미래사회
Journal of Future Society
2026. Vol. 17, No. 3, 54～84.
https://doi.org/10.22987/jifso.2026.17.3.54

AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점에 관한 질적 내용 분석 연구

정 진*

서울사이버대학교

본 연구의 목적은 AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점을 폭넓게 이해하는 데 있다. 이를 위하여 AI(인공지능)와 인간상담자에 의한 심리상담을 모두 경험한 내담자 7인을 대상으로 심층 인터뷰를 실시하였으며, 인터뷰 자료를 질적 내용 분석 방법을 활용하여 분석하였다. 분석 결과, 14개의 구성요인 및 28개의 하위 구성요인이 도출되었다. 연구결과를 종합하면, AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점은 "접근성이 용이하여 심리적 도움을 쉽게 얻을 수 있지만 비밀보장과 책임성 등의 윤리적 신뢰의 필요성을 지각함"으로 요약할 수 있다. 본 연구는 AI를 활용한 심리상담 과정에서 내담자가 경험하는 다양한 윤리적 쟁점을 상담윤리의 관점에서 체계적으로 탐색함으로써, 향후 상담 현장에서 인공지능을 활용할 때 고려해야 할 윤리적 기준과 실천적 방향을 모색하기 위한 기초자료를 제공할 수 있을 것으로 기대한다.

주요어: AI(인공지능), 심리상담, 키츠너, 상담윤리, 질적 내용 분석

* 주저자: 정진/서울사이버대학교 분당심리상담센터 특임교수/경기도 분당구 성남대로 884-9(야탑동) 2층
/Tel: 02-944-5932/ E-mail: jin8021@hanmail.net

Ⅰ. 서론

인공지능(Artificial Intelligence; 이하 AI) 기술은 최근 대규모 언어모델과 자연어 처리 기술의 발전을 기반으로 인간과 유사한 대화 수행 능력을 구현하며 다양한 전문직 영역으로 확장하고 있다(Hsu, S. L., et al., 2025; Maurya, R. K., 2024; 강태정 외, 2024; 김태완 외, 2024; 여수진 외, 2025; 황매향, 2025). 특히 보건·의료 및 정신건강 서비스 영역에서는 AI 기반 챗봇, 디지털 치료제, 정서 분석 시스템 등이 도입되며 심리상담의 접근 방식에 구조적 변화를 가져오고 있다(강수정 외, 2024; 김도연 외, 2020; 김요완, 2026; 김훈민 외, 2025). 이러한 변화는 단순한 기술 도입을 넘어 심리상담의 본질적 경험과 윤리적 경계의 재검토가 요구되고 있다(Ping, Y., 2024; 이승록 외, 2026; 송용섭, 2021; 정진, 2025).

급변하는 AI 기술 발달과 더불어 AI 기반 심리적 개입의 효과에 대한 검증이 활발히 이루어지고 있으며 예컨대, Fitzpatrick et al.(2017)은 AI 챗봇 기반 인지행동치료 프로그램이 대학생의 우울 및 불안 증상 감소에 유의한 효과를 보였음을 보고하였다. 또한, Inkster et al.(2018) 연구에서는 AI 대화형 상담 시스템이 정서적 지지 및 자기 보고식 스트레스 감소에 긍정적 영향을 미친다고 제시하였다. 이러한 국외 연구에서는 AI 기반 심리상담은 접근성 및 즉시성 측면에서 장점을 지니며 경증 수준의 심리적 어려움에 대해 보완적 자원으로 기능할 수 있음을 시사하고 있다.

이상의 AI 기반 심리상담에 대한 효과성 논의와는 별개로, AI 기반 심리상담은 본질적으로 윤리적 질문을 수반한다(강수정 외, 2024; 여수진 외, 2025; 이윤희 외, 2021; 정진, 2025). 심리상담은 인간의 심리적 취약성과 삶의 의미를 다루는 관계적 실천이며(Rogers, 1957), 공감, 진정성, 무조건적 긍정적 존중과 같은 관계적 요인이 치료적 변화의 핵심으로 간주된다(Rogers, 1957; 이지언 외, 2023). 이러한 맥락에서 비인격적 알고리즘이 수행하는 '공감적 반응'이 인간상담자의 관계적 공감과 동일한 윤리적 지위를 가질 수 있는지에 대한 논의는 지속적으로 제기되고 있다(Mittelstadt, B. D., et al., 2016; 안세훈 외, 2021; 이윤희 외, 2021; 정진, 2025).

상담윤리(Counseling Ethics)는 전문직 윤리의 한 영역으로, 내담자의 권리 보호와 복지 증진을 핵심 목적으로 한다(American Psychological Association, 2014/2017; 강수정 외, 2024). Kitchener(1984)는 상담윤리를 자율성 존중, 선행, 무해, 정의, 충실성의 다섯 원칙으로 체계화하였으며, 이는 American Counseling Association(2014)와 American Psychological Association(2024/2017), 그리고 한국상담심리학회 윤리강령(2018/2020)에도 반영되어 있다. 특히 국내 한국상담심리학회의 윤리강령(2018/2020)은 자율성 존중, 비밀보장, 선행, 무해, 신뢰성, 책임성, 정의의 원칙을 명시하고 있으며, 이는 상담 실천의 핵심 가치 체계를 형성하고 있다(Kitchener, K. S., 1984; 강수정 외, 2024). 특히 상담윤리에 대한 가장 기본적인 이론적 틀로서 Kitchener(1984)는 상담자의 윤리적 의사결정을 안내하는 보편적 도덕 원칙을 제시하고 있으며 자율성(autonomy), 선행(beneficence), 무해(nonmaleficence), 정의(justice), 충실성(fidelity)의 다섯 가지 원칙으로 체계화하여 소개하였다(Kitchener, K. S., 1984; 강수정 외, 2024; 이윤희 외 2021). 그는 상담윤리를 단순한 규정 준수의 차원이 아니라 도덕철학에 근거한 규범적 판단 체계로 이해하였으며, 상담자는 내담자의 자기 결정권을 존중하고(자율성), 내담자의 복지를 증진하며(선행), 해를 최소화하고(무해), 공정하고 평등하게 서비스를 제공하며(정의), 신뢰에 기반을 둔 책임 있는 관계를 유지해야 한다고(충실성) 보았다(Kitchener, K. S., 1984; 이윤희 외, 2021). 특히 Kitchener, K. S.(1984)는 실제 상담현장에서 윤리적 딜레마가 발생하면 이 다섯 원칙 간의 균형과 우선순위를 숙고하는 과정이 중요하다고 강조하였

으며, 윤리적 판단은 기계적 규칙 적용이 아니라 상황적 맥락을 고려한 도덕적 숙고의 과정임을 제시하였다. 이러한 원칙윤리 접근은 상담자의 전문적 책임과 도덕적 책무성을 동시에 강조한다는 점에서 현대 상담윤리의 이론적 토대를 형성하였으며, 전통적 대면상담뿐 아니라 기술 기반 상담환경에서도 윤리적 쟁점을 분석하는 유용한 개념적 틀을 제공한다(이윤희 외, 2021; 이소윤, 2025).

국내 상담윤리 연구는 주로 상담자와 내담자 간의 관계를 중심으로 발전해 왔다. 예컨대 김형수 외(2009)는 상담윤리 의사결정 모형을 제시하며 윤리적 판단 과정의 다각적 측면을 분석하였다. 송용섭(2021)은 상담관계에서 발생할 수 있는 개인정보와 책임 소재, 프라이버시 등의 윤리적 문제를 분석하고 있으며 장성식(2020)은 사전 동의(informed consent)의 실제적 이해 정도가 자율성 인식에 중요한 변인임을 보고하였다. 이와 같은 연구는 윤리 원칙이 인간 대 인간의 상담 관계 속에서 경험적으로 구현된다는 점을 확인할 수 있다.

또한, 온라인 상담 및 사이버 상담의 윤리 문제에 관한 연구도 축적되어 왔다. 강수정 외(2018)에서는 디지털 환경에서 개인정보 보호와 상담자-내담자 경계에 대한 윤리적 판단이 필요함을 제안하였으며, 강수정 외(2024)에서는 비대면 상담과 AI 축어록, 소셜미디어 등 디지털 미디어의 실제 활용 실태를 조사하여, 상담자의 정보보호, 사전 동의, 내담자 권리와 상담관계 등에 관한 윤리적 실천과 교육의 필요성을 시사하였다. 김도연 외(2020)에서는 해외 AI 활용 사례를 검토하고, AI 상담의 발전 가능성과 자연어처리 및 공감적 상호작용 등의 기술적 과제를 제안하고 있다. 이상의 연구에서 AI 기술을 매개로 한 상담환경에서 윤리적 재해석이 필요함을 시사하고 있지만, AI와 같은 비인격적 행위자가 상담을 수행하는 상황에 대한 논의는 제한적임을 알 수 있다.

AI 윤리 분야에서는 투명성, 책임성, 공정성, 설명 가능성을 핵심 원칙으로 제시한다(Floridi & Cowls, 2019). Garcia, E.(2021)는 UNESCO(2022)의 AI 윤리 권고안을 통하여 인간의 존엄성과 기본권 보호를 강조하고 있으나 이러한 기술적 담론은 주로 정책 및 시스템 설계 차원에 머무르며, 상담 관계적 맥락에서 내담자가 경험하는 윤리적 이슈를 충분히 설명하지 못하는 한계가 있다.

국내 AI 윤리에 관한 선행연구를 살펴보면, 국내 정보 및 AI 윤리 연구에서도 개인정보 보호와 알고리즘의 책임성 문제가 제기되고 있음을 알 수 있다. 김명주(2017)는 『정보와통신』에서 AI 알고리즘의 책임성과 투명성 확보 필요성을 강조하였으며, 김도연 외(2020)는 『한국심리학회지』에서 인공지능의 철학 자체보다는 상담자의 역할과 책임에 중점을 두고 실제 상담·심리치료에서 AI가 어떻게 활용되는지를 탐색하였다. 그러나 실제 상담 관계적 맥락에서의 구체적 경험을 탐색하지 못한 한계를 지닌다. 최근 국내에서는 디지털 치료제 및 AI 기반 정신건강 서비스에 관한 윤리적 탐색 연구도 수행되고 있다. 이동훈 외(2015)는 디지털 상담의 장점과 한계를 제시하고 있으며 김성현 외(2022)에서 정신건강 영역에서의 디지털 치료제의 동향과 활용 가능성에 대해서는 충분한 논의가 이루어지고 있으나 윤리적 관점에서는 체계적으로 탐색하지 못하고 있음을 확인할 수 있었다.

AI 상담 경험에 관한 선행연구를 살펴보면, 국내에서는 해당 주제에 관한 연구가 매우 미흡한 상황이나 국외 질적 연구에서 AI 상담 경험에 대한 탐색이 일부 보고되고 있다. Maurya, R. K.(2024)는 챗봇을 활용하여 상담자와 내담자가 어떠한 상담관계를 시연해 가는지에 대한 탐색하는 연구가 수행되기도 하였다. Keung, W. M., et al.(2025)은 AI 개입의 진정성(authenticity)에 대한 문제를 제기하며, 사용자가 정서적 지지를 경험하면서도 동시에 알고리즘의 한계를 인식한다고 분석하였다. 이상의 연구들은 AI 상담에서 나타나는 관계적 한계와 윤

리적 미흡성을 지적하고 있으나, 이러한 문제를 실질적으로 보완하기 위한 구체적인 대안과 개입 방안은 여전히 충분히 마련되지 않은 실정이다.

국내외 선행연구를 종합하면, AI 상담의 효과성과 기술적 적용에 관한 탐색은 활발하게 수행되고 있으나(김창옥 외, 2025; 주우영 외, 2024; 황매향, 2025), AI 기반 심리상담에서의 내담자의 윤리적 경험을 탐색한 연구가 매우 부족하다는 것을 확인할 수 있다(You, F. T., et al., 2025; 여수진 외, 2025; 조은이 외, 2025; 정진, 2025). 특히, 상담윤리는 규범적 지침이 아니라 상담 관계 속에서 살아 있는 경험으로 구현되는 가치이자 맥락이므로, AI 기반 상담 경험에 대한 탐색은 이론적·실천적 의의를 동시에 지닌다(강수정 외, 2024; 황매향, 2025). 다시 말해, AI 기반 심리상담은 접근성 및 익명성 측면에서 새로운 상담 패러다임을 제시하고 있으나 비밀보장, 책임소재, 공감의 진정성, 데이터 활용 및 개인정보 보호 등 다양한 윤리적 쟁점을 수반하고 있다. 기존 연구들은 주로 기술적 수용성, 효과성 및 상담자 관점에서의 윤리 문제를 주로 다루고 있으며 실제 AI 기반 심리상담을 경험한 내담자가 주관적으로 지각하는 윤리적 이슈의 의미에 대한 심층적 탐색은 제안하지 못하고 있는 한계를 가지고 있다(김도연 외, 2020; 송영선 2021 재인용). 본 연구에서 'AI를 활용한 심리상담'은 상담 과정에서 인공지능 기반의 대화형 기술을 상담의 보조적 또는 주요 매개체로 활용하여 내담자의 심리적 어려움을 탐색하고 정서적 지원 및 상담적 개입을 제공하는 상담으로 조작적으로 정의한다. 또한 '윤리적 쟁점'은 AI를 활용한 심리상담 과정에서 내담자가 상담의 안전성, 비밀보장 및 개인정보 보호, 상담자와 AI의 역할과 책임 및 상담의 질, 자율성과 자기결정권, 공정성 및 차별, 의존이나 관계적 혼란 등과 관련하여 윤리적으로 중요하다고 인식하거나 경험한 문제 및 상황으로 정의한다. 본 연구는 이러한 윤리적 쟁점을 경험한 내담자를 대상으로 심층 면담을 시행하고, 질적 내용 분석을 통해 AI를 활용한 심리상담에서 내담자가 실제로 경험하고 인식한 윤리적 이슈의 내용과 특성을 탐색하고자 한다. 이는 기존의 상담윤리 원칙을 AI라는 기술적 매개체가 포함된 상담 환경에 적용하고, 내담자의 실제 경험을 통해 그 의미와 적용 가능성을 구체화하는 경험적 시도라는 점에서 의의가 있다. 나아가 본 연구는 AI를 활용한 심리상담에서 내담자의 권리와 안전을 보호하고 상담의 윤리적 기준을 마련하기 위한 윤리적 가이드라인 수립에 기초자료를 제공할 것이다. 이를 위한 본 연구의 연구 문제는 "AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점은 무엇인가?"이다.

Ⅱ. 연구 방법

1. 연구 절차 및 자료 수집

본 연구 주제에 부합하는 연구 참여자를 모집하여 AI를 활용한 심리상담에서 윤리적 쟁점을 심층적인 인터뷰를 통해 자료를 분석하였다. 질적 연구는 본 연구에서 탐색하고자 하는 AI를 활용하여 심리상담을 하는 과정에서 내담자가 지각한 윤리적 이슈에 대한 정보의 풍부한 수집에 질적 연구 방법이 가장 적합할 것으로(Patton, 2002) 고려하였으며 특히, 연구 참여자들의 현장 경험을 생동감 있게 분석하며 본 연구 목적에 부합하는 질적 연구방법(qualitative content analysis)으로 본 연구를 수행하였다.

