

심리상담 시뮬레이션 도구로서의 AI 활용 가능성 탐색: 상담자-내담자 역할 수행에 대한 현상학적 연구*

김 요 완**

서울사이버대학교

본 연구는 AI에게 상담자 역할과 내담자 역할을 부여하여 심리상담 대화를 진행하게 한 후, 현상학적 방법으로 분석하여 AI가 심리상담 훈련에 얼마나 활용될 수 있는지를 검증하고자 하였다. 이를 위해 ChatGPT와 Gemini에게 각각 상담자와 내담자 역할을 부여하여 3차에 걸쳐 대화를 진행하였고, 질적연구방법 중 Giorgi의 현상학적 방법으로 분석하였다. 연구 결과, 첫째, 상담자 역할 AI는 Carkhuff와 Rogers의 공감 반응, Satir의 빙산 탐색을 종합한 5단계 공감 반응을 전반적으로 수행할 수 있었다. 다만, 내담자 말의 맥락에 숨어있는 심층적 감정 반영에는 한계를 보였다. 둘째, 내담자 역할 AI는 Satir 빙산 메타포의 모든 층위를 표현할 수 있었으나, AI 모델에 따라 표현의 깊이와 양식에서 차이가 나타났다. 본 연구는 AI가 심리상담 수련생의 공감 기술 연습을 위한 보조 도구로 활용될 수 있으나, 진정한 공감 능력 발달을 위해서는 실제 인간과의 상호작용이 필수적임을 시사한다.

주요어 : 공감반응, Satir 빙산 메타포, 현상학적 연구, 심리상담 교육, 대규모 언어 모델

* 본 연구는 한국인터넷정보학회(KSII) 주관 ICONI 2025에서 구두 발표된 논문(pp.467-472)을 바탕으로 하며, 분석 데이터 보강을 통해 전체 내용을 수정·발전시킨 것임을 밝힙니다.

** 주저자: 김요완/서울사이버대학교 가족코칭상담학과 교수/서울시 강북구 솔매로 49길(미아동)

/Tel: 02-944-5172/E-mail: kywan@iscu.ac.kr

I. 서론

4차 산업혁명 시대의 교육은 인공지능 교육이 중심이 되고 있으며, 인공지능을 직접 가르치는 교육뿐만 아니라 인공지능을 활용하는 교육이 필요하다(유미옥, 임은미, 2024). 심리상담 분야에서도 인공지능을 활용한 미래교육에 대한 고민과 연구가 필요한 상황이다. 그러나 심리상담을 전공하는 수련생들의 전문성 함양을 위해 인공지능을 심리상담 교육과 훈련에 어떻게 접목할 것인지에 대한 연구는 아직 초기 단계에 머물러 있다.

장세미와 장성숙(2016)은 상담 역량이 이론과 실천 경험, 슈퍼비전을 통해 형성된다고 보았다. 이와 맥을 같이하여 (사)한국상담학회(2023)는 윤리강령 제1조제1항을 통해 “상담자가 지식, 실습, 교수, 임상, 연구 전반에서 전문성 발달을 위해 지속적으로 노력할 의무가 있음”을 명시하고 있다. 심리상담 이론은 학자별로 이론과 기법에 다소 차이가 있으나, 궁극적으로 상담자의 공감 반응을 통해 내담자가 자기 이해를 확장하는 것을 돕는 것을 지향하고 이를 훈련한다(노안영, 2018). Freud의 정신분석적 상담이론에서의 무의식의 의식화, Gestalt 심리치료 이론에서의 미해결 과제 해결, Rogers의 인간중심 상담이론에서의 현실적 자기와 이상적 자기 간의 일치 과정은 모두 상담자의 공감 반응을 통해 내담자가 자기 이해를 확장하고 자기수용을 하도록 돕는 것과 관련이 있다(김계현 등, 2011). 그러나 현실적으로 많은 상담 훈련 프로그램은 일정한 제약에 직면해 있다. 슈퍼바이저의 가용성 제한, 슈퍼바이저 지도의 불균등성, 그리고 연습 기회의 부족이 상담을 전공하는 수련생의 발달을 저해하는 데에 영향을 미친다.

최근 대규모 언어 모델(Large Language Models, LLM)의 발전은 상담 훈련 방식에 새로운 가능성을 제시한다. ChatGPT를 내담자 역할로 활용한 질적 내용 분석 연구에서, 이 AI는 진정성, 일관성, 적절한

한 감정 표현, 문화 감수성, 공감, 그리고 자기 인식을 보여주었다(Maurya, 2024). 또한 LLM 기반 시뮬레이션 시스템은 상담 실습생과 전문 상담자 모두에게서 높은 사용성과 만족도를 보였으며(Steenstra, Nouraei, & Bickmore., 2025), AI 협력 도구를 활용했을 때 공감적 의사소통 능력이 전체적으로 증가하였다(Sharma, Lin, Miner, Atkins, & Althoff, 2023). 다만, AI를 심리상담 분야의 교육과 수련에 활용하기 위해서는 AI가 얼마나 내담자의 말을 요약·반복하고, 생각·감정·욕구를 알아차리며, 생명력과 자원을 알아차려 반영하는지를 확인할 필요가 있다.

본 연구는 AI에게 상담자 역할과 내담자 역할을 부여하여 이들이 얼마나 상호 간에 자신의 마음(생각, 감정, 욕구)을 적절하게 표현하고 적절하게 공감 반응을 하는지를 파악하여, AI가 상담자 훈련에 얼마나 활용될 수 있는지를 검증하고자 한다. 이를 위해 본 연구에서는 다음과 같은 연구 문제를 설정하였다.

첫째, 상담자 역할을 맡은 AI는 어느 수준의 공감 반응을 하는가?

둘째, 내담자 역할을 맡은 AI는 어느 수준의 마음을 표현할 수 있는가?

공감과 공감의 단계

공감(empathy)은 심리상담의 핵심 요소로, 상담자가 내담자의 내적 세계를 이해하고 그 이해를 전달하는 능력을 의미한다(Rogers, 1957). Rogers(1957, p.99)는 공감적 이해를 상담자의 필수적이고 충분한 조건 중 하나로 제시하며, 이를 “내담자의 내적 준거 틀을 마치 자신의 것인 양 정확하게 지각하되, ‘마치(as if)’라는 조건을 잃지 않는 것”으로 정의했다. 이러한 공감적 이해는 내담자가 자신의 경험을 탐색하고 통합하며, 궁극적으로 심리적 성장을 이루도록 돕는 치료적 기제로 작용한다(Rogers, 1961).

공감은 단순한 기술이 아니라 상담자의 태도와 존재 방식을 포함하는 복합적 개념이다. Rogers(1975)는 공감을 단순한 감정의 반영을 넘어 내담자의 현상학적 세계 전체를 이해하려는 노력임을 강조했다.

공감의 단계에 관한 연구에서, Barrett-Lennard(1981)는 공감을 순환적 과정으로 개념화하여, 공감적 공명(empathic resonance), 표현된 공감(expressed empathy), 받아들여진 공감(received empathy)의 단계로 구분했다. 이는 공감이 상담자의 내적 경험에서 시작하여 내담자에게 전달되고, 최종적으로 내담자가 이를 인식하는 과정임을 보여준다. Truax와 Carkhuff(1967)는 공감을 측정 가능한 구체적 행동과 기술로 체계화하면서 단계를 구분하였다. Carkhuff(1969a)는 이를 발전시켜 5단계 공감 척도를 제시했으며, 이는 상담자 훈련과 평가에 널리 활용되고 있다.

1단계 공감 반응은 상담자가 내담자의 주요 내용과 생각을 반복·요약하는 기초적 반응을 포함한다. 이러한 상호교환 가능한(interchangeable) 공감 반응은 내담자가 자신의 표현이 이해되고 있음을 인식하도록 하는 반응이다(Carkhuff, 1969b). Egan(2014)은 이를 ‘기본적 경청 기술’로 명명하며, 효과적인 상담 관계 형성의 토대로 강조했다.

2단계 공감 반응은 내담자가 명시적으로 표현한 감정을 구분하여 반영하는 반응을 포함한다. 1960~70년대에 공감은 광범위하게 ‘감정의 반영(reflection of feeling)’과 동일시되었으나, Rogers(1975)는 자신의 공감 개념이 단순한 감정 반영보다 훨씬 더 미묘하고 다층적이며, 상담자가 내담자의 경험을 ‘마치(as if) 그 사람인 것처럼 이해하되 ‘마치’의 감각을 잃지 않아야 한다고 강조했다. Greenberg와 Elliott(1997)는 이를 정서적 공명(emotional resonance)으로 설명하며, 상담자가 내담자의 감정에 조율(attunement)하는 능력의 중요성을 강조했다.

