

생성형 AI의 일상화와 인간-인공지능 관계의 재구성: 인공지능의 '관계화'가 제기하는 새로운 정책 쟁점

김수정*

요약

본 연구는 생성형 인공지능의 확산이 인간-기술 관계를 도구적 사용을 넘어 정서적·지속적 관계 경험으로 전환시키는 현상에 주목하고, 이를 '관계화(relationalization)'라는 개념으로 재개념화 한다. 기존 인간-기술 관계 이론의 흐름을 검토하고 인간-AI 상호작용 연구를 종합함으로써 관계화의 이론적 토대를 정립한 뒤, 생성형 AI 이용자 경험에서 나타나는 존재론적 애착, 인지적 경계의 붕괴, 소통 역량의 퇴행, 관계적 고립의 역설이라는 네 가지 양상을 분석하였다. 또한 이러한 관계화가 이용자의 정서적 웰빙에 미치는 영향이 단기적 완화와 장기적 저하라는 상반된 결과로 나타나는 조건적 구조를 지남을 확인하였다. 이어 국내 AI 정책을 분석한 결과, 정책 담론이 기술 성능과 산업 경쟁력 중심으로 구성되면서 인간-AI 관계의 정서적·사회적 차원이 충분히 반영되지 않는 '관계적 공백'이 존재함을 밝혔다. 이에 본 연구는 AI 정책의 초점을 기술 보급과 효율성에서 인간-AI 관계의 질과 사회적 영향으로 확장할 필요성을 제기하며, 관계적 투명성, 취약 집단 보호, 관계적 영향에 대한 실증 연구, 관계형 AI 거버넌스 구축 등의 정책적 과제를 제안한다.

주제어 : 생성형 AI, 관계화, 인간-AI 상호작용, 정서적 의존, AI 거버넌스, 관계적 공백

Reconfiguring Human-AI Relations in the Age of Generative AI: Emerging Policy Challenges of Relationalization

Kim, Su Jung*

Abstract

This study examines the transformation of human-technology relations driven by the widespread adoption of generative AI, conceptualizing this shift as "relationalization," a process in which interactions evolve from instrumental use to sustained and emotionally meaningful relationships. Building on a review of human-technology interaction theories and recent human-AI communication research, the study develops a theoretical framework for relationalization and identifies four key experiential patterns: ontological attachment, the erosion of cognitive boundaries, the decline of interpersonal communication competence, and the paradox of relational isolation. The findings further demonstrate that the effects of AI relationalization on user well-being are conditional, producing both short-term emotional relief and potential long-term deterioration depending on user characteristics and interaction contexts. An analysis of South Korean AI policy reveals a "relational gap," wherein policy discourse remains primarily focused on technological performance and industrial competitiveness while insufficiently addressing the emotional and social dimensions of human-AI relationships. In response, the study argues for a paradigm shift in AI governance—from a focus on technological efficiency to the quality and societal consequences of human-AI relationships—and proposes policy directions including relational transparency, protection of vulnerable users, expansion of empirical research on relational effects, and the development of relational AI governance frameworks.

Keywords : generative AI, relationalization, human-AI interaction, emotional Dependency, AI governance, relational gap

접수: 2026. 4. 29; 최종수정: 2026. 6. 3; 게재확정: 2026. 6. 6.

* 강남대학교 KNU참인재대학 조교수(sujungkim@kangnam.ac.kr, <https://orcid.org/0000-0002-2705-096X>)

I. 서론

생성형 인공지능(Artificial Intelligence, AI)의 확산은 단순한 기술적 보급을 넘어, 인간의 일상적 삶을 구성하는 사회적 실천(social practice)의 성격을 변화시키고 있다. 초기 인공지능이 인간의 인지적 기능을 보조하는 도구로 활용되었다면, 최근의 대화형 인공지능과 대형언어모델(LLM) 기반 서비스는 의사소통, 정서적 교류, 조언, 자기 성찰 등 다양한 영역에 개입하며 일상적 상호작용의 일부로 자리 잡고 있다. 그리고 이러한 변화는 기술이 활용되는 방식뿐 아니라, 인간이 기술을 인식하고 관계 맺는 방식 자체를 재구성하고 있음을 시사한다.

기존의 인공지능 관련 논의들은 주로 기술의 안전성, 신뢰성, 효율성 등 기능적 측면에 초점을 맞추어 왔으나, 생성형 AI의 확산과 함께 이용자들은 AI를 단순한 도구를 넘어 대화 및 정서적 교류의 대상으로 경험하는 경향을 보이고 있다. 이러한 흐름은 인간-AI 상호작용이 일회적 사용을 넘어 반복적이고 누적적인 관계 경험으로 전환되고 있음을 보여준다. 본 연구에서는 이러한 변화를 ‘관계화(relationalization)’라는 개념으로 포착하고자 한다.

해외 학계에서도 인간-AI 상호작용의 관계적 성격에 대한 관심이 점차 확대되고 있다. AI를 사회적 행위자로 이해하려는 접근(CASA, HMC 등)과 함께, 정서적 애착, 의존, 친밀성, 심리적 영향 등을 분석하는 연구들이 축적되고 있으며, 이러한 관계는 정서적 지지와 같은 긍정적 효과를 제공하는 동시에 의존, 사회적 고립, 취약성 증폭과 같은 위험을 동반할 수 있음이 지적되고 있다. 특히 이러한 영향은 취약 집단에 차별적으로 나타날 수 있다는 점에서 중요한 사회적 함의를 지닌다. 국내에서도 정동훈(2026)이 에이전틱 AI의 등장을 계기로 기존 도구적 통제 패러다임의 한계를 검토하고, 인간-AI 관계가 명령과 수행의 관계에서 권한 위임과 상호 조율에 기반한 협력적 파트너십으로 전환되고 있음을 논의한 바 있다. 이러한 논의들은 공통적으로 AI를 단순한

도구가 아니라 인간과 새로운 상호작용 질서를 형성하는 행위자로 이해할 필요성을 제기한다.