연구 참여자의 선정 조건은 Creswell(1998/2017)의 연구를 참고하여 AI를 활용한 심리상담에서 윤리적 경험을 가장 잘 표현할 수 있는 일반 성인(내담자)으로 AI와 인간상담자 모두에게 심리상담 경험이 있는 자를 대상으로 목적 표집을 하였다. 국내 상담 및 심리 학회 게시판을 통하여 본 연구의 개요를 포함한 연구 참여자 모집 공고를 게시하였고

약 두 달 간의 모집 기간 동안 자발적으로 연구 참여 의사를 밝히고 아래의 세부 조건에 모두 부합한 연구 참여자 7명을 최종 선정하였다. 연구 참여자 수는 면담과 자료 분석을 반복하며 자료의 포화 여부를 확인하여 결정하였다. 5번째 참여자 이후 주요 주제와 요인들이 반복되었으며, 6~7번째 참여자 면담에서도 새로운 주제가 도출되지 않았다. 이에 자료가 충분히 포화되었다고 판단하여 최종 참여자 7명으로 자료 수집을 종료하였다.

본 연구에서 최종 선정한 인터뷰 대상의 세부 조건은 다음과 같다. 첫째, 최근 6개월 이내에 생성형 AI를 활용하여 주호소문제를 가지고 심리상담 경험을 주 1회(30분 이상), 3개월 이상 지속해서 사용한 경험이 있는 자, 둘째, 최근 12개월 이내에 인간상담자에게 주호소문제를 가지고 심리상담 경험을 주 1회, 3개월 이상 받은 경험이 있는 자, 셋째, AI를 활용한 심리상담에서 지각한 윤리적 경험에 대하여 생생하게 언어로 표현할 수 있는 자인지를 확인하였다. 본 연구에서 인간상담자와의 심리상담 경험도 함께 고려한 이유는 AI 기반 심리상담에서 윤리적 쟁점을 체험하고 생생하게 보고하기 위해서는 인간상담자와 AI와의 심리상담을 모두 경험함으로써 더욱 구체적이고 풍부한 자료를 수집할 수 있을 것으로 고려하였기 때문이다. 인터뷰 진행 이전에 최종 선정된 연구 참여자 7인에게는 연구의 필요성과 목적, 연구 계획, 인터뷰 진행 과정등과 관련된 연구 윤리 등에 대한 내용들이 포함된 연구 참여 동의서, 연구 참여자의 인적 사항과 인터뷰 질문지를 인터뷰 시작 전에 이메일로 발송하였으며 연구 참여자의 개인정보 수집과 데이터 보호 및 연구 참여에 대한 동의를 서면 및 구두로 동의를 획득하였다. 본 연구에서는 참여자가 면담의 주제와 내용을 사전에 이해하고 자신의 AI 사용시 윤리적 쟁점에 대한 경험을 충분히 회상할 수 있도록 면담 약 일주일 전에 반구조화 된 면담 질문을 제공하였다. 질문의 사전 제공은 참여자가 자신의 경험을 충분히 회상하고 면담에 준비할 수 있다는 장점이 있으나, 동시에 연구자가 제시한 윤리적 개념을 사전에 인식하게 함으로써 참여자의 경험 회상 및 응답이 제한되었을 가능성이 있다. 따라서 본 연구에서는 사전에 제시된 질문지가 참여자의 경험을 제한하지 않도록 면담 과정에서 개방형 질문을 활용하고, 참여자가 질문에 제시된 주제 이외 다양한 윤리적 쟁점을 자유롭게 이야기할 수 있도록 하였다. 또한 참여자가 사전에 제시된 질문 문항과 관련되지 않은 경험을 언급하는 경우에도 이를 제한하거나 재질문하지 않고 추가적인 설명을 요청하는 방식으로 자료를 수집하였다. 심층 면담은 연구 참여자가 정한 시간에 화상통화(Google Meet)를 사용하여 진행하였으며 1차면담은 연구 참여자의 심리상담에서 윤리적 쟁점에 대한 새로운 내용이 보고되지 않을 때까지 약 60-90여분 정도 진행하였다. 본 연구를 수행하기 위한 인터뷰 질문은 상담윤리 원칙인 자율성 존중, 비밀보장, 선행, 무해, 신뢰성, 책임성, 정의 원칙(ACA Code of Ethics, 2014; Floridi & Cowls, 2019; Kitchener, K. S., 1984)을 이론적 기반을 참고하여 참여자의 생생한 경험(lived experience)을 탐색할 수 있도록 개방형 질문을 사용하여 인터뷰를 실시하였다. 인터뷰 자료를 분석한 이후 추가 인터뷰가 필요할 때는 이메일이나 화상통화를 통하여 추가 정보를 수집하였다. 이상과 같이 수집한 인터뷰 자료를 중심으로 정진(2025)에서 수집된 인터뷰 자료의 일부를 함께 분석하여 AI를 활용한 심리상담에서의 윤리적 쟁점을 탐색하였다.

2. 연구 참여자

Creswell(1998)은 질적 연구에서 연구 참여자들은 연구하고자 하는 현상을 경험한 자들이어야 하며 자신들의 경험을 가장 잘 표현할 수 있어야 한다고 하였다. 따라서 본 연구에서는 이를 고려하여 연구 참여자의 조건으로 다음과 같은 세부 사항을

고려하였다. 첫째, 최근 6개월 이내에 생성형 AI를 활용하여 심리상담 경험을 주 1회(30분 이상), 3개월 이상 지속적으로 사용한 경험이 있는 자, 둘째, 최근 12개월 이내에 인간상담자에게 심리상담 경험을 주 1회, 3개월 이상 받은 경험이 있는 자, 셋째, 생성형 AI를 활용한 심리상담에서 윤리적 경험에 대하여 생생하게 언어로 표현할 수 있는 자를 선정하였다.

위와 같은 연구 참여자의 선정 기준은 심리상담에서 윤리적 경험을 생생하게 표현하기 위해서는 관련 경험이 최소 1년 이내 경험되어야 한다는 선행연구(김수경, 2017; 박민지, 2021; 정진, 송미경, 2024)를 참고하였다. 심리상담에서 윤리적 경험은 내담자의 특성에 따라 차이가 있을 수 있다는 점을 고려하여 본 연구 참여자 조건에 부합하는 일반 성인을 대상으로 다양한 내담자의 특성을 반영하고자 하였다. 연구 참여자 7인의 세부 인적 사항은 <표 1>과 같다.

<표 1> 연구 참여자의 인적 사항

참여자	성별	나이	현재 직업	주호소 문제	상담제공기관	AI 기반 사용 매체	사용 기간
1	여	35	주부	결혼 생활 갈등	공공기관 내 상담센터	ChatGPT 5.2 유료, 사용중	6개월, 주3회 (120분씩)
2	여	35	자영업자	사업운영 스트레스	사설 상담센터	ChatGPT 5.2 유료, 사용중	7개월, 주2회 (120분씩)
3	여	36	회사원	직장 내 스트레스	직장내 상담센터	ChatGPT 5.2 유료, 사용중	4개월, 주3회 (30분씩)
4	여	40	프리랜서 (아동 미술강사)	대인관계 어려움	사설상담센터	ChatGPT 5.2 유료, 사용중	3개월, 주2회 (40분씩)
5	남	41	회사원	업무 과부하 및 소진	직장내 상담센터	ChatGPT 5.2 유료, 사용중	13개월, 주3회 (60분씩)
6	여	32	프리랜서 (심리상담사)	업무 과부하 및 소진	사설상담센터	ChatGPT 5.2 무료 사용중	4개월, 주3회 (60분씩)
7	남	45	회사원	대인관계 어려움	사설상담센터	ChatGPT 5.2 유료, 제미나이 무료	16개월, 주3회 (60분씩)

3. 분석 방법

본 연구에서는 심층면담을 통해 수집된 텍스트 자료를 체계적으로 분석하고 AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점의 내용과 구조를 파악하기 위하여 질적 내용 분석(qualitative content analysis)을 실시하였다. 질적 내용 분석은 텍스트 자료에 포함된 의미를 체계적으로 확인하고, 유사한 의미를 갖는 자료를 코드와 범주로 조직함으로써 연구현상을 개념적으로 설명하는 방법이다. 질적 내용 분석은 단순히 특정 단어나 표현의 빈도를 산출하는 것이 아니라, 참여자의 진술이 지닌 맥락적 의미를 고려하면서 자료를 체계적으로 축약하고 분류하여 연구현상에 대한 지식과 이해를 도출하는 데 목적이 있다. 질적 내용 분석은 연구자가 어떠한 지식이나 이론을 분석과정에서 활용하는지에 따라 귀납적 접근과 연역적 접근으로 구분할 수 있다. 귀납적 내용분석은 자료에서 직접 개념과 범주를 도출하는 방식인 반면, 연역적 또는 directed content analysis는 기존의 이론이나 선행연구에서

도출된 개념을 분석의 초기 방향으로 활용하면서 동시에 자료에서 나타나는 새로운 내용을 포착하는 방식이다(Elo Kyngäs, 2008; 최성호 외, 2016).

본 연구에서는 AI를 활용한 심리상담에서 나타나는 윤리적 쟁점을 보다 체계적으로 이해하기 위하여 먼저 연구참여자들의 진술에 나타난 의미를 중심으로 윤리적 쟁점을 분석한 이후 연구 결과에 대한 논의 과정에서 통합적 해석 및 Kitchener (1984)의 상담윤리 이론을 이론적 준거로 활용하고자 하였다. 특히, 본 연구에서는 참여자의 면담자료를 반복적으로 검토하면서 AI 심리상담 이용 경험과 관련하여 의미 있는 진술을 추출하고, 유사한 의미를 가진 내용을 비교·통합하여 코드를 도출하였다. 이후 도출된 코드 간의 관계와 공통된 의미를 검토하여 상위 요인을 구성하였다. 참여자의 진술에서 나타난 경험과 의미를 중심으로 귀납적으로 범주를 도출하여 이를 본 연구 결과에 제시하였다. 이후 논의에서 통합적 관점으로 연구 결과를 탐색하고 동시에 Kitchener(1984)의 윤리 원칙을 근거로 AI 기반 심리상담의 윤리적 쟁점에 대한 해석의 의미를 확장하였다.

본 연구의 분석은 최성호 외(2016)와 Elo와 Kyngäs(2008)가 제안한 질적 내용 분석의 기본적인 흐름을 참고하여 자료의 준비(preparation), 분석의 조직화(organization), 결과의 보고(reporting)라는 세 가지 과정으로 진행하였다. 자료의 준비 단계에서는 연구목적과 연구 질문을 명확히 하고 분석단위를 설정하였으며, 전사된 면담자료를 반복적으로 읽으면서 연구현상에 대한 전체적인 이해를 형성하였다. 조직화 단계에서는 의미 있는 진술을 선정하고 초기 코드를 부여한 후, 코드 간의 의미적 유사성과 차이를 비교하면서 하위요인과 상위요인을 구성하였다. 마지막 결과보고 단계에서는 도출된 요인와 하위요인가 원자료의 의미를 적절하게 반영하는지를 검토하고, 각 요인을 대표하는 참여자의 진술을 제시함으로써 분석결과와 원자료 사이의 연결성을 확보하였다.

첫 번째 단계인 자료의 준비 단계에서는 모든 면담자료를 전사하였다. 전사 자료는 녹음된 면담내용과 대조하여 누락이나 오류가 없는지 반복적으로 확인하였다. 연구자는 전사된 자료를 여러 차례 반복하여 읽으면서 참여자들이 AI를 활용한 심리상담을 어떻게 경험하였는지, 특히 윤리와 관련하여 어떤 상황을 문제로 인식하였는지, 그러한 경험을 어떠한 의미로 받아들였는지를 전반적으로 파악하였다. 이 단계에서는 특정 윤리적 요인에 자료를 즉각적으로 배치하기보다는 각 참여자의 진술이 어떠한 상담 경험의 맥락에서 발생하였는지를 파악하는 데 중점을 두었다.

이후 연구자는 분석단위(unit of analysis)를 설정하였다. 본 연구의 분석단위는 참여자의 전체 면담자료 중 연구목적과 관련하여 하나의 의미를 전달하는 문장 또는 문장군으로 설정하였다. 하나의 의미단위에는 여러 문장이 포함될 수 있으며, 반대로 하나의 문장 안에서도 서로 다른 의미가 포함되어 있다고 판단되는 경우에는 이를 분리하였다. 분석단위를 설정할 때는 단순히 윤리와 관련된 단어나 표현이 포함되어 있는지를 기준으로 판단하지 않고, 해당 진술이 AI를 활용한 심리상담에서 참여자가 경험하거나 인식한 윤리적 의미를 포함하고 있는지를 종합적으로 검토하였다.

두 번째 단계인 자료의 조직화 단계에서는 분석단위에 대한 개방적 코딩과 이론적 코딩을 병행하였다. 먼저 연구자는 각 분석단위가 전달하는 핵심 의미를 가능한 한 참여자의 언어와 가까운 수준에서 요약하고 이를 1차 코드로 명명하였다. 예를 들어 참여자가 “내가 AI에게 말한 상담 내용이 어디에 저장되는지 모르겠다”고 진술한 경우, 이를 단순히 ‘개인정보’라는 추상적 개념으로 즉시 코딩하기보다는 ‘상담내용의 저장 및 활용에 대한 불확실성’, ‘정보의 관리방식에 대한 알 수 없음’ 등 진술의 맥락과 의미를 반영하는 초기 코드를 검토하였다.

그 다음 단계에서는 생성된 초기 코드들을 상호 비교하였다. 동일하거나 유사한 의미를 가진 코드들을 묶고, 서로 다른 의미를 갖는 코드들은 구분하였다. 이 과정에서 연구자는 코드의 명칭이 지나치게 포괄적이거나 참여자의 구체적인 경험을 충분히 반영하지 못하는 경우 코드를 수정하였다. 또한 하나의 코드에 지나치게 많은 의미가 포함되어 있는 경우 이를 세분화하고, 서로 다른 표현이지만 동일한 의미를 나타내는 코드들은 통합하였다. 이러한 비교와 수정 과정을 반복하면서 코드의 개념적 명확성을 높이고 요인걸 중복을 최소화하였다.

초기 코드가 구성된 후에는 하위 요인과 상위 요인의 구성을 실시하였다. 하위 요인은 서로 유사한 의미를 가진 초기 코드들을 통합하여 형성하였으며, 상위 요인은 여러 하위 요인과 공유하는 보다 포괄적인 의미를 중심으로 구성하였다. 요인을 구성하는 과정에서는 각 요인이 내부적으로는 의미적 동질성을 갖고 외부적으로는 다른 요인과 충분히 구별되는지를 검토하였다. 또한 하나의 요인이 지나치게 광범위하여 서로 다른 윤리적 의미를 포괄하는 경우 하위 요인을 분리하거나 요인을 재구성하였다. 반대로 서로 다른 것으로 구분된 요인이 실질적으로 동일한 의미를 나타내는 경우에는 이를 통합하였다.