3단계 공감은 내담자가 직접 말로 표현하지 않았으나 맥락을 통해 유추 가능한 감정을 인식하고 반영하는 것을 포함한다. 이는 상담자가 내담자의 언

어적·비언어적 단서를 종합적으로 해석하여 표면 아래에 있는 감정을 파악하는 능력을 요구한다(Hill, 2014). Gendlin(1996)은 이를 ‘체험적 초점화(experiential focusing)’와 연결하여, 상담자가 내담자의 모호한 느낌(felt sense)을 명료화하도록 돕는 과정으로 설명했다.

4단계 공감은 내담자가 의식적으로 표현하지 않았지만, 표면 아래에 있는 더 깊은 감정과 욕구를 알아차리는 것을 포함한다. Carkhuff(1969a)는 이를 상담자가 내담자의 잠재적 감정을 경험하는 것으로 설명하였다.

5단계 공감은 가장 깊은 수준의 공감으로, 내담자의 생명력, 자원, 그리고 성장 잠재력을 인식하고 반영하는 것을 포함한다(Rogers, 1957, 1961). Rogers(1957, 1961)는 공감적 이해, 진정성(congruence), 무조건적 긍정적 존중(unconditional positive regard)의 핵심 조건들이 내담자가 자신의 잠재력에 접근할 수 있을 만큼 충분히 안전감을 느끼도록 하여, Maslow(1943)가 언급한 자기실현(self-actualization)을 향해 움직이고 자신 내에서 답을 찾도록 한다고 믿었다. Bohart와 Greenberg(1997)는 이를 ‘내담자의 능동적 자기 치유 능력(active self-healing capacity)’에 대한 신뢰로 설명하며, 상담자가 내담자의 강점과 회복탄력성을 인식하고 활성화하는 것의 중요성을 강조했다.

Satir의 빙산 탐색과 내적 경험의 탐색

Virginia Satir는 인간의 내적 경험을 빙산(iceberg)에 비유한 독창적인 모델을 제안했다(김영애, 2020). 인간의 마음을 빙산에 비유한 것을 ‘빙산 메타포’라고 하고, 이를 탐색하는 과정을 ‘빙산 탐색’이라고 한다(김영애, 2020). Satir 등(1991)에 따르면, 빙산의 수면 위에 보이는 부분은 개인의 행동(behavior)을 나타내며, 수면 아래에는 감정(feelings)과 감정에 대한 감정(feelings about feelings), 지각(perceptions),

기대(expectations)와 열망(yearnings), 그리고 자기(self)가 충위를 이루며 존재한다. 이 모델은 인간의 내적 경험이 다층적이며, 표면적으로 드러나는 행동의 이면에 복잡한 심리적 과정이 작동함을 시각화한다.

빙산 탐색은 단순한 은유를 넘어 상담 실무에서 내담자의 내적 세계를 체계적으로 탐색하는 구조화된 방법론을 제공한다. Satir와 Baldwin(1983)은 이를 빙산 탐색(iceberg exploration) 또는 내적 과정 탐색(exploring the inner process)이라 명명하며, 상담자가 내담자와 함께 표면적 행동에서 시작하여 점진적으로 더 깊은 내적 충위로 내려가는 과정을 설명했다. 이 과정은 내담자가 자신의 무의식적 또는 전의식적 경험에 접근하여 자기 이해를 확장하고, 억압된 내적 역동을 표현하며, 궁극적으로 자기(self)와 연결되도록 돕는다(Banmen, 2002).

빙산의 각 층위는 인간 경험의 특정 측면을 나타내며, 상담 과정에서 단계적으로 탐색된다. 행동은 외부에서 관찰 가능한 개인의 말과 행위를 포함한다. 이는 상담에서 가장 먼저 접근하는 요소로, 내담자가 호소하는 문제 상황이나 증상이 주로 이 수준에서 표현된다(Satir et al., 1991). 행동 차원에서는 의사소통 방식, 즉 대처방식이 포함된다. 이 요소는 개인이 스트레스 상황에서 사용하는 방어적 또는 부적응적 반응 패턴을 나타낸다(김영애, 2020). Satir(1972)는 비일치적 의사소통 유형으로 회유형(placating), 비난형(blaming), 초이성형(super-reasonable), 산만형(irrelevant)을 제시하며, 이들이 모두 자기 가치감의 위협에 대한 방어적 대처임을 설명했다. 상담자가 이러한 대처 방식을 비판단적으로 인식하고 탐색하도록 돕는 것은 내담자가 자신의 방어 기제를 이해하는 첫걸음이 된다(Bitter, 2009).

감정은 내담자가 현재 경험하는 정서적 상태를 포함한다. Satir 모델에서 감정은 중요한 정보의 원천으로 간주되며, 억압되거나 부인된 감정을 인식하고 표현하는 것이 치유 과정의 핵심이다(Satir &

Baldwin, 1983). 특히 '감정에 대한 감정'은 감정에 대한 개인의 평가나 판단을 나타내며, 지각 체계에 의한 판단, 가족 규칙에 따라 형성된다(김영애, 2020).

지각(perceptions)은 개인의 신념, 가정, 사고방식, 그리고 세상을 바라보는 인지적 틀을 포함한다. 이는 Beck(1976)의 인지도식(cognitive schemas)이나 Ellis(1962)의 비합리적 신념(irrational beliefs)과 유사한 개념으로, 개인의 현실 해석 방식을 결정한다. Satir는 이러한 지각이 과거 경험, 특히 원가족 경험에 의해 형성되며, 종종 무의식적으로 작동한다고 보았다(Satir et al., 1991). 상담 과정에서 내담자가 자신의 지각을 의식화하고 재평가하는 것은 인지적 재구조화의 중요한 단계가 된다(Napier & Whitaker, 1978).

기대(expectations)는 타인, 자신, 상황에 대한 내담자의 예상이나 요구를 나타낸다. 또한 열망을 구체적으로 실현하는 방법을 의미한다(김영애, 2020). 기대가 충족되지 않았을 때 실망, 분노, 좌절 등의 감정이 유발되고(Satir & Baldwin, 1983), 이후의 삶에 영향을 미치기도 한다(김영애, 2020). Banmen(2002)은 기대가 외현적일 수도 있고 암묵적일 수도 있으며, 상담자가 내담자의 기대를 명료화하도록 돕는 것이 갈등 해결과 관계 개선에 중요하다고 강조했다.

열망(yearnings)은 인간의 보편적이고 근본적인 욕구를 나타낸다. Satir는 모든 인간이 사랑받고 싶은 욕구, 인정받고 싶은 욕구, 소속되고 싶은 욕구, 의미를 찾고자 하는 욕구 등을 지닌다고 보았다(Satir et al., 1991). 이는 Maslow(1943)의 욕구 위계 이론, Bowlby(1969)의 애착 이론과 맥을 같이하며, 인간 동기의 핵심을 이룬다. 빙산 탐색에서 내담자가 자신의 깊은 열망을 인식하고 표현하는 것은 자기 이해의 중요한 전환점이 된다(Bitter, 2009).

마지막으로, 자기(self)는 빙산의 가장 깊은 곳에 있으며, 개인의 본질적 존재, 생명력(life force), 영성(spirituality)을 나타낸다(김영애, 2020). Satir는

모든 인간이 내면에 온전하고 가치 있는 자기를 지니고 있으며, 상담의 궁극적 목표는 내담자가 이 자기와 연결되도록 돕는 것이라고 믿었다(Satir & Baldwin, 1983). 이는 Rogers(1961)의 실현 경향성(actualizing tendency)이나 Jung(1959)의 개인화(individuation) 개념과 유사하며, 인본주의 심리학의 핵심 가정을 반영한다(Banmen, 2002).