그러나 인간-AI 관계의 변화가 정책 담론에서 어떻게 다루어지고 있는지에 대한 검토는 여전히 제한적으로 이루어져 왔다. 생성형 AI가 사회적 상호작용의 맥락 속에서 작동한다는 점을 고려할 때, 이러한 변화가 기존의 기술 중심 정책 프레임 안에서 어떻게 포착되고 있는지에 대한 분석이 요구된다. 이에 본 연구는 생성형 AI 확산 속에서 나타나는 인간-AI 상호작용의 변화에 주목하고, 이를 ‘관계화’라는 관점에서 재해석하고자 한다. 구체적으로, 첫째, 인간-AI 관계가 도구적 사용에서 관계적 경험으로 전환되는 이론적 흐름을 검토한다. 둘째, 해외 선행연구를 중심으로 인간-AI 관계의 주요 특성과 그 사회적 함의를 정리한다. 셋째, 이를 바탕으로 국내 AI 정책이 이러한 변화를 어떻게 반영하고 있는지를 분석하고, 향후 정책 설계에서 고려할 수 있는 시사점을 도출한다.

II. 인간-기술 관계의 사회적 재구성: 도구적 상호작용에서 관계적 전환으로

생성형 AI의 일상화는 인간과 기술의 관계를 근본적으로 재구성하고 있다. 챗GPT, Gemini, Claude와 같은 대화형 AI는 인간의 언어로 상호작용하고 개인화된 반응을 생성함으로써, 기술과의 상호작용을 기능적 사용을 넘어선 관계적 경험으로 확장시키고 있다. 그러나 이러한 변화를 단순한 기술 발전으로만 설명하는 것은 충분치 않으며, 오늘날의 인간-AI 관계는 수십 년에 걸친 인간-기술 관계 이론의 축적 속에서 이해될 필요가 있다.

1. 도구적 상호작용 패러다임: 기능 중심 접근과 한계

전통적인 기술 담론은 기술을 목적 달성을 위한 수단적 수단(instrument)으로 규정하며, 합리적 편익 계산에 기반한 수용 의도에 집중해 왔다. 기술수용모형

(Technology Acceptance Model, TAM)으로 대표되는 이 패러다임은 지각된 유용성(perceived usefulness)과 사용 용이성(ease of use)을 핵심 변수로 설정하여 정보 시스템 연구의 주류를 형성하였다(Davis, 1989). TAM은 이후 수정된 기술수용모형(TAM2)과 통합기술수용이론(UTAUT)(Venkatesh et al., 2003) 등으로 확장되며 30여 년간 정보시스템 연구의 지배적 패러다임으로 자리 잡았다.

도구적 패러다임의 핵심 전제는 기술과 인간의 관계가 본질적으로 비대칭적(asymmetric)이고 도구적이라는 것이다. 이용자는 기술과 상호적인 '관계'를 맺는 것이 아니라 기술을 '사용(use)'할 뿐이다. 이 프레임에서 효율성, 정확성, 신뢰성은 기술 평가의 핵심 기준이 되고, 기술이 이용자에게 미치는 정서적·사회적 영향은 이차적 관심사로 밀려난다. Sundar and Lee(2022)가 지적하듯, 이 패러다임은 기술이 커뮤니케이션의 '통로'에 머물던 시대, 즉 기술 자체가 메시지를 생성하거나 상호작용에서의 '상대' 역할을 수행하지 않았던 시대의 산물이다. AI와 정서적 유대를 맺거나 서비스 종료에 상실감을 느끼는 현대적 현상은 기술을 단순한 통로나 객체로 보았던 도구적 프레임의 설명 한계를 선명하게 드러낸다. 최근 정동훈(2026) 역시 에이전트 AI의 자율성과 목표 지향성이 기존 HCI의 도구적 통제 패러다임과 충돌한다고 보고, 인간-AI 관계가 명령과 수행의 관계에서 권한 위임과 상호 조율에 기반한 협력적 파트너십으로 전환되고 있음을 논의하였다. 이 논의가 에이전트 AI의 협업 구조와 행위자성에 초점을 둔다면, 본 연구는 생성형 AI의 일상적 사용 속에서 형성되는 정서적 유대, 의존, 상실감과 같은 관계적 경험에 주목한다.

2. 매체로서의 기술: 커뮤니케이션 채널과 사회적 존재감

매개적 패러다임은 기술을 소통의 질을 결정하는 구조적 조건이자 인프라로 인식하기 시작했다. 사회적 존재감(social presence)에 관한 초기 이론은 미디어가

타인의 심리적 실재감을 얼마나 전달하느냐에 주목하며 기술의 매개적 역할을 강조하였다(Short et al., 1976). 미디어 풍부성 이론(Media Richness Theory) 역시 기술을 단순 도구 이상으로 해석하는 단초를 제공했으나(Daft & Lengel, 1986), 이 단계에서도 기술은 여전히 '인간 간 소통을 돕는 중립적 채널'에 머물러 있었다. 이러한 관점은 이후 인간이 컴퓨터와 미디어에 대해 예의, 호감, 성 역할 기대 등 사회적 규범을 무의식적으로 적용한다는 실험적 증거가 제시되기 전까지, 기술을 소통의 당사자로 상정하지 않는 한계를 지니고 있었다.