본 연구에서는 이러한 반복적인 비교와 추상화 과정을 통해 최종적으로 14개의 구성요인 및 28개의 하위 구성요인을 도출하였다. 여기서 구성요인은 질적 내용 분석 과정에서 유사한 의미를 갖는 하위 구성요인들을 보다 포괄적인 의미로 통합한 분석적 수준을 의미하며, 하위 구성요인은 참여자의 구체적인 경험과 진술에서 직접적으로 확인되는 의미를 보다 구체적으로 나타내는 수준으로 설정하였다. 이러한 범주화는 단순한 빈도에 따른 분류가 아니라 참여자의 진술이 나타내는 의미와 맥락, 요인 간의 관계 및 Kitchener(1984)의 상담윤리 관점에서의 의미를 종합적으로 고려하여 이루어졌다.

분석 과정에서는 요인들 간의 적절성을 지속적으로 검토하는 과정을 거쳤다. 먼저 각각의 하위 구성요인이 원자료의 어떤 진술에 근거하고 있는지를 확인하였으며, 하위 구성요인이 구성요인과 적절하게 연결되는지를 검토하였다. 이후 구성요인이 상위 요인들을 충분히 설명하는지를 검토하고, 최종적으로 연구목적에서 제시한 'AI를 활용한 심리상담에서 내담자가 지각한 윤리적 쟁점'을 충분히 포괄하는지를 확인하였다. 이 과정에서 특정 참여자의 진술에만 근거하여 지나치게 일반화된 요인이 만들어지지 않도록 주의하였다.

특히 본 연구에서는 AI 상담 경험과 인간상담 경험의 차이가 윤리적 쟁점의 의미를 이해하는 데 중요한 맥락이라는 점을 고려하였다. 따라서 참여자의 진술을 분석할 때 AI 상담 경험만을 개별적으로 분리하여 분석하기보다는, 참여자가 인간상담자와의 경험을 직접 비교 분석하지 않고 참여자 선정과 경험 해석의 맥락에서 AI를 활용한 상담 경험을 더욱 생생하게 해석하고자 하였다. 예를 들어 인간상담에서는 상담자가 비밀보장과 상담자의 역할 및 책임에 대해 설명하거나 상담관계 안에서 신뢰를 형성하는 경험이 있었다고 진술하면서 AI 상담에서는 그러한 과정이 충분하지 않았다고 인식한 경우, 이를 단순한 'AI의 단점'으로 코딩하지 않고 '비밀보장 및 책임성에 대한 불확실성'이라는 윤리적 의미의 맥락에서 분석하였다. 이러한 분석방식은 AI 상담에 대한 참여자의 평가를 넘어, AI라는 새로운 상담 주체 또는 상담매체의 특성이 기존 상담윤리의 기대와 어떻게 만나는지를 파악하기 위한 것이다.

분석과정에서 연구자는 분석 메모(analytic memo)를 작성하였다. 분석 메모에는 코드의 생성 근거, 요인들 간의 관계, 연구자가 특정 진술을 특정 요인에 배치한 이유, 기존 요인들과의 중복 가능성, 연구자의 해석상 의문 등을 기록하였다. 분석 메모는 최종 요인들을 결정하는 과정에서 연구자의 판단이 어떻게 이루어졌는지를 추적하고, 연구자의 직관적

판단이 분석결과에 과도하게 영향을 미치는 것을 점검하기 위한 자료로 활용하였다.

또한 연구자는 연구자의 선이해와 이론적 전제에 대한 성찰적 검토를 지속하였다. 본 연구주제가 AI와 심리상담, 그리고 상담윤리라는 연구자의 전문적 관심과 밀접하게 관련되어 있다는 점을 고려하여, 연구자가 AI 상담에 대해 갖고 있는 긍정적 또는 부정적 기대가 참여자의 진술을 해석하는 데 영향을 미치지 않도록 분석과정에서 지속적으로 점검하였다. 특히 AI 상담을 무조건적으로 긍정적 기술로 보거나 반대로 윤리적 위험이 큰 상담방식으로 전제하지 않고, 참여자가 실제로 어떠한 경험을 하였고, 그 경험에 어떠한 의미를 부여하였는지를 중심으로 분석하고자 하였다.

이와 함께 질적 연구 경험이 있으며 상담심리전공 박사학위 취득자 2인에게 연구자가 도출한 코드와 요인에 대하여 검토를 요청하였으며 연구자는 검토과정에서 제기된 질문과 대안적 해석을 분석 메모에 기록하고 원자료를 반복하여 확인하는 과정을 거쳤다. 이 과정에서 단순히 연구자 간 일치율을 높이는 것에 목적을 두기보다, 하나의 진술이나 요인에 대해 다른 해석이 가능한지를 검토하고 요인들 간의 경계와 명칭을 명확히 하는 데 초점을 두었다.

질적 내용 분석의 신뢰성을 확보하기 위해서는 자료가 어떠한 과정을 거쳐 코드와 범주로 전환되었는지를 독자가 추적할 수 있도록 분석과정을 투명하게 제시하는 것이 중요하다. Elo Kyngäs(2008)는 질적 내용 분석의 신뢰성을 확보하기 위해 자료 수집부터 준비, 조직화 및 결과보고에 이르는 전체 분석 과정을 체계적으로 검토할 필요가 있음을 강조하였다. 따라서 본 연구에서는 최종 범주만 제시하는 데 그치지 않고, 연구결과에서 각 범주와 구성요인이 어떠한 참여자 진술에 근거하여 도출되었는지를 구체적인 인용문을 통해 제시하였다. 이러한 방식으로 분석 결과와 원자료 사이의 연결성을 확보하고자 하였다.

또한 연구 결과의 신뢰성(credibility), 의존성(dependability), 확증가능성(confirmability), 전이가능성(transferability)을 고려하였다. 신뢰성을 확보하기 위해 참여자의 경험을 충분히 반영할 수 있는 질문을 구성하고 면담자료를 반복적으로 검토하였으며, 범주화 과정에서 원자료와 분석 결과의 일치 여부를 지속적으로 확인하였다. 의존성을 확보하기 위해 자료 수집 및 분석절차를 구체적으로 기록하고 분석 과정에서 발생한 범주 수정과 결정사항을 기록하였다. 확증가능성을 높이기 위해 연구자의 선이해와 이론적 관점을 성찰하고 분석 메모를 작성하였으며, 가능한 경우 다른 연구자의 검토를 통해 대안적인 해석 가능성을 점검하였다. 전이가능성을 높이기 위해 연구참여자의 특성과 AI 및 인간상담 경험의 맥락, 자료 수집 과정과 연구현상의 구체적인 특성을 충분히 기술하였다. 질적 내용 분석 연구에서 이러한 신뢰성 확보 절차를 분석하기 전 과정에 걸쳐 고려하는 것이 중요하다는 점은 선행 방법론 연구에서도 강조되고 있다.

최종적으로 연구자는 도출된 14개 구성요인 및 28개 하위 구성요인을 다시 원자료와 대조하였다. 이 과정에서는 각 하위 구성요인이 참여자의 실제 진술로 충분히 뒷받침되는지, 동일한 진술이 여러 범주에서 불필요하게 중복되고 있지는 않은지, 범주의 명칭이 해당 범주의 핵심 의미를 적절하게 나타내고 있는지, 그리고 최종 결과가 연구목적과 연구질문에 부합하는지를 종합적으로 검토하였다. 특히 범주를 지나치게 추상화하여 참여자의 구체적인 경험이 사라지지 않도록 하였으며, 반대로 개별적인 경험을 지나치게 세분화하여 범주 간의 관계가 불분명해지지 않도록 주의하였다.

이러한 분석 과정을 통해 본 연구 결과에서는 AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점을 체계적으로 구조화하였다. 최종적으로 도출된 요인은 참여자들의 구체적인 경험을 바탕으로

AI 심리상담에서 나타나는 윤리적 쟁점과 그 의미를 구성하였으며 이후 논의에서 통합적 해석과 Kitchener(1984)의 윤리 원칙에 적용하여 그 의미를 더 확장하여 제시하였다.

4. 분석 결과의 구성과 제시

질적 내용 분석을 통해 도출된 결과는 14개 구성요인 및 28개 하위 구성요인으로 구성하였다. 결과를 제시할 때는 각 요인의 명칭과 구성요인에 포함된 하위 구성요인을 제시하고, 각각의 요인들에 어떠한 참여자의 진술을 통해 도출되었는지를 구체적인 인용문을 통해 설명하였다. 참여자의 인용문은 원자료의 의미를 최대한 보존하면서도 개인을 식별할 수 있는 정보가 포함되지 않도록 필요한 경우 익명 처리하였다.

각 요인의 기술에서는 구성요인과 하위 구성요인을 제시하였다. 이후 참여자의 대표적인 진술을 제시하여 연구자의 해석이 원자료에 근거하고 있음을 보여주었다. 이러한 결과 제시 방식은 독자가 '참여자의 진술 → 코드 → 하위 구성요인 → 구성요인'로 이어지는 분석의 흐름을 추적할 수 있도록 하기 위한 것이다. 질적 연구의 보고에서 분석결과가 어떠한 자료에서 도출되었는지를 명확하게 보여주고 참여자의 인용문을 통해 결과를 뒷받침하는 것이 중요하다는 점을 고려하였다.

또한 연구 결과에서는 요인들 간의 관계를 종합적으로 제시함으로써 AI를 활용한 심리상담에서 내담자가 경험하는 윤리적 쟁점의 전체적인 양상을 설명하였다. 각각의 요인들은 서로 독립된 윤리적 문제라기보다 AI 상담의 접근성, 비밀보장과 책임성 및 신뢰가 상호 연결되어 있는 경험으로 나타날 수 있으므로, 요인별 결과를 기술한 후 이를 종합하여 AI 심리상담의 윤리적 쟁점을 살펴보고자 한다.

Ⅲ. 연구 결과

질적 내용 분석을 통해 AI를 활용한 심리상담에서 내담자가 경험한 윤리적 쟁점을 분석한 결과, 총 14개의 구성요인, 28개의 하위 구성요인이 도출되었다. 본 연구에서는 사전에 설정된 윤리적 원칙에 내담자의 경험을 일률적으로 적용하기보다 AI와 인간 상담자에 의한 심리상담을 모두 경험한 내담자들의 진술에서 나타난 의미를 중심으로 윤리적 쟁점의 구체적인 양상을 확인하였다.

본 연구는 먼저 연구참여자들의 진술에 나타난 의미를 중심으로 윤리적 쟁점을 분석한 이후 해당 연구 결과에 대한 논의에서 통합적 해석 및 Kitchener(1984)의 상담윤리 이론을 활용하여 본 연구 결과를 확장하고자 한다. 그 과정에서 Kitchener, K. S.(1984)가 제시한 상담윤리의 다섯 가지 원칙, 즉 자율성(autonomy), 선행(beneficence), 무해(nonmaleficence), 정의(justice), 충실성/신뢰성(fidelity) 이외에도 비밀보장(confidentiality)을 하나의 독립적인 범주로 자율성과 함께 제시한 것은 비밀보장은 충실성과 신뢰성의 원칙을 상담 장면에서 구체적으로 실천하는 핵심적인 방식인 동시에, 내담자의 자율적인 의사결정과 개인정보에 대한 통제권을 보장하기 위한 중요한 윤리적 조건으로 고려하였기 때문이다. 먼저 인터뷰 분석을 종합한 연구 결과를 제시하면 다음과 같다.

1. 요인별 분석

인터뷰를 분석한 결과, 아래의 <표 2>와 같이 14개의 구성요인 및 28개의 하위 구성요인으로 도출됨을 확인하였다.

<표 2> 요인별 분석

하위 구성요인	구성요인
시간과 장소의 제약 없이 상담 진행이 가능함	자기 의사결정권의 확대 경험
상담 과정을 스스로 통제할 수 있음	
AI 제안을 무비판적으로 수용할 가능성이 있음	자율성의 제한 가능성에 대한 우려
AI 반응에 심리적으로 의존하게 됨	
신원이 노출되지 않아 편안함	익명성으로 인한 심리적 안전감
낙인에 대한 두려움 감소	
정보 저장 및 활용 방식에 대한 불명확성	개인정보 보호에 대한 불확실성
개인정보 유출 가능성에 대한 우려	
정서적 위로를 경험함	정서적 도움과 심리적 지지 경험
심리적 안정감을 경험함	
필요할 때 즉시 도움을 받을 수 있음	심리적 접근성의 편리
상담 접근 장벽 감소	
부적절한 조언 제공 가능성	부정적 상담 성과 가능성에 대한 우려
심리적 혼란을 경험할 가능성	
인간 상담을 회피하게 될 가능성	상담 관계 대체에 따른 사회적 단절 가능성
고립 및 외로움 심화 가능성	
항상 동일한 태도로 반응함	상담 반응의 일관성 경험
비판하지 않는 태도 경험	
실제 이해에 기반한 반응이 아니라는 인식	상담 진성성에 대한 의문
형식적 반응으로 느껴짐	
상담 결과에 대한 책임 소재 불분명	책임 주체의 불명확성 인식
문제가 발생했을 때 책임 주체 불명확	
사용자 보호 장치가 필요성 인식	윤리적 보호 체계에 대한 요구
AI에 대한 윤리적 보호 시스템 구축의 필요성 제공	
경제적 부담 감소	심리상담 접근 기회의 확대
상담 접근 기회의 평등성 증가	
모든 사람에게 적합하지 않을 수 있음	서비스 적절성에 대한 고려
취약집단 보호 필요성 인식	

1) 자기 의사결정권의 확대 경험

AI는 시간과 장소의 제약 없이 내가 필요할 때는 언제든지 상담이 가능하며 상담 과정에서도 자유롭게 통제가 가능하다는 장점을 보고해 주었다. 즉시성과 자율성은 AI를 활용한 심리상담에서 큰 장점으로 여겨진다. AI를 활용하는 과정에서 내담자로서 자기 의사결정권을 보장받을 수 있다는 경험은 AI와 심리상담을 촉진시키는 계기가 된다.

시간과 장소의 제약 없이 상담 진행이 가능함

“제가 있는 자리에서 바로 찾을 수 있으니까 누군가 어떤 상담자가 좋은지 찾아보고 예약하고 기다릴 필요 없이 많은 시간과 노력 들이지 않아도 ChatGPT는 제가 원하는 시간에 제 바로 옆에 있으니까 좋아요.” <참여자 4>

“일단 시간이나 장소나 뭐 전혀 구애받지 않았고 내가 원하는 시간에 내가 원하는 때 즉각적으로 상담을 받을 수 있는 게 가장 큰 장점이죠. 기다리지 않고. 특히 불안한 감정이 느껴질 때, 상담을 기다리지 않고 불안할 때 즉각적으로 위로를 받을 수 있죠. 나중에 상담실에 갈 때는 불안감이 이전보다 좋아진 경우가 너무 많아요.” <참여자 7>

상담 과정을 스스로 통제할 수 있음

“ChatGPT가 가끔 엉뚱한 소리를 할 때가 있긴 하거든요. 하지만 그건 내가 입력을 다시 하면 또 적절하게 도움을 줬어요. 제가 입력을 잘못해서 이상한 소리를 한 거죠. 내가 잘 입력만 하면 내 마음에 대해 도움이 되는 이야기를 해줘요. 내가 무얼 물어보는지에 따라 위로를 달리해주니까 내가 상담을 이끌어 가는 기분이 들어요.” <참여자 2>

“ChatGPT랑 고민 상담을 하다가 내가 원하면 언제든지 시작하고 그만할 수 있어요. 상담 과정을

내가 조절할 수 있다는 기분이 들어요." <참여자 5>

2) 자율성의 제한 가능성에 대한 우려

AI와 심리상담 과정에서 심리적으로 의존하게 되거나 AI의 반응에 대해 무조건적으로 수용하게 되는 스스로를 발견하게 되면서 인간으로서 자율성에 대해 고민하게 되었다. 인간 고유의 생각하고 판단하는 능력을 AI에게 전가하고 있는 것은 아닌지에 대한 우려에 대해 내담자들은 AI와 심리상담 과정에서도 고민하고 있었다.