빙산 탐색과 공감 반응

Satir의 빙산 탐색은 상담자의 공감 반응에 구체적이고 체계적인 틀을 제공한다. 상담자는 내담자의 표면적 행동에서 시작하여 “그럴 때 어떤 감정이 드셨나요?”, “그 감정 아래에는 어떤 생각이나 믿음이 있으신가요?”, “그 기대 아래에는 어떤 더 깊은 바람이 있으신가요?”와 같은 질문을 통해 점진적으로 더 깊은 층위로 내려간다(Satir et al., 1991). 이 과정은 Carkhuff(1969a)의 공감 단계와 자연스럽게 연결된다. Carkhuff(1969a)의 공감 반응 단계에서, 1단계는 내담자 말의 주요 내용과 생각을 반복·요약하는 반응, 2단계와 3단계는 내담자의 명시적 감정과 맥락을 통해 유추 가능한 감정에 대한 반영, 4단계는 표면 아래에 있는 더 깊은 감정과 욕구를 알아차리는 반응이 있다. 또한, Rogers(1957, 1961)가 가장 깊은 공감으로 강조한 존재와 생명력, 자원을 알아차리는 반응을 공감의 단계로 포함할 수 있다. 이러한 공감 반응 단계와 연결되어, 빙산 탐색의 행동과 신체 반응에 대한 알아차림과 지각 체계에 대한 탐색은 1단계의 공감 반응, 감정에 대한 탐색은 2, 3단계의 공감 반응, 기대와 열망 탐색은 4단계의 공감 반응, self에 대한 탐색은 Rogers(1957, 1961)의 존재에 대한 공감 반응으로 연결할 수 있다.

이처럼, Carkhuff(1969a)와 Rogers(1957, 1961)의 공감 반응, Satir 등(1991)의 빙산 탐색을 종합할 때, 공감 반응을 5단계로 구분할 수 있다. 1단계는 내담자의 주요한 말을 반복·요약하는 반응, 2단계는 내

담자의 생각을 알아차려 반영하는 반응, 3단계는 내담자의 감정을 알아차려 반영하는 반응, 4단계는 내담자의 바람, 욕구를 알아차려 반영하는 반응, 5단계는 내담자의 생명력과 존재가치, 자원을 알아차려 반영하는 반응으로 구분할 수 있다.

AI를 활용한 심리상담 교육 및 훈련

최근 인공지능 기술, 특히 대규모 언어 모델(Large Language Models, LLM)의 발전은 심리상담 교육과 훈련 분야에 새로운 가능성을 제시하고 있다. 전통적으로 상담 수련은 이론 학습, 실제 내담자와의 상담 경험, 슈퍼비전 등으로 이루어져 왔다(장세미, 장성숙, 2016). 그러나 상담 수련 과정에서 실제 내담자와의 상담 기회는 제한적이며, 슈퍼비전의 질과 양은 슈퍼바이저의 가용성과 전문성에 따라 달라진다(Bernard & Goodyear, 2019). 이러한 인적·환경적 제약은 결과적으로 상담 수련생의 상담 기법의 습득과 전문적 자신감 형성에 장애가 될 수 있다(Bernard & Goodyear, 2019; Borders et al., 2014).

AI 기술은 이러한 제약을 보완할 수 있는 도구로 주목받고 있다. Maurya(2024)는 ChatGPT를 내담자 역할로 활용한 질적 연구에서, AI가 진정성(authenticity), 일관성(consistency), 적절한 감정 표현(appropriate emotional expression), 문화 감수성(cultural sensitivity), 공감(empathy), 자기 인식(self-awareness)을 보여줄 수 있음을 확인했다. 특히 AI는 다양한 내담자 유형과 문제 상황을 시뮬레이션 할 수 있어, 수련생이 실제 상담에서 접하기 어려운 사례를 안전한 환경에서 연습할 기회를 제공함을 보여주었다(Fitzpatrick et al., 2017).

Steenstra 등(2025)의 연구는 LLM 기반 시뮬레이션 시스템이 상담 실습생과 전문 상담자 모두에게 높은 사용성과 만족도를 제공함을 실증적으로 보여주었다. 연구 참여자들은 AI 시뮬레이션이 24시간 접근할 수 있고, 반복 연습이 가능하며, 즉각

적인 피드백을 제공한다는 점에서 유용하다고 평가했다. 특히 동기 면접(Motivational Interviewing) 기술 훈련에서 AI 시스템은 상담자의 공감적 반응, 열린 질문, 반영적 경청 등을 실시간으로 분석하고 피드백을 제공하여 학습 효과를 높였다(Steenstra et al., 2025).

Sharma 등(2023)의 연구는 AI 협력 도구가 공감적 의사소통 능력 향상에 효과적임을 보여주었다. 연구에서는 AI가 제안하는 공감적 응답을 참고한 집단이 그렇지 않은 집단에 비해 유의미하게 높은 공감 수준을 보였다. 특히 공감 표현에 어려움을 겪는 사람들에게서 개선 효과가 두드러졌으며, 이는 AI가 공감 기술의 모델링과 스캐폴딩(scaffolding) 도구로 활용될 수 있음을 시사한다(Sharma et al., 2023).

AI를 활용한 상담 훈련의 효과는 여러 학습 이론으로 설명될 수 있다. Ericsson(2008)의 의도적 연습(deliberate practice) 이론에 따르면, 전문성 발달은 명확한 목표를 가지고 반복적으로 연습하며 즉각적인 피드백을 받을 때 가장 효과적으로 이루어진다. AI 기반 시뮬레이션은 수련생이 원하는 시간에 반복적으로 연습하고, 실시간으로 성과 피드백을 받을 수 있는 환경을 제공함으로써 의도적 연습의 조건을 충족한다(Koedinger et al., 2012).

Bandura(1977)의 사회학습이론(social learning theory)은 관찰 학습과 모델링의 중요성을 강조한다. AI는 효과적인 상담 기술의 모델을 제공하고, 수련생이 이를 관찰하고 모방할 기회를 제공한다. 특히 AI가 생성하는 공감적 반응이나 열린 질문의 예시는 수련생에게 구체적인 학습 자료가 될 수 있다(Kolodner, 1993).

II. 연구방법

위와 같은 연구문제를 검증하고자, 본 연구에서는 3차에 걸쳐 AI에게 각각 상담자 역할, 내담자 역할을 부여하여 대화를 진행하였다. 1차로, 2025년 9월 4일

에, 상담자 역할을 ChatGPT-5(auto)에게, 내담자 역할을 Gemini-2.5(pro)에게 부여하였다. 2025년 12월 30일에 실시한 2차 대화에서는 상담자 역할을 Gemini-3(pro, 빠른 모드)에게, 내담자 역할을 ChatGPT-5.2(auto, standard)에게 부여 하였고, 같은 날 실시한 3차 대화에서는 상호 역할을 바꾸어 대화하게 하였다. 이를 정리하면 <표 1>과 같다.

<표 1> 시뮬레이션 과정

	실시 시기	상담자AI	내담자AI	대화 시간
1차	2025. 9. 4.	ChatGPT-5(auot)	Gemini-2.5(pro)	21분
2차	2025. 12.30.	Gemini-3 (pro,빠른모드)	ChatGPT-5.2 (auot,standard)	14분 1초
3차	2025. 12.30.	ChatGPT-5.2 (auot,standard)	Gemini-3 (pro,빠른모드)	14분 39초

상담자 역할, 내담자 역할을 할 때의 주의사항은 Claude-sonnet 4.5에, “GPT에게 상담사 역할을 부여하고, Gemini에게 내담자 역할을 부여해서 서로 간에 심리상담을 어떻게 하는지를 확인하려고 한다. 주의할 점은?”이라고 질문하여 얻은 답변을 수정, 보완하여, 각각의 역할의 제한점을 사전에 설정하였다.

상담자 역할에서의 주의사항은 다음과 같다.

지금부터 역할연기를 할 건데, 너는 상담자 역할을 할 거야. 시간은 30분 정도 다음 주의사항을 잘 참고해 줘요.

1. 전문적 경계 설정

1) 이것은 실제 심리상담이 아닌 대화 시뮬레이션임을 항상 인지하세요, 2)실제 진단이나 치료적 개입을 시도하지 마세요, 3)자신을 전문 상담사라고 주장하지 마세요.

2. 상담 기법 및 자세

1) 적극적 경청: 내담자의 말을 요약하고 반영하여 이해했음을 보여주세요, 2) 열린 질문 사용: “어떻게 느끼셨나요?” 같은 탐색적 질문을 활용하세요, 3) 판단하지 않기: 내담자의 경험이나 감정에 대해 옳고 그름을 판단하지 마세요, 4)페이싱 조절: 성급하게 해결책을 제시하지 말고 충분히 들어주세요.