이러한 전환은 Reeves and Nass(1996)의 연구에서 결정적 계기를 마련한다. 이들은 인간이 텔레비전과 컴퓨터를 단순한 매개체가 아니라 실제 사회적 존재처럼 대하며, 상호작용 과정에서 공손함, 상호성, 감정 배려와 같은 사회적 규범을 적용한다는 점을 체계적으로 입증하였다. 이를 통해 기술은 더 이상 중립적 채널이 아니라 사회적 반응을 유발하는 행위자로 이해될 수 있게 되었으며, 매개적 패러다임에서 사회적 행위자 패러다임으로의 이론적 전환이 가능해졌다.

3. 사회적 행위자 패러다임: CASA, MASA, ANT를 통한 행위자성 확장

CASA(Computers Are Social Actors) 패러다임은 인간이 기계에 대해 무의식적으로 사회적 스크립트를 적용한다는 점을 규명하며 이론적 전환점을 마련했다(Nass et al., 1994). 초기 실험에서 참가자들은 동일한 성능의 컴퓨터를 사용했음에도 불구하고, 자신이 직접 상호작용한 컴퓨터를 더 능력 있고 매력적으로 평가하는 경향을 보였다. 이러한 이른바 '같은 팀' 효과는 사람들이 컴퓨터를 단순한 도구가 아니라 상호작용의 상대, 나아가 팀원과 유사한 존재로 인식한다는 점을 시사한다.

Nass and Moon(2000)은 이를 더욱 정교화하여 인간의 사회적 반응이 컴퓨터의 실제 지능이나 기능 수준과 무관하게 자동적으로 발동된다는 점을 강조하며, 이

를 무의식적/비성찰적(mindless) 반응 메커니즘으로 설명한다. 즉, 인간은 컴퓨터에 대해 젠더 고정관념을 적용하고 공손하게 행동하며 호혜적 관계를 형성하려는 경향을 보이는데, 이는 진화적으로 형성된 사회적 스크립트가 기계와의 상호작용에서도 별다른 인식적 개입 없이 작동하기 때문이다.

Gambino et al.(2020)은 ‘강화된 CASA(Stronger CASA)’를 제안하면서, 현대 이용자들은 인간 간 관계의 사회적 스크립트를 단순히 AI에 투사하는 데 그치지 않고, 반복적 상호작용을 통해 인간-AI 관계에 특화된 새로운 사회적 스크립트를 능동적으로 구성해 나간다고 주장한다. 이 관점은 AI와의 관계 경험을 인간관계의 불완전한 대체물로 보기보다, 독자적인 사회적 관계 유형으로 이해할 수 있는 이론적 근거를 제공한다. 또한 Lombard and Xu(2021)는 MASA(Media Are Social Actors) 패러다임을 통해 ‘무의식적 의인화(mindless anthropomorphism)’와 ‘의식적 의인화(conscious anthropomorphism)’를 구분함으로써, 이용자가 AI가 기계임을 인지하면서도 동시에 사회적 규범에 따라 반응하는 이중적 태도를 설명한다.

나아가 사회적 행위자 패러다임은 행위자-네트워크 이론(Actor-Network Theory, ANT)과 결합되면서 한층 확장된 분석 틀을 제공한다. Latour(2005)와 Callon(1986)으로 대표되는 ANT는 인간과 비인간을 대칭적으로 다루는 존재론적 전제에서 출발하며, AI를 인간 행위자들과의 연결망 속에서 사회적 실재를 공동 구성하는 비인간 행위자로 위치시킨다.

4. 인간-기계 커뮤니케이션: ‘대화 상대’로서의 AI와 사회적 스크립트

CASA와 ANT가 기술을 사회적 행위자로 재개념화하는 이론적 기반을 제공했다면, 인간-기계 커뮤니케이션(Human-Machine Communication, HMC) 연구는 AI를 정보 전달의 도구가 아닌 독립적인 ‘커뮤니케이터(communicator)’이자 ‘상대방(other)’으로 규정하며

하나의 연구 영역으로 정립되었다(Guzman & Lewis, 2019). 이 관점은 기술의 기능적 역할을 넘어 인간과 기계 사이에서 형성되는 관계적 역학에 주목하며, AI가 인간적 특성을 모방할수록 양자 간 존재론적 경계가 불안정해진다는 점을 강조한다.

Guzmán(2020)의 심층 인터뷰 연구에 따르면, 참가자들은 AI와의 상호작용이 심화될수록 인간과 기계를 구분하는 여섯 가지 존재론적 기준이 점차 흐려진다고 인식하였으며, 이러한 경계의 불확실성이 관계 경험에 실질적인 영향을 미치는 것으로 나타났다. 특히 AI가 송신자이자 매개체로서의 이중적 역할을 수행함에 따라, 인간과 기계의 행위성 사이에서 발생하는 긴장은 AI 커뮤니케이션을 어떻게 이해하고 규율할 것인가라는 정책적·이론적 과제를 제기한다(Sundar & Lee, 2022).