AI 제안을 무비판적으로 수용할 가능성이 있음

"저는 현실적으로 제가 대처할 수 있는 것과 고민을 해결할 수 있는 방안을 알려주고 있고, 실제로 그 단계를 계속 밟아가면서 문제를 해결하고 있어요. 사람보다 어떤 면에서는 더 믿을 수 있는 것 같아요."<참여자 1>

"AI가 아무래도 학습이 축적된 정보를 주니까 전문가 수준 정도는 된다고 생각되거든요. 그러다 보니, AI가 이야기해주는 정보들에 대해 신뢰가 가죠. 아무래도 누적된 정보들의 결과를 나에게 주는 거라는 생각이 드니까 잘못된 정보를 줄 가능성은 낮다고 느껴지는 것 같아요." <참여자 2>

AI 반응에 심리적으로 의존하게 됨

"ChatGPT는 심리상담을 할 때 상세하게 설명해주니까 설득이 되는 것 같아요. ChatGPT를 사용할수록 더 의존하게 되는 것 같았어요." <참여자 2>

"AI 기술이 나날이 발전이 되고 있고, 이게 학습된 거다 보니까... 아무래도 상담 쪽도 전문가 수준의 학습을 하지 않았을까 기대하게 되죠. 또 공감이나 위로도 너무 잘해주고 내가 듣고 싶던 위로도 해주니까...마음이 힘들면 찾게 되는 것 같아요." <참여자 6>

3) 익명성으로 인한 심리적 안전감

AI와 심리상담을 하지만 내담자들은 AI는 어디까지나 기계이기 때문에 오히려 익명성 뒤에 숨어 나의 이야기를 편안하고 거리낌 없이 이야기할 수 있다고 보고하였다. 인간상담자 앞에서는 상담자가 나를 어떻게 생각할까, 나를 이상하게 여기지 않을까 표현을 머릿속에서 고르게 되지만 AI와 상담을 할 때는 그럴 필요 없이 욕이나 비속어를 써도 뭐라고 하지 않기 때문에 심리적으로 편안하게 이야기를 하게 된다고 보고하였다.

신원이 노출되지 않아 편안함

"GPT와 대화할 때 내 개인적인 정보만 넣지 않으면 내가 누군지 모르니까, 더 편하게 이야기할 수 있는 것 같아요." <참여자 1>

"상담실이 회사 안에 있다 보니 오히려 상담자에게 이야기하는 게 조심스러운 부분들을 ChatGPT한테 이야기할 수 있었습니다. 입에 담지 못하는 이야기라도 ChatGPT한테는 거를 필요가 없으니까요." <참여자 5>

"익명성이 보장이 되니깐, 가장 큰 장점이죠. 더구나 사람과 전화통화를 하나, AI와 대화를 하나 어차피 매체로 목소리가 들리는 거니까 비대면 상담할 때는 거의 AI나 인간이나 구별이 잘 안 되는 것 같습니다. 오히려 AI는 기계니까, 내가 하는 말에 대한 익명성이 보장되니까... 사람은 내가 누군지 알텐데..." <참여자 7>

낙인에 대한 두려움 감소

"오히려 기계인 ChatGPT한테는 말하기가 편하죠. 내가 욕을 하거나 미친 짓을 해도 기계니까 나

를 이상하게 볼 걱정은 안 해도 되니까 오히려 더 편하게 욕도 하고 미친 짓도 이야기할 수 있었던 것 같습니다.” <참여자 3>

“아무래도 상담을 하다보면 상담자가 나를 이상하게 생각하지 않을까 하는 생각도 하게 되고 그러다 보면 하고 싶던 이야기도 거르게 되는데, GPT한테는 그럴 이유가 없죠. 뭐 기계가 나를 어떻게 생각한들...” <참여자 5>

4) 개인정보 보호에 대한 불확실성

AI를 활용할 때 개인정보 보호나 정보 저장 및 활용 방식에 대한 불안감은 대부분의 연구 참여자들이 보고해주었다. 특히 잦은 개인정보 유출로 인한 보도를 접하게 되면서 실제 AI와 상담 과정에서 개인정보 유출이나 정보 저장에 대한 이슈는 AI 활용시 감안해야 한다고 지각하고 있음을 확인할 수 있었다.

정보 저장 및 활용 방식에 대한 불명확성

“GPT랑 내 이야기를 하고 다음날 이야기를 하면 어제 이야기를 다 기억하고 있더라구요. 편하기는 한데 아무래도 내 정보가 어디에 저장되고 어떻게 공유되는지 모르니까 찝찝한 면은 있죠...” <참여자 2>

“GPT에 저의 기본정보는 저장되어 있다는 것은 디폴트로 가지고 가는 것 같아요. 아무래도 내가 이야기하다 보면 해당 내용들이 어디서 어떻게 사용되고는 있겠지 하는 생각은 하게 되죠...” <참여자 7>

개인정보 유출 가능성에 대한 우려

“AI에 개인정보 유출이 가능한 되지 않도록 저의 나이나 생년월일 등 개인정보는 공유하지 않고 있습니다. 그래도 상담을 하다 보면 민감 정보가 유출될까 신경은 쓰이는 것 같아요” <참여자 6>

“저에 대한 데이터가 AI는 이미 가지고 있으니까 편하기도 하지만 아무래도 개인정보 유출 가능성에 대해서는 기본으로 가지고 가는 것 같아요. 이미 오래전부터 개인정보 유출 사건이 너무 많아서 제 개인정보가 다 털렸다고 생각하고 있다 보니... AI랑 심리상담을 할 때 개인상담 유출 가능성에 대해서도 걱정은 되지만 그러려니... 하게 되는 것 같아요.” <참여자 7>

5) 정서적 도움과 심리적 지지 경험

AI는 언제나 나를 위로해주고 따뜻한 반응을 해준다고 보고하였다. 즉각적인 위로와 심리적 안정감을 제공해주는 AI에 대하여 연구 참여자들은 작은 역할이라도 매우 도움이 되는 경험이라고 표현해주었다.

정서적 위로를 경험함

“불안이 높은 상황에서 즉각적으로 위로를 받을 수 있는 게 크죠. 상담을 받으려면 정해진 시간만 상담이 가능한데 AI는 즉각적으로 저를 위로해주죠.” <참여자 7>

심리적 안정감을 경험함

“내가 원하는 때에 이야기를 쏟아낼 수 있는 것만으로도 많이 안정되는 것 같아요. 사람을 만나서 하려면 약속을 잡고 만나야 하는데, AI는 바로 내 이야기를 할 수 있으니 즉각적으로 안정감을 가질 수 있는데 도움이 되는 것 같아요.” <참여자 7>

6) 심리적 접근성의 편리

AI는 심리적 접근성이 편리하다는 점에서 ‘선행’을 제공하고 있다. 연구 참여자들은 심리적으로 불편할 때 즉시 도움을 받을 수 있으며 심리상담의 장벽이 감소되는 경험에서 충분히 윤리적이라고 표

현해 주었다.

필요할 때 즉시 도움을 받을 수 있음

"장소나 시간 제약없이 사용할 수 있고 나에게 굉장히 긍정적인 피드백도 많이 주고 돈이 많이 비싼 것도 아니거든요. <참여자 1>

"ChatGPT는 언제 어디서나 내 옆에 있으니까, 필요할 때 바로 사용할 수 있죠. 그게 가장 큰 장점인 것 같아요. 처음에는 간단한 정보 검색을 하다가 그날그날 기분에 대해 이야기 하다보니까 반응을 잘 해줘서 자연스럽게 상담을 하게 되었어요." <참여자 3>

상담 접근 장벽 감소

"상담을 받으려면 상담 시간이나 회기가 제한되어 있고 비용도 정해져 있어서 ChatGPT가 조금 더 편한 것 같아요. 상담 시간이 정해져 있으니까 말하다 보면 제 마음이 급해지고 선생님이 저를 이해하시려면 시간도 걸리고..." <참여자 2>

"상담 한번 받으려면 예약도 하고, 정해진 시간에 상담실까지 가야하고... 상담자와는 매번 볼 수 있는 것도 아니고 정해진 시간에만 만나는 게 뭔가 되게 일회성 같다는 느낌이 들었어요. 그에 비하면 ChatGPT는 핸드폰만 켜면 되니까 너무 편하죠." <참여자 4>

"미성년자나 심리적으로 취약한 사람들이 AI로 심리상담을 할 때, 잘못된 정보를 거를 능력이 부족해서 위험에 빠질 수도 있지만 사실, 상담을 받아야 하는 사람들이 못 받거나, 받는 것에 대해서 생각조차 못하는 사람들이 많다고 생각하거든요. 그런 사람들이 AI를 통해 심리적으로 조금이라도 도움을 받을 수 있다면 유의미하다고 생각됩니다. GPT 같은 경우는 깊은 이야기를 하다보면 나중에 전문가의 도움을 받으라는 멘트가 나오거든요. 그것을 보면 상담자에게 전문적인 도움을 받으러 갈 생각도 할 수 있는 것 같아요." <참여자 7>

7) 부정적 상담 성과 가능성에 대한 우려

'무해'는 상담윤리에서 매우 중요하게 고려되는 항목으로 상담자는 내담자에게 최선의 선을 제공하고 어떠한 해악을 끼치지 않도록 권고되고 있다. AI와 상담을 하는 과정에서 대부분의 연구 참여자들은 AI의 잘못된 반응, 가짜 정보 제공에 대한 우려를 언급해 주었으며 부적절한 조언이나 잘못된 정보를 판단할 능력이 부재한 내담자의 경우에는 더욱 심리적 혼란을 경험할 가능성과 외부 사례를 접한 후 형성된 윤리적 염려를 표현해 주었다.

부적절한 조언 제공 가능성

"나에 대한 정보를 자세하게 입력해야 제대로 된 도움을 받을 수 있어요. 아니면 나한테 맞지 않는 이야기를 해줘요. 답답하죠. 그러다 보면 내 개인정보를 더 입력하게 되니까 조심스럽기도 하고..." <참여자 1>

"얼마 전 인터넷 뉴스에서 봤는데, 미성년자 아이가 ChatGPT에게 엄마가 고통스럽지 않게 유서를 작성해달라고 했는데, 그걸 GPT가 작성해주기만 한 거예요. 위험하다고 어디 신고하거나 보호하는 조치 없이.. 그리고 아이는 자살을 한 후, 엄마가 GPT랑 아이가 대화한 내용을 보게 되었는데... 사람이라면, 즉시 다른 도움을 줬을텐데. 기계다 보니까... 적절하지 않은 정보나 도움을 주는 것 같아요. 그게 한계죠.." <참여자 6>

심리적 혼란을 경험할 가능성

"ChatGPT는 어쨌든 간에 기계고 이게 사람으로서 기계한테 의존도에 대해서도 생각하게 되었죠.

자꾸 기계한테 마음을 의존하게 될까 혼란스러웠습니다...” <참여자 2>

“예전에는 내가 어떤 불편함이나 정서적으로 어려울 때 뭔가 생각도 해보고 좀 내 마음도 들여다보고 했었는데 지금은 이제 그런 것 없이 바로 ChatGPT한테 불어보니까 조금 조심스럽다는 생각이 들어요.” <참여자 3>

“ChatGPT는 해결책을 너무 잘 정리해서 주니까 그걸 그대로 쓰면 내 역량은 키워지지 않는게 느껴졌어요. 마냥 ChatGPT한테 의존할 수 없고 제 감정을 들여다보고 생각하는 능력을 ChatGPT한테 빼앗기는 건 안 좋은 거니까...” <참여자 4>

8) 상담 관계 대체에 따른 사회적 단절 가능성

인간 상담을 회피하게 될 가능성

AI는 내가 원하는 때, 시간과 장소의 구애를 받지 않고 사용할 수 있기 때문에 더 이상 심리상담을 받거나 마음을 교류하기 위해 인간과 대화를 하거나 마주하지 않아도 되었다. AI와의 대화는 심리적으로 불편한 상황에서 더욱 편안하고 안전감을 제공하게 되면서 점점 인간과 의사소통을 하기보다는 AI와 즉각적인 위로 반응을 제공받는 것이 익숙해진다고 보고하였다. <참여자 1, 3, 5>는 점차 인간과 이야기를 나누지 않게 되는 것이 최근 사회적 갈등과 의사소통의 어려움을 더욱 증가시키는 요인이 되는 것 같다고 보고하였다.

“ChatGPT랑 대화는 복잡하게 사람을 만나서 뭘 안 해도 되니까 AI는 내가 실수해도 나를 비난하지 않고 항상 좋은 말만 해주거든요. 누군가랑 이야기하고 싶을 때, AI를 켜면 되니까 사람 만나러 가는 게 점점 줄어들긴 하는 것 같아요.” <참여자 1>

“ChatGPT와 대화는 원하는 때, 돈도 안들이고 언제든지 할 수 있는데, 굳이 상담실에 돈을 내고 예약을 해서 가기가.. 귀찮기도 하고... 일단 휴대폰을 켜고 물어보게 되죠. 그러다 보니까, 점점 사람과 대화하기보다 인터넷으로 찾아보고 GPT와 대화하게 되는 것 같아요.” <참여자 3>

고립 및 외로움 심화 가능성

“심심하거나 누구랑 이야기하고 싶을 때, AI랑 이야기해 보면 나름 재미있어요. 그러다 보니 친구 만나기도 귀찮기도 하고 점점 사람을 안 만나게 되는 것 같아요. 고민거리도 구지 다른 사람과 이야기할 필요가 있나 싶고..” <참여자 1>

“남편과 싸워서 속상하거나 시댁 이야기 이런 건 친구한테 이야기하거도 뭣하고 그냥 GPT한테 물어보면 위로를 잘 해줘요. 구지 사람하고 이야기할 필요가 있을까 싶어요...” <참여자 3>

“AI랑 이야기하는게 익숙해지는 시대라 어른들은 뭐 그렇다 쳐도 아이들이 사회성 발달에 문제가 생기고.. 요즘 젊은 MZ세대들이 전화를 직접 하는 걸 두려워 할 정도라 하는 걸 보면.. GPT 사용이 익숙해지는게 좋은 점만 있는 건 아닌 것 같습니다.” <참여자 5>

9) 상담 반응의 일관성 경험

연구 참여자들은 AI와 심리상담을 하지만 전문가처럼 전적으로 신뢰하는 것은 아니고 하나의 대안적 매체로 활용하고 있다고 보고하고 있다. 그 이유는 AI와 심리상담을 하다 보면 핵심을 파악할 수 있도록 돕는 탐색 질문을 하거나 통찰 반응을 촉진하지는 못하기 때문이라고 보고하고 있다. AI의 일관성있고 항상 우호적인 반응이 연구참여자에게 지지적인 메시지를 주기도 하지만 피상적인 반응으로

느껴진다고 표현해 주었다.