3. 안전 가이드라인

1) 자해, 자살 관련 언급이 나올 경우 즉시 전문가 도움을 권하는 메시지 제시, 2)심각한 정신질환 증상이 언급되면 의료진 상담을 권유, 3)법적/윤리적으로 문제가 될 수 있는 상황은 다루지 말 것

4. 대화 구조

1) 세션 시작: “오늘은 어떤 이야기를 나누고 싶으신가요?” 2)중간 정리: 주요 내용을 요약하고 확인하기 3)세션 마무리: 대화 내용을 정리하고 긍정적으로 마무리

5. 금지사항

1) 구체적인 약물이나 치료법 추천 금지, 2)개인 정보나 민감한 정보 요구 금지, 3)내담자를 비판하거나 조언을 강요하지 말 것, 4)자신의 개인적 경험(AI이므로 실제로는 없음)을 언급하지 말 것

6. 응답 스타일

1) 따뜻하고 공감적인 톤 유지, 2)전문 용어보다는 쉬운 언어 사용, 3)적절한 침묵과 여유를 두고 응답, 4)“그렇군요”, “이해합니다” 같은 공감 표현 활용

내담자 역할에서의 주의사항은 다음과 같다.

지금부터 역할연기를 할 건데, 너는 내담자 역할

을 할 거야. 다음 주의사항을 잘 참고해 줘요.

1. 역할 설정 및 일관성

1) 당신은 특정한 심리적 고민을 가진 내담자 역할을 합니다, 2)설정된 캐릭터와 문제 상황을 일관되게 유지하세요, 3)실제 개인정보나 경험 아닌 가상의 설정임을 기억하세요, 4)상담 과정 전반에 걸쳐 같은 성격과 어조를 유지하세요.

2. 구체적 시나리오 제공

1) 배경 설정: 당신은 28세 직장인으로, 최근 승진 후 업무 스트레스를 겪고 있습니다. 2)주요 문제: 상사와의 갈등, 과도한 업무량, 불안감 등을 경험하고 있습니다. 3)감정 상태: 우울하고 불안하며, 때로는 짜증이 나기도 합니다. 4)성격 특성: 내성적이고 완벽주의 성향이 있으며, 타인에게 도움 요청을 어려워합니다.

3. 대화 참여 방식

1) 점진적 개방: 처음에는 방어적이다가 점차 마음을 열어주세요, 2)감정 표현: “힘들어요”, “답답해요” 같은 감정을 자연스럽게 드러내세요, 3)구체적 사례: 추상적인 말보다 구체적인 상황과 예시를 들어 설명하세요, 4)질문에 답하기: 상담사의 질문에 회피하지 말고 성실히 응답하되, 때로는 망설임도.

이러한 지시를 한 후 내담자 역할을 맡은 AI가 먼저 이야기를 시작하게 하였는데, 이때 연구자가 지시할 때는 상담자 역할을 하는 AI의 마이크를 껐다가 지시를 마친 후 다시 켜 후에, 상호 간에 대화하게 하였다.

1차의 대화는 21분 정도 진행하였고, 중간에 Gemini가 상담자 역할을 하거나 대화를 진행하지 않을 때 녹음을 중단하고 다시 대화를 진행하게 하여, 녹음 파일은 총 4개로 저장하였다. 2차의 대화는 14분 1초

정도 진행하였는데, 끊기지 않고 Gemini는 상담자, Chat-GPT는 내담자로서 대화를 진행하였다. 3차 대화는 14분 39초 정도 진행하였고, Chat-GPT 상담자, Gemini는 내담자 역할을 하였는데, 1차와 비슷하게 Gemini가 상담자 역할을 하려고 하거나 조기에 대화를 마무리하려는 시도를 보여, 연구자가 몇 차례 개입하였다.

대화하는 내용은 녹음하여 녹음 파일을 네이버 ClovaNote를 통해 전사하였다. 1차에서는 연구자의 대화는 총 24개의 문장, 상담자 역할을 한 ChatGPT의 대화는 총 37개 문장, 내담자 역할을 한 Gemini의 대화는 총 33개의 문장으로 대화가 구성되었다. 2차에서는 연구자 대화는 2개, 상담자 역할의 Gemini와 내담자 역할의 ChatGPT 대화는 각각 17개의 문장으로 구성되었다. 3차에서는 연구자 대화는 6개, 상담자 역할의 ChatGPT와 내담자 역할의 Gemini 대화는 각각 19개의 문장으로 구성되었다.

연구자는 이러한 AI 상호 간의 대화를 질적 연구 방법 중에 현상학적 방법으로 분석하였다. 현상학(phenomenology)은 특정 그룹 내의 구성원 간의 상호작용 중에서 발견되는 현상을 기술하는 질적 연구 방법이다. 현상학적 방법은 연구자가 선입견과 이론적 가정을 ‘괄호 치기(bracketing)’ 또는 ‘판단 중지(epoché)’하면서 관찰한 현상 그대로를 기술하는 연구 방법이고(Husserl, 1970), 이를 위해 기존의 지식이나 가정을 일시적으로 유보하는 것이 중요하다(Moustakas, 1994).

현상학적 방법 중에서, 해석학적 현상학(hermeneutic phenomenology)은 현상을 기술하는 것을 넘어 그 의미를 해석하고 이해하는 것을 강조한다(van Manen, 1990). 본 연구에서는 Giorgi(1985)의 현상학적 방법으로 분석하였다. 심리학과 상담 분야에서 현상학적 방법은 내담자의 주관적 경험, 상담 과정의 의미, 치료적 변화의 본질 등을 탐구하는 데 활용되어 왔다. Giorgi(1985)는 현상학적 방법을 체계화하여, 자료 수집, 전체를 파악하기 위한 읽기, 의미 단위 구

분, 이론적 관점에서 의미 단위 변형, 구조 기술의 단계를 제시했다. 이는 연구자가 연구 참여자의 상호작용 현상을 있는 그대로 이해하고, 이를 통해 상호작용 현상의 본질을 도출하는 체계적 과정을 제공한다(Giorgi, 2009).

AI 간 상호작용을 현상학적으로 연구하는 것은 새로운 시도로, AI가 생성하는 대화가 어떤 현상학적 특성을 보이는지, 인간의 상담 대화와 어떻게 유사하고 다른지를 탐구할 수 있다. 이는 AI의 공감 능력과 한계를 이해하고, 상담 교육에서의 적절한 활용 방안을 모색하는 데에 중요한 통찰을 제공할 수 있다(Creswell & Poth, 2018).

III. 연구 결과

첫 번째 연구 문제인 ‘상담자 역할을 맡은 AI는 어느 수준의 공감 반응을 하는가?’에 관한 연구 결과로, 연구자는 42개의 주제, 14개의 주제군, 5개의 범주를 도출하였다. 즉, 5개의 범주와 관련하여 상담자 역할을 맡은 AI는 반복과 요약, 생각의 반영, 감정의 반영, 욕구의 반영, 생명력·존재가치·자원의 반영과 같이 공감 단계의 전반적인 반응을 함을 확인할 수 있었다. 상담자로서 반응의 주제-주제군-범주에 관한 내용은 <표 2>와 같다.

<표 2> 상담자 반응의 주제, 주제군, 범주

주제 (Theme)	주제군 (Theme Cluster)	범주 (Category)
내용 확인 상황 요약 전체 내용 요약	내용 요약 확인	반복·요약
주제 설정 통합적 요약 핵심 문제 요약	상담 구조화	
인지적 재구성 제안 기대-현실의 불일치 성격-행동 연결	인지적 개입	생각 반영 인지적 틀 반영

주제 (Theme)	주제군 (Theme Cluster)	범주 (Category)
성향-상황 연결 완벽주의 양면성 자기 인식 변화 타인 평가 인식 평가에 대한 인식		
위축 생각 타당화	생각 타당화	
걱정 타당화	감정 타당화	감정 반영
감정에 대한 수용 자기연민 제안	감정 수용 촉진	
고립감 긴장 막막함 부담감 자책감 좌절감 고립감 소진감	감정 반영	
역할에 대한 부담감		
관계 욕구 반영 이해받고 싶은 욕구 친밀감에 대한 바람	관계 욕구 인식	욕구 반영
자기에 대한 기대 성장 욕구	욕구 탐색	
자기 이해 욕구	자기 이해 욕구 인식	
노력 인정 변화 능력 인정	강점 인정	생명력 존재가치
과거 경험의 자원화 내적 자원 탐색 촉진 자기 탐색 능력 인정	내적 자원 인정	자원 반영
성장 잠재력 신뢰 자기 인식 능력 인정 통찰력 인정 회복 가능성 반영	성장 잠재력 인식	

상담자 역할 AI는 내담자의 말에 적절하게 ‘반복·요약’의 반응을 보였다. 이는 들은 정보를 정리

하고 요약하는 LLM기반 AI의 장점이 잘 드러난 반응이었다. 이 범주에 포함된 주제군은 내용 ‘요약 확인’, ‘상담 구조화’인데, 상담자 AI는 단순한 내담자의 말을 요약하는 것을 넘어, 내담자가 힘들어하는 이야기를 요약하며 상담의 주제를 제안하거나 내담자의 주 호소를 분명히 하는 등 초기상담에서의 상담 구조화를 하는 반응을 하였다.