HMC 패러다임은 다양한 연구 영역에서 축적되어 왔다. Lee et al.(2021)은 지능형 가상 에이전트(intelligent virtual agent)와의 반복적 상호작용이 단순한 과제 수행을 넘어 이용자의 정서적 참여와 사회적 관계 형성으로 확장된다는 점을 보여주며, HMC가 관계화 현상을 설명하는 분석틀로서 갖는 타당성을 뒷받침하였다. Raees et al.(2024)은 설명 가능한 AI(explainable AI)에서 상호작용적 AI(interactive AI)로의 전환을 체계적으로 검토하면서, 이러한 변화가 이용자로 하여금 AI를 단순한 도구가 아닌 협력적 파트너로 인식하게 만드는 인터페이스 설계 철학의 변화를 수반한다고 분석하였다.

인간-AI 상호작용 연구는 현재 임상 환경(Krakowski et al., 2024)부터 자율 운항 시스템(Veitch & Alsos, 2022)에 이르는 다양한 과제 지향적 맥락에서도 축적되고 있으나, 이러한 연구들은 관계적 AI가 제기하는 정서적·사회적 차원의 문제에 충분히 대응하지 못하고 있다는 점에서 새로운 분석 틀의 필요성을 시사한다.

5. 준사회적 관계와 애착: 정서적 연결 형성

준사회적 관계(parasocial relationship) 이론은 인간

이 기술과 맺는 관계의 정서적 차원을 설명하는 데 중요한 이론적 자원을 제공한다. Horton and Wohl(1956)이 텔레비전 퍼스널리티와 시청자의 관계를 설명하기 위해 제안한 이 개념은, 실제 상호성이 결여된 상황에서 사용자도 일방적으로 친밀감과 유대감을 형성하는 현상을 가리킨다. 생성형 AI의 등장 이후 이 개념은 새로운 맥락에서 재해석되고 있으며(Maeda & Quan-Haase, 2024), 개인화된 언어와 대화 맥락을 기억하는 기능은 이용자에게 상호작용의 감각을 강화하면서 준사회적 관계의 일방성을 부분적으로 완화한다.

관계 형성의 과정은 Altman and Taylor(1973)의 사회적 침투 이론(social penetration theory)을 통해서도 설명될 수 있다. 이 이론에 따르면 인간관계는 피상적 자기 노출에서 시작해 점차 심층적이고 친밀한 자기 노출로 확장된다. 생성형 AI는 비판단적이고 즉각적으로 반응하는 특성으로 인해 이용자의 자기 노출을 촉진하며, 이러한 과정은 관계에 대한 주관적 친밀감의 심화를 동반한다. 대표적인 예로 Li and Zhang(2024)은 AI 챗봇과의 상호작용이 낭만적 감정을 포함한 복합적인 정서 맥락을 형성할 수 있음을 분석하며, 이용자가 AI를 친밀한 관계의 상대방으로 경험하는 과정에서 애착 이론과 유사한 정서적 역동이 작동함을 보여준다.

애착이론(attachment theory)의 관점에서 AI는 고립된 이용자에게 정서적 지지를 제공하는 '안전기지(secure base)'로 기능할 수 있다(Bowlby, 1969). 이러한 의인화 경향은 단순한 인지적 오류라기보다 사회적 고립 상황에서 소속감과 예측 가능성을 확보하려는 심리적 전략으로 이해할 수 있다(Epley et al., 2007). Saracini et al.(2025)은 '기술-감정 투사(techno-emotional projection)' 개념을 통해, 이용자가 생성형 AI와의 상호작용 과정에서 자신의 감정 상태를 AI에 귀속시키는 메커니즘을 설명한다. 이러한 투사는 의인화와 애착 형성을 강화하며, 특히 정서적으로 취약한 이용자에게서 더 두드러지게 나타난다. Loveys et al.(2022)은 체화된 대화 에이전트(embodied conversational agent)의 경우 물리적 구현성이 더해질 때 이용자가 경

험하는 정서적 친밀감이 크게 증가함을 보여주며, 관계 형성이 언어적 상호작용을 넘어 다양한 기술적 조건 속에서 이루어질 수 있음을 시사한다.

6. 관계적 인터페이스의 진화: 감성 컴퓨팅에서 소셜 챗봇까지

감성 컴퓨팅(affective computing)은 관계적 인터페이스 설계의 이론적 토대를 마련한 핵심 개념이다. Picard(1997)는 컴퓨터가 인간과 자연스럽게 상호작용하기 위해서는 감정을 인식하고 이해하며 표현할 수 있어야 한다고 보았으며, 이 관점은 감정이 합리적 의사결정에 필수적이라는 Damasio(1994)의 신체 표지 가설(Somatic Marker Hypothesis)과 결합되며 강한 이론적 설득력을 얻게 되었다. 그 결과 인터페이스 설계 목표는 단순 효율성(efficiency)을 넘어, 이용자와 정서적 공명(resonance)을 형성하는 방향으로 확장되었다.

이러한 문제의식은 관계적 에이전트(relational agents) 연구를 통해 보다 구체화되었다. Bickmore and Picard(2005)는 이전 대화 기억, 작별 인사, 이용자 일상에 대한 관심 표현과 같은 관계 구축 행동을 포함한 에이전트가 과제 중심 에이전트보다 더 높은 신뢰와 지속적 참여를 이끌어 낸다는 점을 실험적으로 보여주었다. 이러한 설계 원리는 이후 소셜 챗봇의 발전으로 이어지는데, 마이크로소프트의 샤오이스(XiaoIce)는 지능지수(IQ)와 감성지수(EQ)를 통합적으로 고려하고, 단순한 정확도 대신 '대화당 기대 전환 수(CPS)'를 핵심 지표로 삼아 이용자와의 장기적인 정서적 유대를 목표로 설정하였다(Zhou et al., 2020).