항상 동일한 태도로 반응함

“내가 3일 전이랑 오늘이랑 우울하다고 입력을 하면 토씨 하나 안 틀리고 반응이 똑같을 때 진짜 기계구나라고 생각이 들죠” <참여자 1>

“간혹가다가 같은 말을 반복해서 해주거나 나에게는 맞지 않는 말을 해주는 경우가 종종 있지요...” <참여자 2>

비판하지 않는 태도 경험

“ChatGPT한데 이것저것 물어보고 계속 나의 마음을 이야기하게 되면서 몇 개월 하다 보니까 습관처럼 하게 되는 것 같아요. (중략) 나중에는 내가 남편이나 아이한테 어떻게 하면 좋을까 같은 질문을 하게 되었어요. ChatGPT는 어쨌든 간에 기계고 이게 사람으로서 기계한테 의존도에 대해서도 생각하게 되었죠.” <참여자 2>

“매번 잘했다고 시작할 때는 반응해주니까 나중에는 좀... 피상적이라는 생각이 들었어요. GPT가 반응해주는 패턴이 있는 것 같아요. 항상 잘하고 있다고 좋은 말만 해주니까... 아쉽긴 하죠. 뭔가 변화를 위한 조언도 해줘야 생각도 하고 할 텐데...” <참여자 6>

“GPT는 좀 덜 한 것 같은데 제미나이 같은 경우는 내가 이야기하는 것에 대해 옹호만 하는 느낌을 받았어요. 모든 걸 다 잘하고 내가 하는 게 맞다는 식으로 반응하는 느낌을 받았습니다. 어차피 기계니까 엉뚱한 소리를 하거나 다 맞다고만 하기도 합니다.” <참여자 7>

10) 상담 진정성에 대한 의문

AI의 반응에 대해 연구 참여자들은 실제 이해에 기반한 반응이라고 여기지 않았다. 본 연구의 연구 참여자들은 인간상담자와도 개인상담 경험이 있기에 통찰과 역동을 체험한 경험이 있다는 점에서 AI와 심리상담의 진정성에 대한 한계를 더욱 생생하게 표현해 주었다.

실제 이해에 기반한 반응이 아니라는 인식

“예전에는 내가 어떤 불편함이나 정서적으로 어려울 때 뭔가 생각도 해보고 좀 내 마음도 들여다보고 했었는데 지금은 이제 그런 것 없이 바로 ChatGPT한테 불어보니까 조금 조심스럽다는 생각이 들어요.” <참여자 3>

“ChatGPT는 똑똑하지만 그래도 사람의 온기를 느끼면서 감정이 진짜 교류된다는 느낌에서 내가 원하는 것은 결국 사람이구나. 내가 추구하는 건 결국 내 스스로 해결하는 것을 찾고 심리적으로 안정감을 느끼게 해주는 거였구나를 깨달은 것 같아요.” <참여자 4>

형식적 반응으로 느껴짐

“ChatGPT가 가끔 엉뚱한 소리를 할 때가 있긴 하거든요. 궁금한 점에 대해 자세하게 입력을 해야 자세한 답변이 나와요. 그냥 우울하다고 하면, 형식적이거나 이상한 소리를 하죠..” <참여자 1>

“제가 느끼기에 ChatGPT의 반응은 그냥 포장 같은 것.. 감정이 없는 통계적인 산물이라고 여겨져서 제가 ChatGPT와 교감을 한다는 생각조차 안하는 것 같아요. ChatGPT와는 감정이 없기 때문에 제 감정을 확실히 알아주고 있는 건지... 제 개인적으로는 다가가기 어려운 점이 있어요 정서적인 벽이 있는.. 가까이 다가가지 못하는 단점이 있는 것

같아요.” <참여자 5>

“GPT도 그렇긴 한데 특히 제미나이 같은 경우는 내가 무슨 말을 듣고 싶어 하는지 파악해서 반응해 주는 느낌이 들었어요. 좋은 말은 해 주는데 핵심을 파악하거나 진짜 중요한 반응은 한계가 있는 거죠.” <참여자 7>

11) 책임 주체의 불명확성 인식

연구 참여자들은 AI와 심리상담을 하지만 어디까지나 기계라는 인식을 가지고 있으며 상담성과나 결과에 대해서도 AI는 실체가 없다는 사실에 대해서도 감안하고 있었다. 이러한 사실은 일반 성인에게는 특별한 어려움이 예상되지 않지만, 심리적으로 취약한 내담자의 경우에는 그 위험이 크기 때문에 해당 윤리적 이슈에 대한 보완이 필요함을 언급해 주었다.

상담 결과에 대한 책임 소재 불분명

“AI는 실제가 없는 기계인데, 상담 내용에 대해서 책임을 질 수도 없고, 책임질 필요도 없죠. 실체가 없으니까...” <참여자 6>

문제가 발생했을 때 책임 주체 불명확

“인터넷 뉴스에서 나온 GPT를 사용한 미성년자의 자살에 관한 이야기는 충격적이었습니다. 기계가 어떻게 책임지겠습니까.. 아휴, 그 부모 생각하면... 끔찍하죠...” <참여자 3>

“AI가 제공한 정보가 옳고 그른지에 대한 책임을 질 필요가 없죠. AI는 인간 질문에 대한 답만 해주면 되니까.. 실체가 없는 거죠.” <참여자 5>

12) 윤리적 보호 체계에 대한 요구

AI를 활용한 심리상담의 경우, 내담자를 보호하거나 위험으로부터 예방하기 위한 시스템적인 보완이 시급하다고 언급해 주었다. 대부분의 연구 참여자들은 AI의 정보 제공을 상담전문가와 공유하여 개인상담에서 활용하는 방안으로 윤리적인 보호 체계를 마련할 수 있으며 AI는 어디까지나 전문가의 보완재 역할로 기능할 필요가 있다고 표현해 주었다. 특히, 윤리적 보호를 위하여 심리적으로 매우 취약하거나 의사결정이 어려운 내담자와 미성년자들의 경우에는 AI 사용에 대한 보호 장치가 필요함도 보고해 주었다.

사용자 보호 장치의 필요성 인식

“상담을 하다 보면 저의 개인정보가 많이 입력되는데 아무래도 걱정이 됩니다. 시스템적으로나 법적으로나 보완이 필요할 것 같아요” <참여자 1>

“AI가 발전하면서 개인정보 유출이나 특히 요즘 보이스 피싱도 그렇고, 음성이나 영상도 조작이 가능하니까 무섭죠. 특히 상담같이 개인적 이야기를 하는데 비밀보장에 대해서는 보완이 필요한 것 같습니다” <참여자 2>

“AI가 제공해주는 반응이 가짜인지 진짜인지 판단하기 어렵거나 무조건적으로 수용하는 내담자의 경우에는 매우 위험할 수 있다고 봅니다. 얼마 전 뉴스에서 자살 관련 이야기를 AI와 하다가 실제 자살로 이어진 사례를 보면, 보호 장치나 시스템이 매우 필요하다고 생각됩니다” <참여자 6>

AI에 대한 윤리적 보호 시스템 구축의 필요성 제고

“AI가 제공한 정보가 옳고 그른지에 대한 책임을 질 필요가 없죠. AI는 인간 질문에 대한 답만 해주면 되니까.. 실체가 없는 거죠.” <참여자 5>

“실제 제 분야에서 GPT를 사용해 보면, 아직 전문가 수준은 아니라는 것을 체험하다 보니까, GPT랑 상담을 할 때도 전문적이지 않을 수 있겠다. 잘못된 정보도 있을 수 있겠다고 생각을 하지만, 어린 아이들이나 노인분들과 같이 잘 모르시는 분들은 위험할 수 있을 것 같긴 합니다. 이에 대한 보완이 필요할 것 같아요.” <참여자 7>

13) 심리상담 접근 기회의 확대

상담자는 내담자에게 안전한 상담 환경을 제공할 책무가 있다. AI는 인간상담자보다 심리상담 접근 기회를 확대하고 있으며, 경제적 부담을 실질적으로 감소시키는 도움과 상담 접근 기회를 확대하고 있다는 점에서 내담자에게 분명한 윤리적 이익을 제공하고 있음을 확인할 수 있었다.

경제적 부담 감소

“상담료보다 GPT 유료 사용이 더 저렴하니까 아무래도 경제적으로 부담이 덜하죠.” <참여자 1>

“남편하고, 아이들하고 싸울 때마다 심리상담을 받는 게, 경제적으로 부담이 되죠.” <참여자 2>

“상담을 할 때도 AI 사용 구독료 말고는 다른 비용을 지불하지 않으니까 실제 심리상담료 보다 아주 저렴하고..” <참여자 7>

상담 접근 기회의 평등성 증가

“은퇴하고 돈을 벌지 않는 노인분들은 거동도 어렵고, 상담받기가 어렵죠. 그에 비해 휴대폰을 켜서 이야기라도 할 수 있는게 정서적으로 도움이 많이 될 것 같습니다. 실제 부모님들께서 AI 사용을 많이 하시더라구요.” <참여자 2>

“상담료가 저렴한 편은 아니거든요. 무료상담도 요즘은 많아졌지만, 보통 8~10만원 정도 회당 하니까... 경제적으로 어려운 사람들은 상담은 사실 엄두가 안 나죠.. 정작 필요한 사람들은 경제적으로 어려운 사람일 수도 있는데... 그런 면에서 GPT는 돈이 안 들거나 저렴하니까, 좀 더 많은 사람들이 상담을 받을 수 있는 건 장점인 것 같아요.” <참여자 4>

14) 서비스 적절성에 대한 고려

AI와 심리상담을 하는 과정에서 연구 참여자들은 AI의 반응이 모든 사람에게 적절하지 않을 수 있으며 형식적인 반응으로 도움이 되지 않았던 경험을 보고해주었다. 이는 취약한 내담자들에게 부적절한 의사결정을 내릴 수 있도록 할 수 있다는 점에서 취약집단의 보호 필요성과 AI 심리상담에 대한 위험성에 대한 외부 사례 등에 대해서도 언급해주었다.

모든 사람에게 적합하지 않을 수 있음

“AI는 실체가 없는 기계입니다. 정서적 위로 정도는 해 줄 수 있지만 모든 사람들에게 도움이 될 수 없죠. 당연한 겁니다. 필요한 정보만 사람이 받아들이면 되는거죠.” <참여자 5>

“AI와 심리상담에 백프로 정확하거나 전문적이라고 느껴지지는 않습니다. 아직 현 단계에서는 잘못된 정보도 자주 알려주니까 위험할 수도 있다고 봅니다. 제가 GPT한테 상담을 받는 것은 세컨 어드바이스 정도죠. 기계는 이 상황을 어떻게 생각하는지에 대한 어드바이스를 얻는 정도... 미성년자나 심리적으로 취약한 사람들에게는 매우 위험할 수도 있다고 봅니다.” <참여자 7>

취약 집단 보호 필요성 인식

“정신과에서 진료를 볼 때 가능하면 GPT랑 진료에 대해서는 이야기하지 말고 병원에 와서 물어보

라고 하더라구요. 아직 학습되지 않은 부분도 있어서 진단받은 내용에 대해서는 잘못된 정보도 많이 준다고 했어요.” <참여자 1>

“실제로 뉴스에서 본 것 같아요. ChatGPT한테 자살하고 싶어 방법을 물어봤을 때, 어떻게 죽을 수 있다고 ChatGPT가 이야기를 해줘서 결국에 생명을 잃게 되는 상황이 발생하는... 무분별하게 ChatGPT를 사용하는 건 아직은 위험한 것 같아요. 심리적으로 취약한 사람들이거나 아이들 같은 경우에는 반드시 성인이나 전문가랑 같이 사용해야 할 것 같아요.” <참여자 6>

이상의 연구 결과를 요약하면, AI 인공지능을 활용한 심리상담에서 내담자는 개인이 심리적 어려움을 경험하는 상황에서 형성되었음을 알 수 있다. 특히 연구 참여자들은 시간적, 경제적, 심리적 제약으로 인해 대면상담에 접근하기 어려운 상황에서 AI를 활용한 심리상담을 선택하게 된 것으로 나타났다. 연구 참여자들은 AI를 활용한 심리상담이 시간과 장소의 제약 없이 이용 가능하며, 이용 여부와 상담의 지속 여부를 스스로 결정할 수 있다는 점에서 자율성이 존중되는 경험을 하고 있었다. 특히 언제든지 상담을 시작하고 종료할 수 있다는 점은 자기결정권이 강화되는 경험으로 보고하였다. 동시에 일부 참여자들은 AI의 반응이 반복적으로 제공되는 과정에서 심리적으로 의존하거나 AI의 제안을 충분히 비판적으로 검토하지 않고 수용할 가능성에 대해 우려를 표현하였다. 또한, 참여자들은 익명성이 보장된 환경에서 자신의 경험과 감정을 보다 자유롭게 표현할 수 있었으며, 이는 심리적 안전감을 증진시키는 요인으로 작용하였다. 그러나 자신이 입력한 개인정보가 어디에 저장되고 어떻게 활용되는지에 대한 명확한 정보를 제공 받지 못했다고 인식하였으며, 이는 개인정보 보호와 비밀보장에 대한 불확실성과 불안을 보고하였다.

AI를 활용한 심리상담에서 참여자들은 정서적 위로와 심리적 안정감을 경험하였으며, 이는 AI를 활용한 심리상담이 내담자들의 심리적 복지를 돕고 긍정적 역할을 수행하고 있음을 보여준다. 특히 필요할 때 즉시 접근할 수 있다는 점은 심리적 도움의 접근성을 크게 향상시키는 요소로 작용하였다. 그러나 AI가 제공하는 조언이 항상 적절하거나 개인의 상황을 충분히 반영하지 못할 수 있다는 점에서 연구 참여자들은 AI의 반응에 대한 무조건적인 활용에 대하여 잠재적인 위험 가능성을 인식하고 있었다. 일부 참여자들은 AI 상담에 의존함으로써 인간과 심리상담을 회피하게 될 가능성에 대해서도 우려를 표현하였다. 참여자들은 AI가 항상 비판하지 않고 일관된 태도로 반응한다는 점에서 신뢰감을 경험하기도 하였으나, 이러한 반응이 실제 이해에 기반한 것이 아니라 알고리즘에 의해 생성된 것이라는 점에서 상담 관계의 진정성에 대한 의문을 경험하였다. 또한, AI를 활용하여 심리상담을 하는 과정에서 문제가 발생할 경우, 누구에게 책임이 있는지 명확하지 않다는 점에서 윤리적 책임성에 대한 불확실성을 보고하였다. 연구 참여자들은 이러한 윤리적 이슈에 대한 경험을 통해 AI를 활용한 심리상담은 향후 윤리적 기준과 보호 체계가 매우 필요하다는 인식을 제기해 주었다. AI를 활용한 심리상담이 경제적 부담을 줄이고 심리상담에 대한 접근 기회를 확대한다는 점에서 정의의 원칙이 실현되는 장점이 있으나 이러한 서비스가 모든 사람에게 적합하지 않을 수 있으며, 특히 심리적으로 취약한 집단에 대해서는 보다 신중한 적용이 필요할 수 있음을 제안하였다.

