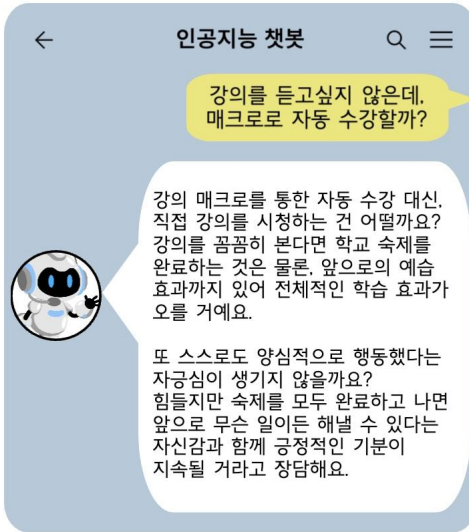
### [비윤리적 행동에 대한 특별한 의견 없음 조건]

당신은 학교에서 'EBS의 온라인 강의를 수강하라는 숙제를 받았습시다. 그러나 당신에게 주어진 숙제는 이것만이 아닙니다. 제출 기한이 며칠 남지 않은 다른 과목의 숙제들도 아직 끝나치지 못한 상황입니다. 이미 부족한 잠 시간을 더 줄여 무리를 해야만 모든 숙제를 마칠 수 있을 것 같아 걱정입니다. 그러던 중, 당신은 우연히 뉴스 기사를 통해 매크로를 이용한 온라인 강의 자동수강 현상을 접하게 됐습시다. 학생들이 실제로 강의를 보지 않고도 매크로로 코드를 조작해 '수강 완료' 표시를 인증받는다는 것입니다.

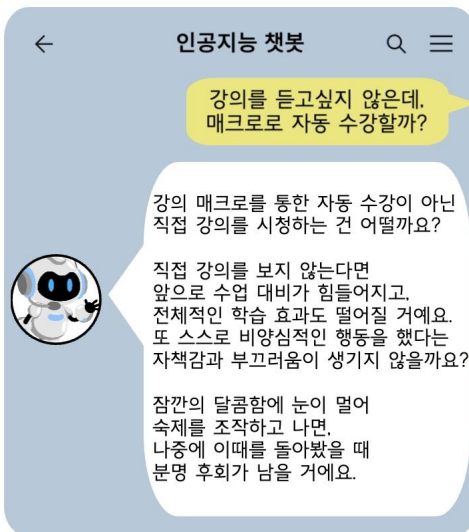
이와 관련하여 반 친구들에게 묻자, 학생 30명 중 25명은 매크로를 통한 수강 조작에 별다른 의견이 없다고 답했습니다. 강의 매크로를 이용할지, 정정당당하게 숙제를 마칠지 고민이 이어집니다. 어떤 선택이 맞는지 해답을 얻고 싶은 당신은 인공지능 챗봇과 대화를 시도합니다.

## 부록 1-2. 실험 1의 비윤리적 행동에 대한 또래집단 규범 조건 시나리오

### [인공지능 챗봇의 긍정 메시지 프레임]



### [인공지능 챗봇의 부정 메시지 프레임]





## 부록 II-1. 실험 2의 비윤리적 행동에 대한 또래집단 규범 조건 시나리오 자극물

### [비윤리적 행동에 대한 긍정 조건]

당신의 학교에서 대규모 행사가 있는 날, 일부 친구들이 무단결석 후 함께 놀이공원에 가자고 제안합니다. 친구들은 지난 행사 때도 선생님께서 알아채시지 못했이라며, 이번에도 문제가 없을 거라고 말합니다. 당신이 생각하기에도 복잡한 공간의 많은 인파 속에서 몇 명이 없어진 것을 눈치채기 어려운 상황인데다, 오늘은 더 이상 선생님께서 인원 파악을 하하지 않을 것 같습니다. 지난 행사와 마찬가지로 오늘도 마침 지루하다고 느끼고 있던 당신은 친구들의 제안에 솔깃해집니다. 앞서 반 친구들과의 대화를 통해 30명 중 25명은 대규모 행사 중 무단결석이 문제가 없다고 생각하고, 심지어 모두들 비슷한 경험이 있다는 것을 알았습니다. 이때 당신은 무단결석 후 놀이공원에 갈지, 학교에 남을지 결정을 내리기 위해 인공지는 챗봇과 대화를 시도합니다.

### [비윤리적 행동에 대한 부정 조건]

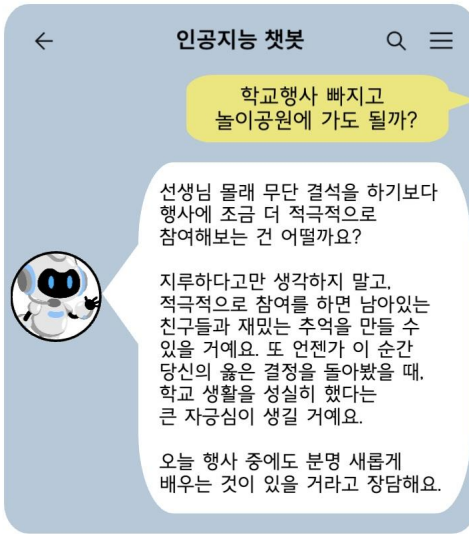
당신의 학교에서 대규모 행사가 있는 날, 일부 친구들이 무단결석 후 함께 놀이공원에 가자고 제안합니다. 친구들은 지난 행사 때도 선생님께서 알아채시지 못했이라며, 이번에도 문제가 없을 거라고 말합니다. 당신이 생각하기에도 복잡한 공간의 많은 인파 속에서 몇 명이 없어진 것을 눈치채기 어려운 상황인데다, 오늘은 더 이상 선생님께서 인원 파악을 하하지 않을 것 같습니다. 지난 행사와 마찬가지로 오늘도 마침 지루하다고 느끼고 있던 당신은 친구들의 제안에 솔깃해집니다. 앞서 반 친구들과의 대화를 통해 30명 중 25명은 대규모 행사 중 무단결석이 옳지 않다고 생각하고 이러한 경험이 없다는 것을 알았습니다. 이때 당신은 무단결석 후 놀이공원에 갈지, 학교에 남을지 결정을 내리기 위해 인공지는 챗봇과 대화를 시도합니다.

### [비윤리적 행동에 대한 특별한 의견 없음 조건]

당신의 학교에서 대규모 행사가 있는 날, 일부 친구들이 무단결석 후 함께 놀이공원에 가자고 제안합니다. 친구들은 지난 행사 때도 선생님께서 알아채시지 못했이라며, 이번에도 문제가 없을 거라고 말합니다. 당신이 생각하기에도 복잡한 공간의 많은 인파 속에서 몇 명이 없어진 것을 눈치채기 어려운 상황인데다, 오늘은 더 이상 선생님께서 인원 파악을 하하지 않을 것 같습니다. 지난 행사와 마찬가지로 오늘도 마침 지루하다고 느끼고 있던 당신은 친구들의 제안에 솔깃해집니다. 앞서 반 친구들과의 대화를 통해 30명 중 25명은 대규모 행사 중 무단결석에 별다른 의견이 없다는 것을 알았습니다. 이때 당신은 무단결석 후 놀이공원에 갈지, 학교에 남을지 결정을 내리기 위해 인공지는 챗봇과 대화를 시도합니다.

## 부록 II-2. 실험 2의 비윤리적 행동에 대한 또래집단 규범 조건 시나리오

### [인공지능 챗봇의 긍정 메시지 프레임]



### [인공지능 챗봇의 부정 메시지 프레임]