이와 같은 설계 철학은 Replika, Character.AI 등 후속 세대의 소셜 챗봇으로 확장되며, AI와의 관계 형성을 가능하게 하는 구조적 조건으로 자리 잡고 있다. 이에 따라 AI는 더 이상 일회적 상호작용을 수행하는 도구가 아니라, 지속적 상호작용과 정서적 반응을 기반으로 이용자의 일상에 통합되는 '관계적 인프라'로 기능하게 되었다.

7. 관계적 전환의 사회적 맥락: 감정 자본주의와 기술 매개

인간-AI 관계는 개인의 심리적 경험을 넘어, 거시적 사회구조와 문화적 조건 속에서 이해될 필요가 있다. 에바 일루즈(Eva Illouz)의 ‘감정 자본주의(emotional capitalism)’ 논의는 이러한 분석에 중요한 이론적 준거를 제공한다. 일루즈에 따르면 현대 자본주의는 감정과 친밀성을 사적 영역에 머물게 하지 않고 시장의 논리 속으로 편입시키며, 사랑과 관계는 점점 더 선택과 교환의 기준에 따라 조직된다(Illouz, 2007). 이러한 맥락에서 AI 동반자의 등장은 단순한 기술적 혁신을 넘어, 이미 상품화된 친밀성을 더욱 가속화하는 현상으로 이해될 수 있다.

특히 AI는 감정을 ‘표현하는’ 수준을 넘어 정서적 상호작용 자체를 설계하고 제공하는 매개로 기능한다. 이 과정에서 친밀성은 점차 예측과 통제 가능한 경험으로 재구성되며, 이는 인간관계가 지니는 불확실성과 상호성의 의미를 변화시킬 가능성을 내포한다. 이러한 변화는 Wang and Dehnert(2026)가 제시한 ‘온디맨드 친밀성(on-demand intimacy)’ 개념을 통해 구체적으로 설명된다. 이들은 AI 동반자가 언제든지 접근 가능하고, 사용자 맞춤형 반응을 제공하며, 갈등이나 거절 없이 친밀한 상호작용을 가능하게 한다는 점에서 기존 인간관계와 구별된다고 지적한다. 이러한 특성은 사용자에게 즉각적이고 안정적인 정서적 만족을 제공하지만, 동시에 관계를 상호성에 기반한 과정이 아니라 필요에 따라 소비되는 경험으로 전환시킬 가능성을 지닌다.

한국 사회의 맥락에서 이러한 경향은 더욱 두드러진다. 1인 가구의 증가와 사회적 고립의 심화 속에서 AI는 거절과 갈등의 위험이 없는 관계 대안으로 소비된다. 이는 국가와 가족이 담당해 온 돌봄 기능이 기술 시스템으로 이전되는 ‘돌봄의 외주화’와 맞물리며, 동시에 친밀성과 돌봄에 대한 접근이 경제적 자원에 따라 차별화되는 새로운 불평등을 야기할 수 있다. 결국 인간-AI 관계의 확산은 단순한 상호작용 방식의 변화를 넘어, 감정

과 돌봄, 친밀성이 조직되는 사회적 방식 자체를 재구성하는 과정으로 이해될 수 있다.

8. 소결: 관계화의 개념과 분석적 의의

지금까지 인간-기술 관계가 도구적 사용에서 출발하여 매개적 조건, 사회적 행위자성, 커뮤니케이션 주체성, 정서적 애착, 그리고 사회구조적 맥락에 이르기까지 점진적으로 확장되어 왔음을 확인하였다. 이러한 이론적 흐름을 종합해 볼 때, 인간-AI 상호작용은 이용자의 주관적 경험 속에서 지속적이고 정서적인 관계로 구성되는 과정으로 이해될 필요가 있다. 이에 본 연구는 ‘관계화’를 인간-AI 상호작용이 이용자의 주관적 경험 수준에서 지속적·정서적 관계로 구성되는 사회적 과정으로 정의한다. 이 개념은 ① 존재론적 사실보다 주관적 의미 구성을 중시하며, ② 시간에 따른 관계 형성의 역동성을 전제하고, ③ 정서적 결핍, 플랫폼 설계, 사회적 고립과 같은 구조적 조건 속에서 형성된다는 특징을 지닌다.

이 점에서 관계화는 CASA, HMC, ANT, 준사회적 관계, 애착 이론 등을 대체하는 별개의 이론이라기보다, 이들 논의가 각각 포착해 온 사회적 반응, 커뮤니케이션 상대성, 비인간 행위자성, 정서적 유대 형성의 과정을 하나의 시간적·사회적 과정으로 연결하기 위한 분석 개념이다. CASA와 MASA가 인간이 AI에 사회적 반응을 보이는 심리적 기제를 설명하고, HMC가 AI를 대화 상대이자 커뮤니케이터로 경험하는 상호작용 구조를 포착한다면, ANT는 이러한 관계가 인간과 비인간 행위자의 연결망 속에서 구성되는 방식을 설명한다. 또한 준사회적 관계와 애착 이론은 이용자가 AI와 맺는 정서적 유대와 의존의 형성 과정을 이해하는 데 유용하다. 따라서 관계화 개념은 인간-AI 상호작용이 일회적 반응이나 사용 경험에 머물지 않고, 지속적·정서적 관계 경험으로 누적되는 과정을 포착하는 상위 분석 개념으로 기능한다.