2. 핵심적 의미의 도출

이상의 요인별 분석을 종합하여 AI를 활용한 심리상담에서 내담자가 지각한 윤리적 이슈의 핵심적 의미를 살펴보면 “접근성이 용이하여 심리적 도움

을 쉽게 얻을 수 있지만 비밀보장과 책임성 등의 윤리적 신뢰의 필요성을 지각함"'으로 도출할 수 있다. 내담자는 AI를 활용한 심리상담을 통해 시간과 장소의 제약 없이 자신의 필요에 따라 상담을 선택하고 이용할 수 있었으며, 이는 자기결정권과 자율성이 존중되는 경험으로 인식되었다. 또한 익명성이 보장된 환경은 심리적 안전감을 증진시키고 자유로운 자기개방을 가능하게 하였다.

AI를 활용한 심리상담은 즉시성을 강점으로 정서적 위로와 심리적 안정감을 제공함으로써 내담자의 심리적 복지에 기여하는 선행의 기능을 수행하였다. 동시에 심리상담에 대한 접근성을 확대함으로써 보다 많은 사람들이 심리적 도움을 받을 수 있도록 하는 정의의 원칙을 실현하는 가능성을 보여주었다. 그러나 내담자는 자신의 개인정보가 어떻게 보호되고 활용되는지에 대한 명확한 이해가 부족한 상태에서 AI를 활용하여 심리상담을 받았으며 이는 비밀보장에 대한 신뢰 형성을 제한하는 요인으로 작용하였다. 또한 AI가 제공하는 상담적 반응이 실제 이해와 책임에 기반한 것이 아니라 기술적 알고리즘에 의해 생성된 것이기에 사용자들은 상담 관계의 진정성과 신뢰성에 대한 근본적 의문을 제기하였다. 연구 참여자들은 AI를 활용한 심리상담이 심리적인 지지와 실질적인 도움을 제공할 수 있는 유용한 자원임을 인정하면서도 상담 과정에서 발생할 수 있는 위험과 책임의 불명확성에 대해 윤리적인 한계가 있음을 보고하고 있다. 따라서 AI를 활용한 심리상담에서 연구 참여자들의 윤리적 경험의 본질은 자율성, 선행, 정의의 윤리적 가능성과 비밀보장, 무해, 신뢰성, 책임성에 대한 윤리적 불확실성이 상호작용하는 가운데, 심리적 도움과 윤리적 안전성 사이에서 균형을 모색하여야 함을 확인할 수 있다.

마지막으로 본 연구에서 도출된 AI를 활용한 심리상담에서 내담자의 윤리적 이슈의 핵심적 의미는 '"접근성이 용이하여 심리적 도움을 쉽게 얻을 수 있지만 비밀보장과 책임성 등의 윤리적 신뢰의 필요성을 지각함"으로 이해할 수 있다. 즉 AI 기반 상담을 통해 연구 참여자들은 시간과 장소의 제약 없이 심리적 도움에 접근할 수 있었으며, 상담 이용 여부와 상담의 지속 여부를 스스로 결정할 수 있다는 점에서 자율성이 존중되는 경험을 하였다. 이러한 경험은 심리적 자기결정권과 심리적 도움에 대한 접근권이 확장되는 윤리적 가능성을 의미한다. 또한, 익명성이 보장된 환경은 사용자의 평가나 낙인에 대한 두려움 없이 자신의 경험을 표현할 수 있도록 하였으며 이는 심리적 안전감을 증진시키는 중요한 요인으로 작용하였다. 이는 연구 참여자에게 정서적 위로와 심리적 안정감을 제공함으로써 사용자들의 심리적 복지에 기여하는 선행의 기능을 수행하였다.

동시에 연구 참여자들은 자신의 개인정보가 어떻게 보호되고 활용되는지에 대한 명확한 이해 없이 AI를 활용하여 심리상담을 하였으며 이는 비밀보장에 대한 신뢰 형성을 제한하는 요인으로 작용하였다. 또한, AI의 반응이 실제 이해와 책임에 기반한 것이 아니라 기술적 알고리즘에 의해 생성된 것이라는 인식은 상담 관계의 신뢰성과 책임성에 대한 근본적 불확실성을 야기하였다.

AI를 활용한 심리상담 과정에서 사용자들은 이러한 윤리적 불확실성에도 불구하고 AI를 활용한 심리상담의 상담 반응에 대한 유용성과 접근성을 고려하여 제한적으로 신뢰하고 활용하였다. 따라서 AI를 활용한 심리상담에서 내담자의 윤리적 경험은 자율성, 선행, 정의의 윤리적 가능성과 비밀보장, 무해, 신뢰성, 책임성의 윤리적 불확실성이 상호작용하는 가운데, 심리적 도움에 대한 욕구와 윤리적 안전성에 대한 요구 사이에서 균형을 형성하는 경험으로 이해할 수 있다.

Ⅳ. 논의 및 결론

본 연구에서는 AI를 활용한 심리상담에서 내담자가 지각한 윤리적 이슈는 어떠하며 핵심적 의미는 무엇인지 살펴보기 위하여 질적 내용 분석으로 생성형 AI와 인간상담자와 심리상담 경험이 모두 있는 일반 성인 7인을 인터뷰하여 분석하였다. 그 결과 14개의 구성요인과 28개의 하위 구성요인을 확인하였으며 “접근성이 용이하여 심리적 도움을 쉽게 얻을 수 있지만 비밀보장과 책임성 등의 윤리적 신뢰의 필요성을 지각함”이라는 핵심적 의미를 도출하였다.

참여자들은 시간과 장소의 제약 없이 필요할 때 AI를 이용할 수 있다는 점을 긍정적으로 경험하였다. 이러한 접근성과 즉시성은 심리적 어려움이 발생했을 때 도움을 쉽게 요청할 수 있다는 가능성을 제공하였다. 그러나 일부 참여자들은 이러한 편리성이 반복적인 AI 이용이나 의존으로 이어질 가능성에 대해서도 우려를 나타냈다. 즉, AI의 높은 접근성은 심리적 지원에 대한 접근을 확대하는 가능성과 동시에 이용자의 자기판단이나 자율적 문제해결을 약화시킬 수 있는 위험이 함께 존재하는 양면적 특성으로 나타났다.

본 연구에서 나타난 AI 심리상담 경험은 단순히 긍정적 또는 부정적인 특성으로 구분되기보다는, 하나의 특성이 심리적 지원의 가능성을 확대하는 동시에 새로운 윤리적 위험을 발생시킬 수 있는 양면적 특성을 보였다. 이에 본 연구에서는 앞서 도출된 결과를 바탕으로 AI 심리상담의 주요 특성을 ‘가능성과 위험의 양면성’이라는 관점에서 통합적 관점에서 다음과 같이 논의하였다.

<표 3> 연구 결과의 통합적 해석

AI 심리상담의 특성	긍정적 가능성	잠재적 위험 및 윤리적 쟁점
접근성과 즉시성	필요할 때 쉽게 심리적 지원 이용	AI 의존 및 자기판단 약화 가능성
익명성	심리적 편안함, 낙인 감소	개인정보 저장·활용에 대한 불확실성
즉각적인 정서적 반응	위로와 정서적 지원	부정확하거나 맥락을 고려하지 못한 조언
일관되고 비판하지 않는 반응	심리적 안전감과 편안함	실제 이해와 진정성에 대한 의문
낮은 비용과 높은 접근성	심리적 지원의 접근성 확대	취약집단 및 디지털 격차
자유로운 이용	도움 요청의 자율성 확대	책임주체와 보호 체계의 불명확성

첫째, AI의 높은 접근성과 즉시성은 시간과 장소의 제약 없이 필요할 때 심리적 지원을 받을 수 있다는 가능성을 제공하고 있다. 특히 전문적인 도움을 요청하기 어려운 상황에서도 비교적 쉽게 자신의 어려움을 표현하고 도움을 탐색할 수 있다는 점에서 내담자들의 자율적인 선택과 심리적 지원에 대한 접근성을 확대할 수 있다. 그러나 이러한 편의성과 접근성이 반복적인 이용이나 AI에 대한 과도한 의존으로 이어질 경우, 내담자의 자기판단이나 스스로 문제를 해결하려는 능력을 약화시킬 가능성도 함께 고려할 필요가 있다. 따라서 AI의 접근성은 그 자체로 윤리적 쟁점이라기보다, 내담자의 자율성과 자기결정에 어떠한 영향을 미치는가에 따라 윤리적 의미를 고려할 필요가 있다.

둘째, 익명성은 내담자가 타인의 평가나 낙인에 대한 부담 없이 자신의 심리적 어려움을 표현할 수 있도록 하여 심리적 편안함을 높이는 가능성을 제

공하였다. 반면 내담자는 자신이 제공한 개인정보나 대화 내용이 어떻게 저장되고 활용되는지에 대해 불확실성을 경험할 수 있었다. 이는 익명성이 심리적 안전감을 높이는 동시에 개인정보 보호와 관련된 새로운 우려를 발생시킬 수 있음을 보여준다. 따라서 AI 심리상담에서 익명성의 장점은 개인정보의 수집·저장·활용에 대한 투명성과 함께 고려될 필요가 있다.

셋째, AI의 즉각적이고 일관된 반응은 내담자가 필요할 때 위로와 정서적 지지를 받을 수 있도록 한다는 점에서 긍정적으로 경험되고 있다. 그러나 생성형 AI가 내담자의 개인적·맥락적 상황을 충분히 이해하지 못한 상태에서 부정확하거나 부적절한 조언을 제공할 가능성도 존재함을 확인하였다. 특히, 심리적 어려움이 심각하거나 복합적인 경우에는 AI의 응답이 내담자의 고유한 특성과 상황을 충분히 반영하지 못할 수 있으며, 잘못된 정보나 조언이 오히려 위해로 이어질 가능성을 고려해야 한다. 따라서 AI가 제공하는 정서적 지원의 가치는 응답의 정확성 및 안전성과 함께 검토될 필요가 있다.

넷째, AI가 비판이나 평가 없이 일관되게 반응한다는 특성은 내담자에게 심리적 부담을 낮추고 자신의 경험을 자유롭게 표현할 수 있는 환경을 제공하였다. 그러나 이러한 상호작용이 실제 인간 상담자와 같은 이해나 공감, 관계적 진정성을 의미하는 것은 아니다. 내담자가 AI의 반응을 실제 이해나 공감으로 과도하게 해석할 경우 AI와의 관계에 대한 기대와 실제 기능 사이에 차이가 발생할 수 있다. 따라서 AI의 비판 없는 반응과 일관성은 심리적 지원의 가능성을 제공하지만, 그 한계와 비인간적 상호작용의 특성 또한 함께 고려할 필요가 있다.

다섯째, 낮은 비용과 높은 접근성은 심리적 지원에 대한 경제적·물리적 접근성을 높일 수 있다는 가능성을 보여주고 있다. 그러나 디지털 기기와 정보 활용 능력의 차이는 이용 기회의 차이로 이어질 수 있으며, 특히 디지털 접근성이 낮은 집단은 이러한 혜택에서 상대적으로 배제될 가능성이 있다. 따라서 AI 심리상담의 확산은 단순히 서비스 접근성을 확대하는 측면뿐만 아니라, 디지털 격차와 취약집단의 접근성 문제를 함께 고려할 필요가 있다.

마지막으로, AI를 이용하여 자유롭게 도움을 요청할 수 있다는 특성은 기존의 전문적 상담서비스에 접근하기 어려운 내담자에게 새로운 심리적 지원의 가능성을 제공한다. 그러나 AI 사용 과정에서 문제가 발생했을 때 누가 책임을 지며 어떠한 보호체계를 통해 이용자를 지원할 수 있는지는 불분명할 수 있다. 이는 AI 심리상담이 인간상담자와의 전문적 상담관계와 달리 책임과 보호의 주체가 명확하지 않을 수 있음을 보여준다. 따라서 AI 심리상담의 자율적 이용 가능성이 확대되기 위해서는 이용자의 자유로운 선택뿐만 아니라 위험 상황에서의 보호체계와 책임소재에 대한 제도적·윤리적 고려가 함께 이루어질 필요가 있다.

이상의 연구 결과를 종합하면, AI 심리상담의 윤리적 특성은 개별적인 장점이나 단점으로 분리하기보다 하나의 기술적 특성이 긍정적 가능성과 잠재적 위험을 동시에 발생시키는 양면적 구조로 이해할 필요가 있다. 즉, 접근성은 지원의 확대와 의존 가능성을 동시에 포함하고, 익명성은 심리적 편안함과 개인정보 보호의 불확실성을 동시에 수반하며, 즉각적인 정서적 지원은 도움의 가능성과 부정확한 조언으로 인한 위해 가능성을 함께 내포한다. 이러한 양면성은 AI 심리상담의 윤리적 쟁점이 기술의 특정 특성 자체에 존재하는 것이 아니라, 해당 특성이 이용자의 자율성, 안전, 개인정보 보호, 형평성, 책임성 등 윤리적 가치와 상호작용하는 과정에서 발생함을 시사한다.

다음으로, 이상의 연구 결과를 <표 4>에서는 Kitchener(1984)의 윤리 원칙과 연결하여 그 이론적 의미를 더욱 확장하여 논의해 보고자 한다.

이는 Kitchener의 윤리 원칙을 기준으로 참여자의 진술을 재분류한 것이 아니라, 귀납적으로 도출

된 요인들이 각각 어떠한 윤리적 원칙과 관련되는지를 사후적으로 검토한 것이다.

Kitchener, K. S. (1984)의 상담 윤리의 다섯 가지 원칙을 적용하면서도 비밀보장을 별도의 범주에 포함한 이유는 AI 기반 심리상담이라는 기술적 맥락에서 비밀보장이 단순히 충실성(fidelity)의 하위 요소를 넘어 독립적이고 구조적인 윤리 쟁점으로 경험되었기 때문으로 앞서 언급한 바 있다. Kitchener, K. S. (1984)의 원칙은 전통적 대면상담 관계를 전제로 하여 신뢰와 약속 이행의 도덕적 책임을 강조하지만, AI를 활용한 인공지능을 심리상담에서는 상담 내용이 데이터화되어 저장·처리·활용되는 기술적 과정을 포함하므로 개인정보 보호, 데이터 보안, 제3자 접근 가능성 등 관계적 차원을 넘어선 구조적 문제로 확장될 필요가 있으며(윤석주 외, 2025; 조은이 외, 2025). 본 연구의 참여자들은 상담 이용의 통제 가능성은 자율성으로 인식한 반면, 자신의 정보가 어떻게 관리되는지에 대한 불확실성은 별도의 불안과 우려를 보고하며 비밀보장이 자율성이나 충실성에 종속되기보다 독립적 판단 기준으로 작용하고 있음을 확인할 수 있음을 고려하였다(Kitchener, K. S., 1984; 강수정 외 2024).