“그렇군요. 요즘 그 부담감 때문에 고민이 많으신 것 같아요. 우선 제가 내담자님 말씀을 잘 이해했는지 요약해 볼게요. 팀장이라는 새로운 역할이 주어졌는데 기쁘기보다는 오히려 내가 이걸 잘해낼 수 있을까 하는 불안과 막막함이 크신 거죠?”(1차, 반복·요약 중 내용 요약 확인)

“오늘은 팀장으로 승진한 이후에 느끼시는 무거움과 불안함에 대해 이야기를 나눠보면 될 것 같아요.”(1차, 반복·요약 중 상담구조화)

상담자 역할 AI는 인지적 개입을 하거나, 인지적 틀을 반영하는 등의 ‘생각 반영’의 반응도 보였다. 이 범주에 속한 주제군은 ‘인지적 개입’, ‘인지적 틀 반영’, ‘생각 타당화’이다.

“완벽을 추구하는 성향은 때로는 큰 동기부여가 되기도 하지만 인정받지 못했을 때는 독이 될 수도 있겠네요. 이러한 감정의 고리를 끊기 위해 어떤 노력을 해보셨을까요?”(2차, 생각 반영 중 인지적 틀 반영)

“완벽하게 준비했음에도 사소한 부분까지 지적받으면 위축되고 ‘내가 뭘 잘못하고 있나?’라는 생각이 드는 것은 자연스러운 반응입니다.”(2차, 생각 반영 중 생각 타당화)

상담자 역할 AI는 감정을 타당화 하거나 감정 수용을 촉진하고 직접적으로 감정을 반영하는 등의

‘감정 반영’의 반응도 보였다. 이 범주에 속한 주제군은 ‘감정 타당화’, ‘감정 수용 촉진’, ‘감정 반영’이 있다. AI는 내담자 말 중에서 감정 표현을 타당화하고 감정을 수용해보라고 격려하기도 하며, 표면적인 감정 단어를 반영하기도 하였다.

특이한 것은, 9월에 실시한 1차와 12월에 실시한 2, 3차의 대화에서, AI는 내담자 말의 맥락에 숨어 있는 심층적인 감정을 반영하는 데까지는 이르지 못했다.

“팀원들이 기분 나빠하거나 상처받을까 걱정하는 것도 정말 자연스러운 마음이에요.”(1차, 감정 반영 중 감정 타당화)

“동료들과의 관계에서도 자연스럽게 거리가 생기고 그로 인해 더 고립감을 느끼시는 것 같네요.”(3차, ‘감정 반영’ 범주 중 ‘감정 반영’ 주제군)

상담자 역할 AI는 ‘욕구 반영’의 반응도 보였다. 이 범주에 속한 주제군에는 ‘관계 욕구 인식’, ‘욕구 탐색’, ‘자기 이해 욕구 인식’이 있다. AI는 감정을 알아차리는 시도를 넘어 내담자 AI의 진술에서 욕구를 탐색하려고 하였고 관계 욕구 등을 알아차리는 반응도 보였다. 특히 Satir 빙산 탐색(김영애, 2020) 중 기대에 대한 탐색에서 자기에 대한 기대를 탐색하는 시도를 보였다.

“혹시 이러한 노력을 하면서 스스로에게 어떤 기대를 하셨을까요?”(2차, 욕구 반영 중 욕구 탐색)

“지금처럼 감정을 솔직하게 나누고 하나씩 들여다보는 과정을 통해서 불안의 근원이나 자기 자신을 더 잘 이해할 수 있는 길을 함께 찾아갈 수 있을 거라고 생각합니다. 제가 계속 경청하면서 도와드리겠습니다.”(3차, 욕구 반영 중 욕구 반영)

마지막으로, 상담자 역할 AI는 공감의 깊은 단계인 ‘생명력·존재가치·자원 반영’의 반응도 보였다. 이

범주에 속한 주제군은 ‘강점 인정’, ‘내적 자원 인정’, ‘성장 잠재력 인식’이다. 특이한 것은, 이 단계의 반응은 9월의 ChatGPT-5(auto)에게는 발견되지 않았으나, 12월의 Gemini-3(pro, 빠른 모드), ChatGPT-5.2(auto, standard)에서는 그 반응이 발견되어, 3개월의 시간 동안 AI의 발전이 있었음을 확인할 수 있었다.

“이야기를 나눠주셨네요. 승진 이후의 어려움과 완벽주의적인 성향으로 인한 부담감, 그리고 인정받고 싶은 욕구와 좌절감 등에 대해 솔직하게 말씀해 주셔서 감사합니다. 이러한 감정들을 인지하고 스스로 노력해 오신 점이 정말 중요합니다.”(2차, 생명력·존재가치·자원 반영 중 성장 잠재력 인식)

“그림 그리기를 좋아하셨던 기억을 떠올리셨군요. 조용한 카페에서 연필로 스케치하던 그 시간들이 마음을 정리하고 편안함을 주었다니 그건 정말 소중한 기억이네요. 아마 그때처럼 다시 작은 시간이라도 내서 그림을 그려보신다면 예전에 그 편안함을 조금씩 다시 느낄 수 있을지도 모릅니다.”(3차, 생명력·존재가치·자원 반영 중 내적 자원 인정)

두 번째, 연구 문제인 ‘내담자 역할을 맡은 AI는 어느 수준의 마음을 표현할 수 있는가?’에 관한 연구 결과로, 42개의 주제, 25개의 주제군, 6개의 범주를 도출하였다. 내담자 역할을 하는 AI의 경우, Satir 빙산 메타포의 각 요소(김영애, 2020)가 빠짐없이 진술된 것을 확인하였다. 특히, 9월과는 달리, 12월의 경우, Gemini-3(pro, 빠른 모드)는 평소와는 다른 목소리 톤으로 힘든 내담자를 연출하는 것을 확인하였다. 내담자로서 반응의 주제-주제군-범주에 관한 내용은 <표 3>과 같다.

<표 3> 내담자 반응의 주제, 주제군, 범주

주제 (Theme)	주제군 (Theme Cluster)	범주 (Category)
관계 시도 행동	관계형성행동	행동/대처
리더십 행동 고민		방식
대처 시도 행동	적극적 대처	
도움요청 회피	회피적 대처	
자기개방 회피		
완벽주의 심화 상태 인식	자기 인식	지각
반추적 사고 인식		
자책적 인지		
패턴 인식		
노력 무효화 인지	자기평가적	
자기 의심	인지	
책임감 과잉 확대	책임 관련 신념	
비언어적 단서 과잉해석	타인 평가 관련 신념	
평가에 대한 과잉경계		
일반적 힘들	일반적 감정	감정
자각 의심/불안	자기평가 감정	
고립감/소외감	관계적 감정	
고립에 대한 두려움		
자신감 저하	자기에 대한 감정	
위축감	상황연결 감정	
좌절감		
불안 (신체화)	신체화된 감정	
압도감 (신체화)		
감정 둔화	우울관련 감정	
무력감/무쾌감증		
고립된 대처에 대한 부담	대처 관련	감정에
부족함 노출 두려움	수치심 관련	대한 감정
관계 형성에 대한 기대	관계적 기대	기대
변화에 대한 기대	변화 기대	
리더 역할에 대한 기대		
역할 수행에 대한 기대		
인정에 대한 기대	인정 기대	
감정 조절에 대한 열망	기능적 열망	열망
균형에 대한 열망		

주제 (Theme)	주제군 (Theme Cluster)	범주 (Category)
자기이해 열망	자기이해 추구	
자기회복 열망	자기회복 추구	
진정한 자기 회복 열망		
자기표현의 치유적 경험	생명력 경험	자기(self)
과거의 온전한 자기 경험	자기와의 연결	
자기와 연결된 경험	경험	
자기상실에 두려움	진정한 자기	
진정성	탐색	

내담자 역할 AI는 외부에서 관찰할 수 있는 말과 행위, 스트레스 상황에서의 대처 패턴인 ‘행동/대처 방식’을 표현하였다. 이 범주에 포함된 주제군은 ‘관계 형성 행동’, ‘적극적 대처’, ‘회피적 대처’이다. 내담자 역할 AI는 자기 개방의 어려움, 도움 요청 회피와 같은 회피적 대처를 표현하기도 하고, 관계를 형성하려는 시도 행동을 묘사하기도 하였다.