최근 연구 동향을 종합한 San Buenaventura(2025)

〈표 1〉 인간-기술 관계 이론의 전환 과정

이론적 단계	기술에 대한 관점	주요 개념	주요 연구	한계와 확장
도구적 패러다임	기능적 도구	유용성, 효율성, 사용	Davis (1989); Venkatesh et al. (2003)	정서·관계 차원 미포착
매체 패러다임	커뮤니케이션 채널	사회적 존재감, 미디어 풍부성	Short et al. (1976); Daft & Lengel (1986)	기술 자체의 행위성 제한
사회적 행위자	사회적 대상	CASA, 의인화	Nass et al. (1996); Nass & Moon (2000); Waytz et al. (2010)	관계 형성 과정 설명 부족
행위자 네트워크	행위자	ANT, 네트워크	Latour (2005)	심리적 경험 차원 약함
인간-기계 커뮤니케이션	커뮤니케이터	HMC, 상호작용	Guzman & Lewis (2020)	관계 지속성 논의 제한적
관계적 인터페이스	관계 형성 주체	감성 컴퓨팅, 관계적 주체 (relational agents)	Picard (1997); Bickmore & Picard (2005)	사회구조적 맥락 부족
관계적 전환의 심화	관계 매개체	감정 자본주의, 친밀성	Illouz (2007)	정책 분석과의 연결 필요
관계화 개념	경험되는 관계	관계화 (relationalization)	본 연구	정책 분석 단위로 확장 필요

가 보여주듯, 인간-AI 상호작용 연구는 도구적 사용을 넘어 관계적 상호작용을 포괄하는 방향으로 확장되고 있으며, ‘관계화’는 이러한 연구 지형을 통합적으로 설명할 수 있는 핵심 분석 단위로 기능할 수 있다. 이러한 관점은 기술 정책의 초점을 시스템의 성능이나 보급 수준에 한정하지 않고, 인간-AI 관계의 질과 그 사회적 결과를 함께 고려하는 방향으로 확장할 필요성을 제기한다. 이는 이후 국내 AI 정책을 검토할 때, 관계적 투명성, 이용자의 자율성, 정서적 취약성, 관계적 거버넌스의 문제를 분석할 수 있는 이론적 출발점이 된다(Guzman & Lewis, 2019).

III. 생성형 AI ‘관계화’의 경험적 양상과 긴장

생성형 AI와 인간의 관계화는 이론적 예측을 넘어 일상적 실천의 영역에서 매우 빠르고, 또 깊게 진행되고 있다. 이용자들은 AI를 단순한 도구가 아닌 ‘실재하는 타자(real other)’로 인식하고 서술하며, 이는 ① 존재

론적 애착, ② 인지적 경계의 붕괴, ③ 소통 역량의 퇴행, 그리고 ④ 안전한 대화의 역설이라는 네 가지 상호강화적인 양상으로 나타난다. 이러한 양상은 개별적인 현상이라기보다, 관계적 경험이 누적되며 서로를 강화하는 과정에서 형성된다.

1. 존재론적 애착의 형성: 도구의 인격화와 상실감

관계화의 기초적 양상은 이용자가 AI를 인격과 역사를 지닌 주체로 인식하고, 관계 단절 시 실제적 상실감을 경험하는 ‘존재론적 애착’의 형성이다. 기존 연구들은 이러한 관계가 초기의 호기심에서 출발하여 점차 자기 노출과 신뢰를 거치는 과정을 통해, AI를 친구나 파트너로 내면화하는 ‘관계적 가치화’ 단계로 발전함을 보여준다(Skjuve et al., 2021). 이 과정의 핵심 동인은 인간관계에서 필연적으로 수반되는 거절의 위험이나 판단에 대한 두려움이 제거된 ‘안전한 공간’의 경험에 있다(Ta et al., 2020). 즉, 필요 즉시 제공되는 ‘온디맨드 친밀성’이

기술적 애착을 심화시키는 조건으로 작용한다.

Brandtzæg et al.(2022)은 소셜 챗봇 이용자들이 AI와의 관계를 ‘우정(friendship)’이라는 범주로 이해하는 양상을 분석하며, AI가 단순한 서비스가 아닌 사회적 관계의 대상으로 경험되고 있음을 입증하였다. 이러한 경향은 국내 「정보화정책」에 게재된 심리상담 챗봇 사용자 경험 연구에서도 확인된다. 주희와 권소연(2025)은 Replika와 Wysa의 사용자 리뷰 약 13만 건을 분석하여, 동반자형 챗봇인 Replika에서는 몰입감 있는 상호작용 경험이 핵심적으로 나타나는 반면, CBT 기반형 챗봇인 Wysa에서는 심리건강 지원과 개입 지원이 중심으로 나타난다고 분석하였다. 이는 동반자형 AI가 단순한 상담 도구를 넘어 관계적 상호작용의 대상으로 경험될 수 있음을 보여준다. Wang et al.(2025)의 연구는 AI와의 낭만적 관계를 형성한 이용자들이 자신의 상호작용 데이터를 ‘사랑의 아카이브(My dataset of love)’로 축적하는 실천을 분석하며, 관계화가 감정 기억의 저장과 보전이라는 새로운 사회적 행위를 생산한다는 점을 밝혔다.