<표 4> 연구 결과와 Kitchener 윤리 이론의 적용

구성요인	윤리적 범주
자기 의사결정권의 확대 경험	자율성
자율성의 제한 가능성에 대한 우려	
익명성으로 인한 심리적 안전감	
개인정보 보호에 대한 불확실성	선행
정서적 도움과 심리적 지지 경험	
심리적 접근성의 편리	
부정적 상담 성과 가능성에 대한 우려	무해
상담 관계 대체에 따른 사회적 단절 가능성	
상담 반응의 일관성 경험	신뢰성
상담 진성성에 대한 의문	
책임 주체의 불명확성 인식	책임성
윤리적 보호 체계에 대한 요구	
심리상담 접근 기회의 확대	
서비스 적절성에 대한 고려	정의

1. 자율성 존중 원칙과 내담자의 자기결정권 확대 및 제한 가능성

자율성 존중의 원칙(respect for autonomy)은 상담윤리에서 내담자가 자신의 삶과 관련된 결정을 스스로 내릴 권리를 존중하는 것을 의미한다(Corey et al., 2010). 이는 상담자가 내담자의 선택을 통제하거나 강요하는 것이 아니라, 내담자가 충분한 정보에 기반하여 자유롭게 선택할 수 있도록 지원해야 함을 의미한다. 본 연구에서 연구 참여자들은 AI를 활용한 심리상담이 시간과 장소의 제약 없이 이용 가능하며, 상담의 시작과 종료를 스스로 결정할 수 있다는 점에서 자율성이 향상되는 경험을 하였다. 이는 전통적 상담 환경에서 발생할 수 있는 구조적 제약(예약, 비용, 접근성 등)이 완화됨으로써 내담자의 자기결정권이 확대되는 경험으로 이해할 수 있다. 이러한 결과는 AI 기반 심리상담이 내담자의 주체성을 강화할 수 있는 윤리적 가능성을 지니고 있음을 시사한다.

그러나 동시에 연구 참여자들은 AI의 심리적 반응과 제안을 자율적으로 심사숙고하거나 검토하기보다는 무비판적으로 수용하게 될 가능성을 제기하며 이에 대한 우려를 보고하였다. 이는 자율성이 단순히 선택의 기회를 제공받는 의미가 아니라 충분한 정보와 이해를 기반으로 한 실질적 자율성(substantive autonomy)이 중요함을 보여준다. 상담윤리에서는 내담자의 자율성이 진정으로 보장되기 위해서는 충분한 정보 제공과 이해가 전제되어야 한다고 강조한다(Kitchener, 1984). 이러한 맥락에서

향후 AI를 활용한 심리상담에서 내담자의 자율성을 존중하는 윤리 원칙이 보장되기 위해서는 AI의 한계와 데이터 처리 방식 등에 대한 명확한 설명이 제공될 필요가 있으며 이를 통해 내담자가 충분한 정보를 가지고 자발적인 선택을 할 필요가 있다.

2. 비밀보장 원칙과 심리적 안전성 및 윤리적 신뢰 형성

비밀보장(confidentiality)은 상담윤리의 핵심 원칙 중 하나로, 내담자의 개인정보와 상담 내용이 보호되어야 함을 의미한다. 비밀보장은 상담 관계에서 신뢰 형성의 필수적 전제 조건이며, 내담자의 자기개방을 촉진하는 핵심 요소이다(Corey et al., 2010). 본 연구에서 참여자들은 AI를 활용하여 심리상담을 하면서 익명성이 보장된 환경에서 심리적 안전감을 경험하였으며 이는 자유로운 자기개방을 가능하게 하는 중요한 요인으로 작용함을 보고해주었다. 이는 비밀보장이 단순히 정보 보호의 기능을 넘어, 내담자의 심리적 안전감을 형성하는 윤리적 기반임을 보여준다.

그러나 동시에 참여자들은 AI를 활용한 심리상담 과정에서 자신의 개인정보가 어떻게 저장되고 활용되는지에 대한 명확한 이해 없이 AI를 이용하였으며, 이는 비밀보장에 대한 신뢰 형성을 제한하는 요인으로 작용하였다. 이는 상담윤리에서 강조하는 '사전 동의(informed consent)'의 원칙과 밀접하게 관련된다. 사전 동의는 내담자가 상담 과정과 관련된 정보를 충분히 이해하고 자발적으로 참여하는 것을 의미하며, 이는 자율성과 비밀보장 원칙의 실질적 구현을 위한 필수 조건이다. 향후 AI를 활용하여 심리상담을 할 때는 내담자에 대한 사적이고 민감한 데이터가 기술 시스템에 의해 처리된다는 점에서 전통적 상담과 다른 윤리적 고려가 필요할 것으로 사료된다. 따라서 AI를 활용한 심리상담에서도 인간상담자와 같이 비밀보장과 익명성을 보장해 줄 것으로 기대하고 있으나 실제 상담이 진행됨에 따라 AI 기반 심리상담에서는 비밀보장에 대한 윤리적 염려를 한다는 것을 확인할 수 있었다. 비밀보장에 대한 윤리적 안전을 확보하기 위해서는 향후 내담자들의 데이터 보호와 활용 방식에 대한 명확한 안내 및 정보 제공이 AI 기반 심리상담에서 특히 필수적일 것으로 사료된다.

3. 선행과 무해의 원칙: 심리적 도움의 가능성과 윤리적 위험

선행(beneficence)의 원칙은 내담자의 복지를 증진시키는 방향으로 행동해야 함을 의미하며, 무해(nonmaleficence)의 원칙은 내담자에게 해를 끼치지 않아야 함을 의미한다(Kitchener, 1984). 이 두 원칙은 상담자의 윤리적 의사결정에서 가장 기본적인 기준이 된다. 본 연구에서 참여자들은 AI를 활용한 심리상담을 통해 정서적 위로와 심리적 안정감을 경험하였으며 이는 AI를 활용한 심리상담이 실제 사용자들의 심리적 복지를 증진시키는 선행의 기능을 수행할 수 있음을 보여준다. 특히 접근성이 제한된 상황에서 AI를 활용한 심리상담은 심리적 개입이 제공되는 중요한 대안적 자원으로 기능할 수 있다(Kuhail, M. A., et al., 2024)

그러나 참여자들은 AI의 상담 반응이 개개인의 고유한 상황을 충분히 반영하지 못할 가능성에 대해 우려를 보고하고 있다(Keung, W. M., et al, 2025; Rosen, C., 2025). 부적절한 조언이나 내담자에 대한 충분한 이해가 부족한 상황에서 상담적 제안 및 반응은 사용자에게 심리적 혼란이나 부정적 영향을 미칠 가능성이 제기된다. 이는 무해의 원칙과 관련된 중요한 윤리적 고려 사항이다. 상담윤리에서는 상담자의 전문성이 내담자와의 복지 및 보호에 핵심적 역할을 한다고 강조한다(Corey et al., 2015). 실제 AI를 활용한 심리상담에서 사용자에게 제공되는 상담적 제안과 반응이 전문성과 신뢰성을

확보한 윤리적인 상담 개입인가에 대해서는 확신하기 어렵다는 것을 본 연구 결과를 통해 확인할 수 있었다. AI의 임상적 판단에 대한 윤리적 고려 및 심리적 개입이 기술적 알고리즘에 의해 이루어진다는 점에서 윤리적 책임과 안전성 확보를 위한 추가적 보완이 시급할 것으로 고려된다.

4. 신뢰성과 책임성의 원칙: 상담 관계의 윤리적 기반

신뢰성(fidelity)은 상담자가 내담자와의 관계에서 신뢰를 유지하고 책임감 있는 태도를 유지해야 함을 의미한다(Kitchener, 1984). 이는 상담 관계의 윤리적 기반을 형성하는 핵심 원칙이다. 본 연구에서 참여자들은 AI의 심리적 개입이 일관되고, 비판적이지 않다는 점에서 일정 수준의 신뢰감을 보고하고 있으나 동시에 AI가 실질적으로 내담자인 자신을 이해하는 존재가 아니라는 인식을 통하여 상담 관계의 진정성 대한 근본적 의문을 보고하였다. 이는 상담 관계가 단순한 정보 교환이 아니라 상호주관적 이해(inter-subjective understanding)에 기반한 관계적 과정임을 보여준다(M.Buber, 1923; Rogers, 1957; Rosen, C., 2025). 관계윤리(relational ethics)의 관점에서 상담은 상호 책임성과 진정성을 기반으로 한 관계적 과정이며, 이러한 관계적 요소는 상담의 효과성과 윤리성에 중요한 역할을 한다. 또한 참여자들은 상담 과정에서 발생할 수 있는 문제에 대한 책임 주체가 불명확하다는 점에서 윤리적 불확실성을 보고해 주었다. 이는 AI를 활용한 심리상담에서 책임성(accountability)의 구조가 명확히 규정될 필요가 있음을 시사한다. 생성형 AI는 개발자, 서비스 제공자, 운영 기관 등 그 주체가 다양하기 때문에 향후 AI 기반 심리상담 서비스를 내담자에게 제공할 경우에는 제공하는 심리적 개입에 대한 책임성 소재를 명확히 구축할 필요가 있겠다.

5. 정의의 원칙: 심리상담 접근성의 확대와 윤리적 형평성

정의(justice)의 원칙은 상담 서비스가 공정하고 평등하게 제공되어야 함을 의미한다(Kitchener, 1984). 이는 심리상담에 대한 접근 기회가 특정 집단에 제한되지 않고 모든 사람에게 공정하게 제공되어야 함을 강조한다. 본 연구에서 참여자들은 AI를 활용한 심리상담이 경제적 부담을 줄이고 상담 접근성을 확대한다는 점에서 긍정적 경험을 보고하였다. 이는 AI를 활용한 심리상담이 접근의 불평등을 완화하는 데 이바지할 수 있는 가능성을 보여준다. 특히 심리상담 서비스에 접근하기 어려운 개인들에게 AI를 활용한 심리상담은 심리적 지원을 위한 효과적인 자원이 될 수 있을 것으로 예상된다. 이는 상담윤리에서 강조하는 사회적 정의(social justice)의 관점에서 중요한 의미를 가진다. 그러나 동시에 AI를 활용한 심리상담에 모든 내담자에게 동일하게 적합하지 않을 수 있으며 특히 심리적으로 취약한 집단에 대한 보호가 선행되어야 한다는 점에서 미성년자, 위기상황의 내담자, 임상적으로 유의미한 진단을 받은 내담자와 같이 독립적인 의사결정이 취약할 수 있는 내담자들에게 AI를 활용한 심리상담이 접근 불평등을 완화할 수 있다는 명목으로 AI를 활용한 심리상담을 독려하기에는 다양한 위험성과 한계가 있다는 점을 고려할 필요가 있다.

Ⅴ. 연구의 한계 및 제언

1. 정책적 함의

이상의 연구 결과와 논의를 고려하여 AI를 활용한 심리상담 과정에서 비밀보장, 신뢰성, 책임성, 무해의 원칙 등과 관련된 윤리적 과제를 확인할 수 있었다. 이는 AI를 활용한 심리상담이 기존의 전통적 상담윤리의 관점을 확장하여 급변하는 기술 환경에 적합한 새로운 윤리적 기준과 실천 지침이 필

요함을 제안하고 있다. 구체적으로 살펴보면 다음과 같다.

첫째, 본 연구에서 참여자들은 AI를 활용한 심리상담을 통해 사용자의 필요에 따라 상담을 선택하고 이용할 수 있다는 점에서 자율성이 존중되는 경험을 하였다. 이는 AI를 활용한 심리상담이 사용자들의 자율성을 증진시키는 중요한 심리지원을 돕는 자원으로 기능할 수 있으나 AI의 상담적 반응을 사용자가 무비판적으로 수용하는 경향에서 사용자의 자기결정권을 향상시키는데 도움이 되지 않을 수 있음을 고려할 수 있다. 사용자의 심리적 도움에 대한 접근 장벽을 낮추고 즉시적인 심리적 지지를 제공한다는 점에서 도움받을 수 있으나 사용자의 주체적인 의사결정 능력을 침해할 수 있을 것으로 보여진다. 따라서 윤리강령에서 강조하는 내담자의 자율성과 자기결정권 존중의 원칙이 기술적 환경에서도 구현될 수 있도록 보완될 필요가 있음을 시사한다.

둘째, 참여자들은 AI를 활용한 심리상담에서 기계와의 대화가 익명성을 보장하여 심리적 안정감을 제공하고 있으며 이는 자기개방을 촉진하는 중요한 요인으로 보고하였다. 이러한 결과는 비밀보장의 원칙이 내담자의 심리적 안정과 상담 참여에 중요한 역할을 한다는 기존 연구를 고려할 때 AI를 활용한 심리상담이 사용자에게 안전감을 제공할 수 있음을 예상할 수 있다. 하지만 이는 동시에 사용자들의 개인정보 보호와 사용자 고유의 특성에 대한 비밀보장 불확실성의 문제를 확인할 수 있다. AI를 활용한 심리상담에서 개인정보 보호와 관련된 투명성과 안정성은 사용자들의 윤리적 신뢰 형성에 핵심적인 요소로 작용할 것으로 예상된다. 따라서 향후 AI기반 심리상담서비스 과정에서 내담자의 개인정보 보호에 대한 윤리적 한계를 고려하여 관련 대책이 시급히 마련될 필요가 있을 것으로 사료된다.

셋째, 참여자들은 AI를 활용한 심리상담을 통해 정서적 위로와 심리적 안정감을 경험하였으며, 이는 선행의 윤리 원칙이 기술 기반 상담 환경에서도 실현 가능성을 확인할 수 있었다. 그러나 내담자의 고유한 특성과 상황을 충분히 반영하지 못한 상담적 개입은 사용자의 심리적 혼란과 부정적 영향의 위험이 존재하기에 무해의 원칙을 고려할 때 비윤리적이다. 따라서 AI를 활용한 심리상담이 향후 내담자들에게 윤리적인 한계가 있음을 사전에 충분히 안내할 필요가 있으며 이는 향후 상담 현장에서 관련 연구를 통해 AI를 활용한 심리상담의 윤리적 한계에 대한 보완책이 마련되어야 할 것이다.

넷째, 참여자들은 상담 과정에서 발생할 수 있는 다양한 윤리적 문제에 대한 책임 주체가 불명확하다는 점에서 윤리적 공백을 우려하고 있었다. 연구 참여자 중 개인정보 및 윤리적 문제의 책임이 AI를 개발 및 보급한 회사에 있을 것이라고 보고해주었지만 실제 세부적인 AI의 반응에 해당 회사에 대한 책임소재는 불분명한 상황이다. AI가 제공하는 정보에 대하여 사용자의 무비판적인 수용으로 인한 부정적인 결과가 발생하고 있으며 이에 사회적인 경고가 제기되고 있다(Ping, Y., 2024; 윤석주 외, 2025; 유지은, 2023). 따라서 AI를 활용한 심리상담에서 윤리적 책임 구조를 명확히 사용자들에게 안내될 필요가 있으며 공신력 있는 상담 관련 기관 및 학회에서도 관련 규정을 안내할 필요가 있다.