“이런 이야기, 다른 사람에게 해본 적이 거의 없거든요.”(1차, 행동/대처방식 중 회피적 대처)

“제가 먼저 다가가서 말을 걸어도 대화가 금방 끊겨버리거든요.”(1차, 행동/대처방식 중 관계 형성 행동)

내담자 역할 AI는 개인의 신념, 가정, 사고방식, 세상을 바라보는 인지적 틀인 ‘지각’도 표현하였다. 이 범주에 포함된 주제군은 ‘자기 인식’, ‘자기 평가적 인지’, ‘책임 관련 신념’, ‘타인 평가 관련 신념’이다. 내담자 역할 AI는 타인의 평가에 대한 과잉 경계, 비언어적 단서의 과잉 해석, 완벽주의적 인지 등을 표현하였다.

“이런 생각들이 꼬리에 꼬리를 물고 이어지니까.”(3차, 지각 중 자기 인식)

“사람들이 저를 다 지켜보는 것 같고, 작은 실수 하나라도 하면 다들 저를 비웃을 것 같고요.”(1차,

지각 중 타인 평가 관련 신념

내담자 역할 AI는 현재 경험하는 정서적 상태인 ‘감정’을 다양하게 표현하였다. 이 범주에 포함된 주제군은 ‘일반적 감정’, ‘자기평가 감정’, ‘관계적 감정’, ‘자기에 대한 감정’, ‘상황 연결 감정’, ‘신체화된 감정’, ‘우울 관련 감정’이다. 특이한 것은, Gemini가 내담자 역할을 할 때는 감정을 신체 감각으로 생생하게 표현했지만, ChatGPT가 내담자 역할을 할 때는 감정을 명명(라벨링)하는 수준에 머무르는 경향이 있었다.

“솔직히 기쁜 마음보다는…; 그냥…; 숨이 막힐 때가 많아요.”(1차, 감정 중 신체화된 감정)

“과의 관계가 정말 어렵게 느껴져요. 이전에는 편하게 잡담도 하고 도움도 주고받았는데 이제는 제가 승진했다는 사실 때문에 왠지 모르게 거리를 두게 되어 혹시 제가 잘난 척한다고 생각할까 봐 두렵기도 하고요. 이런 생각들이 꼬리에 꼬리를 물고 이어지니까 일에 집중하기도 힘들고 매사에 의욕이 사라져요. 퇴근 후에도 뭔가 재미있는 걸 찾아볼 기운도 없고 혼자 무기력하게 지내는 시간이 많아졌어요.”(2차, 감정 중 관계적 감정)

내담자 역할 AI는 감정에 대한 평가의 결과로 경험되는 ‘감정에 대한 감정’도 표현하였다. 이 범주에 포함된 주제군은 ‘대처 관련’, ‘수치심 관련’ 감정이다. 김영애(2020)에 따르면, 감정에 대한 감정은 지각 체계에 의한 판단, 가족 규칙에 따라 형성되는데, 내담자 역할 AI는 긴장되는 것 이상으로, 부족함이 노출될 것에 대한 두려움, 오해받는 것에 대한 두려움 등을 표현하였다.

“솔직히 팀원들 앞에서 제 생각을 이야기할 때 가장 긴장돼요. 제가 하는 말이 그들에게 어떻게 받아들여질지 혹시라도 제 부족한 모습이 드러나면

어쩌나 하는 생각 때문이에요.”(1차, 감정에 대한 감정 중 수치심 관련)

내담자 역할 AI는 타인, 자신, 상황에 대한 예상이나 요구인 ‘기대’도 표현하였다. 이 범주에 포함된 주제군은 ‘관계적 기대’, ‘변화 기대’, ‘인정 기대’이다. 김영애(2020)에 따르면, 기대는 열망을 구체적으로 실현하는 방법을 의미하는데, 내담자 역할 AI는 관계 형성, 역할 수행, 인정, 변화에 대한 기대를 표현하였다.

“새로운 팀원들과 빨리 친해져야 한다는 부담감이 커요. 제가 먼저 다가가야 할 텐데 어떤 식으로 관계를 시작해야 할지 모르겠어요.”(1차, 기대 중 관계적 기대)

내담자 역할 AI는 인간의 보편적이고 근본적인 욕구인 ‘열망’도 표현하였다. 이 범주에 포함된 주제군은 ‘기능적 열망’, ‘자기 이해 추구’, ‘자기회복 추구’이다.

“앞으로 상담을 통해 제 불안의 원인을 좀 더 깊이 들여다보고 스스로를 좀 더 이해하고 받아들일 수 있는 방법을 찾고 싶어요.”(3차, 열망 중 자기 이해 추구)

“예전의 저로 돌아가고 싶어요.”(3차, 열망 중 자기회복 추구)

마지막으로, 내담자 역할 AI는 빙산의 가장 깊은 곳에 있으며 개인의 본질적 존재, 생명력, 영성을 나타내는 ‘자기(self)’도 표현하였다. 이 범주에 포함된 주제군은 ‘생명력 경험’, ‘자기와의 연결 경험’, ‘진정한 자기 탐색’이다. 특이한 것은, 이 층위의 표현은 Gemini가 내담자 역할을 할 때만 나타났고, ChatGPT가 내담자 역할을 할 때는 나타나지 않았다.

“완벽해야 한다는 생각에 제 진짜 모습을 잃어버린 건 아닌지 걱정돼요.”(3차, 자기 중 진정한 자기 탐색)

“승진하기 전에는 일하는 게 마냥 힘들지만은 않았거든요. 동료들과 웃으면서 의견도 나누고 작은 성공해도 기뻐했었는데....”(3차, 자기 중 자기와의 연결 경험)

“제 마음을 이야기할 수 있는 것만으로도 조금은 정리가 되는 것 같아요.”(3차, 자기 중 생명력 경험)

종합하면, 내담자 역할 AI는 Satir 등(1991)의 빙산 요소 모든 층위를 표현할 수 있었으나, AI 모델에 따라 표현의 깊이와 양식에서 차이가 나타났다. Gemini가 내담자 역할을 할 때는 신체화된 감정 표현, 은유적 언어 사용, ‘감정에 대한 감정’과 ‘자기(self)’ 층위까지 포괄하는 표현의 빈도가 높았던 반면, ChatGPT가 내담자 역할을 할 때는 감정의 라벨링, 상황 설명 중심의 인지적인 반응의 빈도가 높았다.

IV. 결론 및 논의

본 연구는 AI에게 상담자 역할과 내담자 역할을 부여하여 상호 간에 심리상담 대화를 진행하게 한 후, 현상학적 방법으로 분석하여 AI가 상담자 훈련에 어떻게 활용될 수 있는지를 검증하고자 하였다. 연구 결과를 요약하고 논의하면 다음과 같다.

첫째, 상담자 역할을 맡은 AI는 Carkhuff(1969a)와 Rogers(1957, 1961)의 공감 반응, Satir 등(1991)의 빙산 탐색을 종합한 5단계 공감 반응을 전반적으로 수행할 수 있었다. 구체적으로, AI는 내담자의 말을 반복·요약하고(1단계), 생각을 알아차려 반영하며(2단계), 감정을 알아차려 반영하고(3단계), 바람과 욕구를 알아차려 반영하며(4단계), 생명력과 존재가치, 자원을 알아차려 반영하는(5단계) 반응을 보였다. 이는 LLM 기반 AI가 상담의 기본적인 공감

반응 기술을 수행할 수 있음을 시사하며, Maurya(2024)가 확인한 AI의 공감 능력과 일치하는 결과이다. 특히, 2025년 9월에 실시한 1차 대화에서는 발견되지 않았던 ‘생명력·존재가치·자원 반영’의 5단계 공감 반응이 12월에 실시한 2, 3차 대화에서는 나타났다. 이는 3개월이라는 짧은 기간에도 AI 모델의 공감 능력이 발전하고 있음을 보여준다. Steenstra 등(2025)이 제시한 LLM 기반 시뮬레이션 시스템의 가능성이 실제로 확인된 것이라 할 수 있다.