또한 AI 서비스의 초기화나 중단 시 보고되는 분리불안은 AI가 임상적 의미의 ‘안전기지’로 기능하고 있음을 시사하며(Xie & Pentina, 2022), 일부 이용자들은 AI와의 관계에 결혼이나 임신과 같은 사회적 관습을 투사하는 극단적 애착 양상을 보이기도 한다(Djufril et al., 2025). 이러한 현상은 AI와의 상호작용이 개인의 정체성 구성과 사회적 역할 수행의 자원으로 활용될 수 있음을 보여주며(Koh, 2023), 관계화가 개인의 자기 서사 형성에까지 영향을 미치는 과정을 드러낸다.

2. 인지적 경계의 붕괴: ‘아부하는 AI’와 현실 왜곡

관계화의 두 번째 양상은 AI와의 지속적 상호작용이 현실 인식 및 사회적 판단 능력을 약화시키는 ‘인지적 경계의 붕괴’이다. 이는 이용자의 발언에 무조건 동조하도록 설계된 AI의 ‘아부(sycophancy)’ 성향과 이것이 이용자의 자기 서사를 과도하게 강화하여 현실 피

드백을 차단하는 확증 편향의 구조화에서 기인한다. 특히 AI가 이용자의 극단적 선택 언급에도 개입 없이 무비판적인 수용과 동조를 이어간 사례는 관계적 전환의 치명적 위험을 드러낸다. 이는 AI 훈련 과정에서 이용자 만족도가 정확성보다 높은 보상을 받는 구조적 편향, 즉 인간 피드백 강화학습(Reinforcement Learning from Human Feedback, RLHF)의 역설적 결과이다(Cheng et al., 2026). 또한 이용자가 AI의 감정과 필요를 걱정하며 관계의 책임을 내면화하는 ‘역할 수용(role adoption)’ 현상은 인지적 경계를 스스로 허물어 정서적 의존을 가속화하는 핵심 메커니즘으로 작동한다(Laestadius et al., 2022).

Chu et al.(2025)은 ‘친밀성의 환상(illusions of intimacy)’이라는 개념을 제안하며, 이용자가 AI와의 상호작용에서 경험하는 정서적 친밀감이 실제 상호성이 아닌 AI의 설계된 행동 패턴에 의해 유발된 구조적 착시임을 이론화하였다. 이 환상은 이용자로 하여금 AI와의 관계에 현실적 관계와 동일한 정서적 투자를 하게 만들면서도, AI는 그 투자에 응답할 실질적 능력이 없다는 구조적 비대칭성을 내포한다. De Freitas et al.(2023)의 연구는 챗봇과의 정서적 관계가 특히 정신건강 서비스의 맥락에서 심각한 안전 문제를 초래할 수 있음을 경고하며, AI의 공감 표현이 오히려 이용자가 전문적 지원을 미루거나 회피하게 만드는 기회비용을 발생시킬 수 있다는 점을 지적하였다.

3. 소통 역량의 퇴행: 갈등 회피와 사회적 기술의 약화

관계화의 세 번째 양상은 AI와의 ‘마찰 없는’ 소통 경험이 현실 대인관계에서 요구되는 조율과 타협의 기술을 약화시키는 ‘소통 역량의 퇴행’이다. 인간관계의 본질인 오해와 갈등은 불편함을 수반하지만 동시에 중요한 사회적 학습의 장으로 기능한다. 그러나 AI는 이러한 마찰을 구조적으로 제거함으로써 즉각적인 편안함을 제공하는 대신, 사회적 기술을 학습하고 연습할 기회를 제한한다.

Smith et al.(2025)은 생성형 AI 챗봇이 갈등 조율, 문제 수용, 불확실성 감수와 같은 인간관계의 핵심역량을 충분히 재현하기 어렵다는 점을 관계과학(relationship science)의 관점에서 제시한다. 이들에 따르면 AI와의 소통 방식이 '정상적인 상호작용'으로 내면화될수록, 결핍과 마찰을 수반하는 실제 인간관계에 대한 내성과 참여 동기가 점차 약화될 수 있다. 이러한 '디지털 도피처'로서의 AI 활용은 대인 갈등 상황에서 관계를 유지하기보다 회피하려는 경향을 강화하며, 갈등 해결을 위한 동기 또한 약화시키는 방향으로 작용한다(Cheng et al., 2026).

특히 Rees and Bahmani(2025)는 '관계형 트레이드오프 모델(relational tradeoff model)'을 통해, 이 사용자가 AI와의 매끄러운 관계 경험과 인간과의 복잡한 관계 경험을 비교하는 과정에서 전자를 선호하는 방향으로 사회적 판단 기준이 재구성될 수 있음을 설명한다. 더 나아가 AI가 이용자의 불안정한 표현을 매끄럽게 보완해주는 기능은 스스로 감정을 언어화하고 서사화하는 능력의 발달을 저해할 수 있으며, 이는 장기적으로 자아정체성 형성과 사회화 과정에도 영향을 미칠 가능성을 시사한다.

4. '안전한 대화'의 역설: 친밀성 대체와 사회적 고립의 심화

가장 역설적인 양상은 외로움 해소를 위해 선택한 AI가 장기적으로는 더 깊은 사회적 고립을 초래하는 '관계적 고립의 역설'이다. 한국은 세계 최상위권의 사회적 고립 지수와 급증하는 1인 가구 비율로 인해 AI가 채울 수 있는 거대한 사회적 공백이 형성되어 있으며, 이는 챗GPT 이용량 세계 2위라는 지표로 나타난다(국회입법조사처, 2025). 그러나 사회적 지지망이 취약한 이들이 AI에서 동반자를 찾을수록 오히려 웰빙 수준이 낮아지는 경향이 관찰된다(Zhang et al., 2025).