다섯째, 참여자들은 AI를 활용한 심리상담이 경제적 부담을 줄이고 심리상담의 접근성을 용이하게 한다는 점에서 정의 원칙의 실현 가능성을 확인하였다. 이는 AI기반 심리상담이 접근 불평등을 완화하는 데 기여할 수 있는 순기능을 보여준다. 하지만 이러한 생성형 AI의 접근 용이성은 정서적 및 의사결정이 취약한 내담자에게는 위험하거나 잘못된 정보를 선별하지 못하고 AI가 제공한 정보를 그대로 수용함으로써 위험한 상황으로 이어질 수 있는 가능성도 동시에 고려해야 할 필요가 있다.

마지막으로, 본 연구는 AI를 활용한 심리상담에서 내담자가 보고한 윤리적 쟁점을 심도 깊게 분석

하였으며 이 과정에서 AI 심리상담의 가능성과 윤리적 쟁점을 비교하여 살펴보기도 하고 키츠너의 윤리 이론에 적용하여 AI를 활용한 심리상담에서 내담자의 윤리적 쟁점을 확장하여 그 의미를 탐색하고 있다는 점에서 의의가 있다.

더불어 생성형 AI가 정신건강 영역에 대체제로 활용 가능성을 논의하는 시대적 흐름 속에서 AI를 활용한 윤리적 활용 가능성과 제안을 논의함으로써 임상적 함의를 갖는다. 이상의 AI를 활용한 심리상담에서 윤리적 쟁점과 논의는 기술적인 허용 여부에 대한 논의를 넘어 향후 어떠한 상황에서 AI를 심리상담 분야에서 윤리적으로 활용할 것인가를 질문과 탐색을 시작하는 계기가 될 것이다. 상담윤리는 기술 발전을 지양하고자 하는 규제를 지향하는 것이 아니라 인간의 존엄과 안전을 보장하기 위한 최소한의 사회적 합의를 촉진할 필요가 있다(Kuhail, M. A., et al., 2024; 정여주 외, 2024; 황매향, 2025). 따라서 본 연구에서 논의된 윤리적 쟁점을 고려하여 AI를 활용한 심리상담이 보다 안전하고 윤리적으로 활용될 수 있도록, 향후 관련 제도와 법적 근거 마련 및 정책적 정교화를 위한 지속적인 논의가 이루어질 필요가 있다.

2. 연구의 제한점 및 후속연구 제언

본 연구의 제한점과 후속연구에 대한 제언을 살펴보면, 본 연구에서는 참여자가 면담 내용을 사전에 이해하고 자신의 경험을 충분히 회상할 수 있도록 면담 질문을 약 일주일 전에 제공함으로써 질문을 사전에 확인한 참여자가 자신의 경험을 특정 방향으로 회상하거나 정리했을 가능성을 완전히 배제하기 어렵다. 본 연구에서 도출된 윤리적 쟁점은 참여자의 생생한 경험을, 있는 그대로 재현하기에는 한계가 있었음을 밝힌다. 본 연구 결과는 인터뷰 질문을 통한 참여자의 경험과 연구자의 해석이 상호작용하는 과정에서 도출된 결과로 이해할 필요가 있다. 따라서 향후 연구에서는 사전 질문 제공 방식을 달리하거나 보다 개방적인 질문으로 구성하여 보다 다양하고 심도 깊은 윤리 경험을 탐색할 것을 제안한다.

또한, 본 연구 참여자의 특성에서 나타난 바와 같이, 다양한 직업군에서 각기 다른 주호소문제를 가지고 있는 인간상담자와 AI상담을 모두 경험한 참여자를 대상으로 AI기반 심리상담의 윤리를 탐색하고 있다. 이는 특정 내담자 특성에 따른 윤리적 쟁점을 충분히 구분하여 살펴보는데 있다. 후속연구에서는 내담자의 특성 및 임상적 특성을 고려한 집단별로 나타나는 윤리적 쟁점을 구체적으로 탐색할 필요가 있다.

시대적 흐름에 따라 AI를 활용한 상담의 효과성과 안전성에 대한 장기적 실증 연구는 다양한 관점에서 지속적으로 후속 연구를 축적할 필요가 있으며 인간과 AI 간의 혼합 상담 모델의 윤리 기준을 구체화할 필요가 있다. 또한 다양한 관점에서 국내외의 고위험 AI 규제 체계를 분석하고, 정신건강 영역에 특화된 규범 설계 방안을 국내에 효과적으로 적용할 수 있는 방안을 탐색하는 연구가 요구된다. 이러한 학문적 축적은 정책 설계의 근거를 강화하고, 기술 발전과 인간 보호의 균형을 이루는 토대를 마련할 수 있을 것으로 기대한다.

참고문헌

강수정, 박성현, 이지은 (2024). 상담자의 비대면 상담 및 디지털 미디어의 윤리적 활용 연구: (사)한국상담심리학회원을 대상으로. 한국심리학회지: 상담 및 심리치료, 36(4), 1613-1644.

강수정, 유금란 (2018). 상담자의 온라인 윤리: 내담자 정보 검사 및 활용과 상담자 자기개방, 한국심리학회지: 상담 및 심리치료, 30(3), 575-600.

강태정, 조항, 지우영 (2024). 대학생의 챗GPT를 활용한 진로탐색의 효과에 대한 탐색적 연구. 연

구방법논총, 9(2), 29-60.
김광옥, 장기덕 (2025). G대학교 재학생의 생성형 AI에 대한 인식 및 경험에 대한 연구, 『상담심리교육복지』, 12(4), 171-182.
김도연, 조민기, 신희천 (2020). 상담 및 심리치료에서 인공지능 기술의 활용: 국외사례를 중심으로. 한국심리학회지: 상담 및 심리치료, 32(2), 821-847.
김명주 (2017). 인공지능 윤리의 필요성과 국내외 동향. 정보와통신, 34(10), 45-54.
김성현, 박효은, 이동훈 (2022). 정신건강 영역에서의 디지털 치료제 동향과 시사점, 한국심리치료학회지: 상담 및 심리치료, 34(4), 1401-1430.
김수경 (2017). 전문상담교사의 상담과정 중 역전이 극복경험에 대한 현상학적 연구, 경성대학교 대학원 박사학위논문.
김요완 (2026). 심리상담 시뮬레이션 도구로서의 AI 활용 가능성 탐색: 상담자-내담자 역할 수행에 대한 현상학적 연구, 미래사회, 17(1), 215-233.
김태완, 이정우, 조주희, 전영진, 이재현, 이동헌, 김은한 (2024). GPT 파인튜닝을 통한 대화형 AI 의료 플랫폼 개발 : 질병 예측 및 대응 웹 서비스 구현. Proceedings of KIIT Conference, 2024(5), 896-899.
김훈민, 최준기, 이재철, 정회경 (2025). 자연어 처리 및 챗GPT를 활용한 정신건강 진단 플랫폼 연구. 한국콘텐츠학회논문지, 25(5), 42-50.
김형수, 김옥진 (2009). 상담자의 윤리적 판단 모형: 철학적 논의와 사례 적용, 상담학 연구, 10(2), 701-717.
박민지 (2021). 상담자의 청소년 내담자 죽음 경험에 관한 현상학적 연구. 숙명여자대학교 대학원. 박사학위논문.
송용섭 (2021). 인공지능 상담의 윤리적 이슈와 대안, 기독교사회윤리, 49, 239-267.
안명숙 (2019). 인공지능(AI) 심리상담에 관한 소셜미디어 빅데이터 분석. 융복합지식학회논문지, 7(3), 65-75.
안세훈, 정옥란 (2021). 감정분석 기반 심리상담 AI 챗봇 시스템에 대한 연구. 한국IT서비스학회지, 20(3), 75-86.
여수진, 손승희 (2025). AI 기반 심리상담 연구의 국내 동향 분석: 주제범위 문헌 고찰을 중심으로. 한국상담심리교육복지학회: 상담심리교육복지 12(5), 11-27
유지은 (2023). Chat GPT를 활용한 재활운동상담사의 가능성과 한계. 한국운동재활학회 학술대회, 2023(11), 29-40.
윤석주, 임상호 (2025). AI 기술과 심리 상담의 융합 가능성과 한계에 대한 고찰, 산업진흥연구, 10(1), 203-216.
이동훈, 김주연, 김진주 (2015). 온라인 심리치료의 가능성과 한계에 대한 탐색적 연구, 한국심리학회지: 상담 및 심리치료, 27(3), 543-582.
이소윤 (2025). AI/로봇 기반 노인 돌봄의 윤리적 딜레마: 돌봄윤리 관점에서의 문헌연구, 상담심리교육복지, 12(5), 235-250.
이승록, 박정환 (2026). 인공지능 기반 상담 및 상담자 교육 연구의 체계적 문헌고찰: 상담심리학과 교육공학의 융합 관점, 지능정보융합과 미래교육, 5(5), 1-11.
이윤희, 김경민, 이상민 (2021). 비대면 상담 윤리 강령 기초 연구, 상담학연구, 22(5), 77-106.
이지언, 이대호, 이재길 (2023). AI 심리 상담 챗봇 라포 형성 전략에 따른 상담 챗봇 디자인 제안. 한국HCI학회 학술대회, 2023(2), 631-635.
장성식, 성승연 (2020). 심리상담 서비스 사전동의 정보에 대한 대학생의 이식과 경험, 한국심리학회지: 상담 및 심리치료, 32(4), 1967-1999.
정여주, 최민아, 김지영, 윤나래 (2024). AI리터러시 교육 동향과 초등학생을 위한 상담적 접근 고찰,

초등상담연구, 23(5), 705-729.
정진, 송미경 (2024). 학대 피해 청소년에 대한 상담자의 역전이 경험, 한국심리치료학회지, 16(1), 69-102.
정진 (2025). ChatGPT를 활용한 내담자의 심리상담 경험: 현상학적 연구 방법을 중심으로, 사회과학 담론과 정책, 18(2), 229-261.
주우영, 유지은 (2024). Chat GPT를 활용한 대학 학생선수의 심리상담 사례연구, 스포츠심리상담연구, 1(1), 87-95.
조은이, 이진실, 최영원 (2025). 상담사 보조용 AI 개발을 위한 탐색적 연구: 상담사의 요구 및 우려와 기술적 대응에 대한 고찰. 상담심리교육복지, 12(1), 117-138.
최성호, 정정훈, 정상훈(2016). 질적 내용 분석의 개념과 절차, 질적탐구, 2(1), 127-155.
한국상담심리학회 (2018). 상담심리사 윤리강령. https://krcpa.or.kr/member/sub01_5.asp
한국상담심리학회 (2020). 매체상담 시 주의사항. https://krcpa.or.kr/bbs/list.asp?bid=bid_3
황매향 (2025), 인공지능(AI) 적용 상담 실제의 가능성 탐색, 교육논총, 45(3), 15-33.
American Psychological Association. (2014). *Ethical principles of psychologists and code of conduct.*
American Counseling Association. (2014). *ACA code of ethics. Author.*
American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct. Author.*
Corey, G. (2010). 심리상담과 치료의 이론과 실제 (8판). (조현춘, 조현재, 문지혜, 이근배, 홍영근 역). 서울: 시그마프레스.
Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions,* CA: Sage.
Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*(5th ed.), Sage publications.
Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process, *Journal of Advanced Nursing, 62*(1), 107-115.
Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).
Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial, *JMIR Mental Health, 4*(2), e19.
Garcia, E. (2021). UNESCO's Recommendation on the Ethics of AI: why it matters and what to expect from it. *URL: https://www. researchgate. net/publication/357074719_UNESCO%27s_Recommendation_on_the_Ethics_of_AI_why_it_matters_and_what_to_expect_from_it.*
Hsu, S. L., Shah, R. S., Senthil, P., Ashktorab, Z., Dugan, C., Geyer, W., & Yang, D. (2025). Helping the helper: Supporting peer counselors via ai-empowered practice and feedback. *Proceedings of the ACM on Human-Computer Interaction, 9*(2), 1-45.
Keung, W. M., & So, T. Y. (2025). Attitudes towards AI counseling: the existence of perceptual fear in affecting perceived chatbot support quality. *Frontiers in Psychology, 16,* 1538387.
Kitchener, K. S. (1984). Intuition, critical evaluation and ethical principles. *The Counseling Psychologist, 12,* 43-55.

Kuhail, M. A., Alturki, N., Thomas, J., & Alkhalifa, A. K. (2024). Human vs. AI counseling: College students' perspectives. *Computers in Human Behavior Reports, 16,* 100534.

Maurya, R. K. (2024). Using AI based chatbot ChatGPT for practicing counseling skills through role-play. *Journal of Creativity in Mental Health, 19*(4), 513-528.

Mittelstadt, B. D., et al. (2016). *The ethics of algorithms.* Big Data & Society, 3(2).

Patton, M. Q. (2002). Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative social work, 1*(3), 261-283.

Ping, Y. (2024). Experience in psychological counseling supported by artificial intelligence technology. *Technology and Health Care, 32*(6), 3871-3888.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21,* 95-103.

Rosen, C. (2025). *The extinction of experience: Being human in a disembodied world.* New York: St. Martin's Press.

United Nations Educational, Scientific and Cultural Organization. (2022). *Recommendation on the ethics of artificial intelligence.* https://unesdoc.unesco.org/ark:/48223/pf0000381378223/pf0000381137

You, F. T., Lin, P. C., Huang, C. L., Wu, J. H., Kabasawa, Y., Chen, C. C., & Huang, H. L. (2025). Artificial intelligence with counseling on the treatment outcomes and quality of life in periodontitis patients. *Journal of Periodontology, 96*(7), 781-793.

투고일자: 2026. 8. 29.
심사일자: 2026. 10. 5.
게재확정일자: 2026. 10. 9.

미래사회
Journal of Future Society
2026. Vol. 17, No. 3, 54～84.
https://doi.org/10.22987/jifso.2026.17.3.54

A Qualitative Content Analysis Study on Ethical Issues Experienced by Clients in AI-assisted Psychological Counseling

Jin Jung

Seoul Cyber University

This study aims to gain a comprehensive understanding of the ethical issues experienced by clients during AI(Artificial Intelligence) assisted psychological counseling. To this end, in-depth interviews were conducted with seven clients who had received psychological counseling from both AI and human counselors, and the data were analyzed using qualitative content analysis. The analysis identified 14 components and 28 sub-components. In conclusion, the ethical issues experienced by clients in AI-assisted counseling can be described as follows: "While easy accessibility allows for the convenient receipt of psychological help, clients perceive a requirement for ethical trust regarding aspects such as confidentiality and accountability." This study is expected to provide foundational data for establishing ethical standards and practical guidelines for the future use of AI in counseling settings by systematically exploring the various ethical issues clients encounter in AI-assisted counseling from the perspective of counseling ethics.

Key words: AI(Artificial Intelligence), Psychological Counseling, Kitchener, Counseling Ethics, Qualitative Content Analysis