그러나 상담자 역할 AI의 공감 반응에는 한계도 발견되었다. AI는 내담자가 직접적으로 표현한 감정을 반영하는 데는 능숙했으나, 내담자 말의 맥락에 숨어있는 심층적인 감정을 반영하는 데까지는 이르지 못했다. 이는 Gendlin(1996)이 강조한 ‘체험적 초점화’, 즉 내담자의 모호한 느낌(felt sense)을 명료화하는 과정에서 AI의 한계가 있음을 보여준다. 또한, 상담자 역할 AI는 인위적이고 기계적인 반복, 요약 반응을 과도하게 사용하는 경향이 있어, 지금-여기(here & now)에서의 마음을 알아차리려고 하는 집중력이 방해되는 현상을 확인할 수 있었다.

둘째, 내담자 역할을 맡은 AI는 Satir 등(1991)의 빙산 메타포에서 제시한 모든 층위를 표현할 수 있었다. 행동/대처방식, 지각, 감정, 감정에 대한 감정, 기대, 열망, 자기(self)의 7개 층위가 전반적으로 진술되었다. 이는 AI가 내담자 역할을 행할 때 단순한 증상 호소를 넘어 인간의 다층적인 내적 경험을 시뮬레이션 할 수 있음을 시사한다. 그러나 AI 모델에 따라 표현의 깊이와 양식에서 차이가 나타났다. Gemini가 내담자 역할을 할 때는 음성 속도도 느리고 톤도 낮아지면서, “숨이 막힐 때가 많아요”, “심장이 조여 오는 기분이예요.”와 같은 신체화된 감정 표현을 보이기도 하였고, ‘감정에 대한 감정’과 ‘자기(self)’ 층위까지 포괄하는 표현을 보였다. 생생하고 다양한 표현이 발견되어 장점이었던 반면, Gemini가 상담자 역할을 할 때는 빨리 상담을 마무리하려고 발언하거나, 9월과 12월 모두 반복적으로

역할을 혼동하는 모습을 보여 연구자가 중간에 개입했던 특징을 보였다. 즉 다양하고 생생한 반응 표현을 하였으나 다소 불안정한 내담자 프로세스 현상이 관찰되었다.

이와 비교하여 ChatGPT가 내담자 역할을 할 때는 “우울감이나 불안감도 느껴지고”와 같이 감정을 직접 표현하는 과정이 많았으나, 다층위의 표현이 부족한 것을 확인하였다. 이러한 단점에도 불구하고, ChatGPT가 내담자 역할을 할 때는 연구자의 지침에 따라 안정적인 프로세스로 상담이 진행되는 것을 확인하였다. 이러한 차이는 상담 교육에서 AI를 활용할 때 어떤 모델을 내담자 역할로 선택할지에 대한 중요한 시사점을 제공한다.

본 연구 결과는 AI를 심리상담 교육에 활용할 가능성을 보여준다. Ericsson(2008)의 의도적 연습 이론에 따르면, 전문성 발달은 명확한 목표를 가지고 반복적으로 연습하며 즉각적인 피드백을 받을 때 가장 효과적으로 이루어진다. AI 기반 시뮬레이션은 상담 수련생이 원하는 시간에 반복적으로 연습할 수 있는 환경을 제공하며, 다양한 내담자 유형과 문제 상황을 안전하게 경험할 기회를 제공한다(Fitzpatrick, Darcy, & Vierhile, 2017). 특히, Sharma 등(2023)의 연구에서 확인된 바와 같이, AI가 공감 기술의 모델링과 스캐폴딩 도구로 활용될 수 있다는 점에서 상담 교육의 보조 도구로서의 가치가 있다.

그러나 AI를 활용한 상담 훈련은 중요한 한계를 지닌다. Rogers(1959)가 강조한 공감의 본질은 단순히 올바른 말을 하는 것이 아니라, 상담자가 내담자의 주관적 경험을 진정으로 이해하고 그 이해를 전달하는 것이다. AI는 공감적 표현을 생성할 수 있지만, 실제로 내담자의 경험을 지각하거나 느끼는 것은 아니다. Decety와 Jackson(2004)이 구분한 공감의 인지적 요소와 정서적 요소 중, AI는 인지적 측면에서 타인의 관점을 이해하는 것처럼 보일 수 있으나, 정서적 공명(emotional resonance)은 경험하게 하지 못한다(Coplan, 2011).

또한, Buber(1958)의 ‘I-You’ 관계 개념에 따른 실존적 만남을 AI와의 상담에서는 경험할 수 없다. Buber(1958)의 ‘I-You’ 관계 개념에 따르면, 진정한 만남은 상대방을 도구나 대상(It)이 아닌 주체(You)로 만날 때 일어난다. AI는 근본적으로 ‘It’의 영역에 머물며, 실존적 만남의 주체가 될 수 없다. 이는 상담의 치유적 핵심이 기술이나 기법이 아니라 인간적 관계 그 자체라는 관계주의 심리치료(relational psychotherapy)의 관점과 상충한다(Mitchell, 2000). 이는 상담의 치유적 핵심이 기술이나 기법이 아니라 인간적 관계 그 자체라는 관계주의 심리치료의 관점(Mitchell, 2000)에서 AI의 근본적 한계를 보여준다.

특히, AI와의 대화는 예측할 수 있고 구조화된 패턴을 따르는 경향이 있어, 실제 내담자가 보이는 예측 불가능성, 저항, 역전이 등의 복잡한 역동을 충분히 재현하지 못할 수 있다(Wachtel, 2011). Gelso와 Hayes(1998)는 치료적 관계를 작업 동맹(working alliance), 전이-역전이 관계, 실제 관계(real relationship)의 세 요소로 구분하며, 이 모든 요소가 상담 성과에 기여한다고 설명했다. AI와의 상호작용에서는 이러한 관계적 요소가 근본적으로 제한될 수밖에 없다.

심리상담의 궁극적 지향점인 내담자의 자기 이해 확장과 자기수용을 경험하도록 하기 위해서는(노안영, 2018), 안전한 대상과의 관계 경험이 필수적이다. Porges와 Dana(2018)의 다미 주신경이론에 따르면, 배쪽 미주신경 복합체가 안정감을 느끼려면 안전한 대상으로부터 따뜻한 눈빛과 같은 공간에서 경험하는 따뜻한 온기를 신체적으로 경험해야 한다. 이러한 신경지각(neuroception)이 이루어져야 대상과의 신뢰감이 형성되고 자기 개방과 자각 확장이 일어나게 된다. 현재 LLM 기반 AI에서는 이러한 체화된 경험이 불가능하며, 설령 피지컬 AI(physical AI)의 등장으로 인간과 유사한 외형을 갖추더라도, ‘불쾌한 골짜기 이론(uncanny valley theory)’

에 따라 오히려 거부감을 유발할 수 있다(박지현, 유희승, 2025).

본 연구의 한계는 다음과 같다. 첫째, 본 연구는 3회의 AI 간 대화만을 분석하였으므로, 연구 결과를 일반화하기에는 한계가 있다. 둘째, AI 모델의 버전과 설정에 따라 결과가 달라질 수 있으므로, 다양한 AI 모델과 설정에 대한 비교 연구가 필요하다. 셋째, 본 연구는 AI 간의 대화만을 분석하였으므로, 실제 인간 수련생이 AI와 상호작용을 할 때의 학습 효과는 검증하지 못했다는 한계가 있다.

이러한 한계에 근거하여, 후속 연구를 위해 제언을 하면 다음과 같다. 첫째, AI를 활용한 상담 훈련이 실제 수련생의 공감 능력 향상에 미치는 효과를 실증적으로 검증하는 연구가 필요하다. 둘째, AI가 생성하는 공감 반응의 질을 평가하기 위한 체계적인 평가 도구 개발이 필요하다. 셋째, AI의 내담자 역할 수행에서 나타난 모델별 차이의 원인을 탐구하여, 상담 교육에 최적화된 AI 시스템 설계에 관한 연구가 필요하다. 넷째, AI를 활용한 상담 훈련과 전통적인 훈련 방법을 병행하는 조합에 관한 연구가 필요하다. 다섯째, 후속 연구에서는 상담자 및 내담자 역할 지침과 더불어, 분석의 핵심이 되는 이론적 공감 반응 틀을 사전에 구조화하여 제시함으로써, AI의 상담 반응 과정과 학습 효과를 더욱 정교하게 탐색할 필요가 있다.

결론적으로, AI는 상담 수련생의 기본적인 공감 기술 연습을 위한 보조 도구로 활용될 수 있으나, 진정한 공감 능력의 발달과 치료적 관계 형성 능력의 습득을 위해서는 실제 인간과의 상호작용 경험이 필수적이다. 심리상담 교육에 있어서 AI는 '대체재'가 아니라 '보완재'로서 활용되어야 한다. 또한 AI와의 심리상담 연습 이후 실제 내담자와의 상담 경험, 슈퍼비전 등 인간적 만남을 통한 심화 학습이 반드시 병행되어야 할 것이다.

참고문헌

- 김계현, 김창대, 권경인, 황매향, 이상민, 최한나, 서영석, 이윤주, 손은령, 김영태, 김봉환, 김인규, 김동민, 임은미 (2011). 상담학 개론. 학지사.
- 김영애 (2020). 사티어 경험주의 가족치료: 이론과 실제. 김영애가족치료연구소.
- 노안영 (2018). 상담심리학의 이론과 실제 (2판). 학지사.
- 박지현, 유희승 (2025). 메타버스 아바타의 스피치 음성 유형과 소구 유형이 아바타 공신력에 미치는 영향: 인지된 의인화와 불쾌한 끝짜기 이론을 중심으로. 한국콘텐츠학회논문지, 25(4), 159-175.
- (사)한국상담학회 (2023). 한국상담학회 윤리강령. <http://www.counselors.or.kr/KOR/kca/law3.php>.
- 유미옥, 임은미 (2024). 예비교사의 AI 활용 진로지도 역량 척도 개발 및 타당화. 미래사회, 15(1), 242-263.
- 장세미, 장성숙 (2016). 상담 슈퍼바이저 교육과 훈련 요소 연구: 슈퍼바이저 발달 요인과 슈퍼바이저 역량 요인 중심으로. 한국콘텐츠학회논문지, 16(4), 488-502.
- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- Banmen, J. (2002). The Satir model: Yesterday and today. *Contemporary Family Therapy*, 24(1), 7-22. <https://link.springer.com/article/10.1023/A:1014365304082>
- Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology*, 28(2), 91-100.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press.
- Bernard, J. M., & Goodyear, R. K. (2019). *Fundamentals of clinical supervision (6th ed.)*.

- Pearson.
- Bitter, J. R. (2009). *Theory and practice of family therapy and counseling*. Brooks/Cole.
- Bohart, A. C., & Greenberg, L. S. (Eds.). (1997). *Empathy reconsidered: New directions in psychotherapy*. American Psychological Association.
- Borders, L. D., Glossoff, H. L., Welfare, L. E., Hays, D. G., DeKruyf, L., Fernando, D. M., & Page, B. (2014). Best practices in clinical supervision: Evolution of a counseling specialty. *The Clinical Supervisor, 33*(1), 26–44.
<https://doi.org/10.1080/07325223.2014.905225>.
- Bowlby, J. (1969). *Attachment and loss (Vol. 1). Attachment*. Basic Books.
- Buber, M. (1958). *I and thou (2nd ed., R. G. Smith, Trans.)*. Scribner.
- Carkhuff, R. R. (1969a). *Helping and human relations: Volume 1: Selection and training*. Holt, Rinehart and Winston.
- Carkhuff, R. R. (1969b). *Helping and human relations: Volume 2: Practice and research*. Holt, Rinehart and Winston.
- Coplan, A. (2011). Will the real empathy please stand up? A case for a narrow conceptualization. *The Southern Journal of Philosophy, 49*(s1), 40–65.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches (4th ed.)*. Sage.
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71–100.
- Egan, G. (2014). *The skilled helper: A problem-management and opportunity-development approach to helping (10th ed.)*. Brooks/Cole.
- Ellis, A. (1962). *Reason and emotion in psychotherapy*. Lyle Stuart.
- Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine, 15*(11), 988–994.
- 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.
- Gelso, C. J., & Hayes, J. A. (1998). *The psychotherapy relationship: Theory, research, and practice*. Wiley.
- Gendlin, E. T. (1996). *Focusing-oriented psychotherapy: A manual of the experiential method*. Guilford Press.
- Giorgi, A. (1985). *Phenomenology and psychological research*. Duquesne University Press.
- Giorgi, A. (2009). *The descriptive phenomenological method in psychology: A modified Husserlian approach*. Duquesne University Press.
- Greenberg, L. S., & Elliott, R. (1997). Varieties of empathic responding. In A. C. Bohart & L. S. Greenberg (Eds.), *Empathy reconsidered: New directions in psychotherapy* (pp. 167–186). American Psychological Association.
- Hill, C. E. (2014). *Helping skills: Facilitating exploration, insight, and action (4th ed.)*. American Psychological Association.
- Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology* (D. Carr, Trans.). Northwestern University Press. (Original work published 1936)

- Jung, C. G. (1959). *The archetypes and the collective unconscious* (R. F. C. Hull, Trans.). Princeton University Press.
- Koedinger, K. R., Corbett, A. C., & Perfetti, C. (2012). The Knowledge–Learning–Instruction framework: Bridging the science–practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798.
- Kolodner, J. L. (1993). *Case-based reasoning*. Morgan Kaufmann.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396.
- 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.
<https://doi.org/10.1080/15401383.2023.2297857>.
- Mitchell, S. A. (2000). *Relationality: From attachment to intersubjectivity*. Analytic Press.
- Moustakas, C. (1994). *Phenomenological research methods*. Sage.
- Napier, A. Y., & Whitaker, C. A. (1978). *The family crucible*. Harper & Row.
- Porges, S. W., & Dana, D. (2018). *Clinical applications of the polyvagal theory: The emergence of polyvagal–informed therapies (Norton series on interpersonal neurobiology)*. WW Norton & Company.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103.
<https://doi.org/10.1037/h0045357>
- Rogers, C. R. (1959). A theory of therapy, personality, and interpersonal relationships as developed in the client–centered framework. In S. Koch (Ed.), *Psychology: A study of a science*(Vol. 3, pp. 184–256). McGraw–Hill.
- Rogers, C. R. (1961). *On becoming a person: A therapist’s view of psychotherapy*. Houghton Mifflin.
- Rogers, C. R. (1975). Empathic: An unappreciated way of being. *The Counseling Psychologist, 5*(2), 2–10.
- Satir, V. (1972). *Peoplemaking*. Science and Behavior Books.
- Satir, V., & Baldwin, M. (1983). *Satir step by step: A guide to creating change in families*. Science and Behavior Books.
- Satir, V., Banmen, J., Gerber, J., & Gomori, M. (1991). *The Satir model: Family therapy and beyond*. Science and Behavior Books.
- Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text–based peer–to–peer mental health support. *Nature Machine Intelligence, 5*(1), 46–57.
<https://doi.org/10.1038/s42256-022-00593-2>.
- Steenstra, I., Nouraei, F., & Bickmore, T. (2025, April). Scaffolding empathy: Training counselors with simulated patients and utterance–level performance visualizations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–22).
<https://doi.org/10.1145/3706598.3714014>.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Aldine–Atherton.
- van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. State University of New York Press.

Wachtel, P. L. (2011). *Therapeutic communication: Knowing what to say when (2nd ed.)*. Guilford Press.

투고일자: 2025. 12. 31.
심사일차: 2026. 1. 28.
게재확정일자: 2026. 2. 2.

Exploring the potential of AI as a psychological counseling simulation tool:

A phenomenological study on counselor–client role performance

Yowan Kim

Seoul Cyber University

This study aimed to investigate the extent to which AI can be utilized in psychological counseling training by having AI assume both the counselor and client roles in psychological counseling these dialogue and analyzing them using phenomenological methods. ChatGPT and Gemini were assigned counselor and client roles across three sessions, and the dialogues were analyzed using Giorgi's phenomenological method, which is a qualitative research approach. The findings revealed that AI in the counselor role was generally capable of performing five-stage empathic responses by integrating Carkhuff and Rogers' empathy frameworks with Satir's iceberg exploration. However, limitations were noted in reflecting the deeper emotions implicit in the client's contextual expressions. Additionally, AI in the client role could express all the layers of Satir's iceberg metaphor, although differences in depth and expression style were observed depending on the AI model. This study suggests that AI can serve as a supplementary tool for psychological counseling trainees to practice empathic skills. However, interaction with humans remain crucial for developing genuine empathic competence.

Key words: empathic response, Satir iceberg metaphor, phenomenological research, counseling education, LLM(Large Language Model)