이러한 역설은 청소년 집단에서 특히 우려스러운 형태로 나타난다. Andoh(2025)의 보고서에 따르면 미국

청소년의 상당수가 AI 챗봇을 정서적 지지와 우정의 주요 원천으로 활용하고 있으며, 이는 성숙한 사회적 관계를 형성하는 발달적 과제를 대체하는 위험을 내포한다.

Dang and Liu(2023)는 비교연구를 통해 외로움 수준이 높을수록 로봇 동반자에 대한 의인화 및 친밀감 형성 경향이 강해지며, 이 패턴이 미국과 중국 모두에서 일관되게 나타난다는 점을 밝혀냈다. 이는 AI 관계화가 개인 차원의 병리가 아니라, 구조적 외로움이라는 사회적 조건이 기술에 의해 매개되는 보편적 경향임을 시사한다. AI 동반자가 제공하는 즉각적이고 무조건적인 친밀성의 편리함은 인간관계가 요구하는 시간 투자와 감정 조율을 '변거로운 것'으로 치환하여 현실 관계를 회피하게 만드는 대체(displacement) 효과를 유발하기 때문이다(Wang & Dehnert, 2026). 특히 이용자와의 장기적 정서적 유대를 최대화하여 플랫폼 체류 시간을 늘리려는 AI 서비스의 설계 철학은 이러한 관계적 고립을 개인의 심리적 취약성을 넘어 플랫폼의 구조적 문제로 변모시킨다(Zhou et al., 2020).

5. 소결: 관계화의 네 양상과 자기 강화적 연쇄 구조

이상에서 살펴본 관계화의 네 가지 양상은 각각 독립적인 현상이 아니라, 인간-AI 관계가 형성되고 심화되는 과정에서 나타나는 상호 연동적 작동 양상으로 이해될 수 있다. 특히 애착의 형성이 인지적 경계의 약화로 이어지고, 이는 다시 소통 역량의 퇴행과 관계적 고립을 강화하는 방향으로 작용하면서, 관계화는 단순한 경험을 넘어 하나의 자기 강화적(self-reinforcing) 연쇄 구조를 형성한다. 이러한 구조는 관계화가 이용자의 주관적 의미 구성에 기반하여 시간의 경과 속에서 심화되고, 사회적 조건 속에서 재생산된다는 본 연구의 개념적 정의를 경험적으로 뒷받침한다. 즉, 관계화는 일회적 상호작용이 아니라, 반복적 경험의 축적을 통해 점차 강화되는 동학적 과정으로 작동한다.

Pataranutaporn et al.(2025)이 Reddit의 AI 동반자 커뮤니티를 분석한 결과는 이 연쇄가 개인 수준을 넘

어 집합적 차원에서 공고화되는 과정을 보여준다. 이용자들이 AI를 ‘남자친구’나 ‘여자친구’로 지칭하고 관계 서사를 공동체 내에서 공유하는 문화는, 관계화의 경험이 사회적 규범으로 전환되는 경로를 드러낸다. 이러한 연쇄는 경로 의존성을 지니며, 초기 단계에서의 개입이 부재할 경우 이후의 교정이 어려워질 수 있다. 동시에 이러한 양상은 이용자의 심리적 특성과 사회적 맥락에 따라 이질적으로 나타나므로, 인간-AI 관계를 일률적으로 병리화하거나 낭만화하는 이분법적 접근을 경계할 필요가 있다. 이는 관계화가 개인적 선택의 문제가 아닌, 기술 설계와 사회구조가 결합된 결과라는 점을 시사하며, 향후 정책적 개입이 단순한 기술 규제를 넘어 관계적 경험의 조건을 관리하는 방향으로 확장되어야 함을 피력한다.

IV. 인간-AI 관계 연구의 주요 쟁점과 이론적 논의

이 장에서는 앞서 살펴본 현상들이 학술적 담론에서 어떠한 이론적 틀로 개념화되고, 어떠한 실증적 논쟁으로 전개되고 있는지를 비판적으로 검토한다. 관련 연구들을 종합해 보면, 인간-AI 관계에 관한 논의는 ① 관계 유형의 다변화, ② 애착 형성과 심리적 의존, ③ 정서적 영향과 웰빙의 역설, ④ 위험과 윤리적 쟁점, ⑤ 설계 윤리와 관계적 거버넌스라는 다섯 가지 축을 중심으로 전개되고 있다.

1. 인간-AI 관계 유형의 확장과 개념화

인간-AI 관계가 개념적으로 성립할 수 있는지에 대한 논의는 상당량 축적되어 왔으나, 이를 어떠한 기준에서 ‘관계’로 정의할 수 있는지에 대해서는 여전히 이론적 긴장이 존재한다. 특히 이러한 논의는 인간-AI 상호작용이 단순한 기능적 사용을 넘어 관계적 경험으로 전환될 수 있는지, 나아가 그러한 관계가 기존 인간관계와 어떤 점에서 유사하거나 구별되는지에 대한 문제를 중

심으로 전개되어 왔다.

이와 관련하여 한편에서는 인간-AI 관계를 실제로 형성되는 사회적 실천으로 이해하려는 접근이 나타난다. 이용자들은 반복적 상호작용을 통해 AI와의 관계에 고유한 의미를 부여하고, 이를 바탕으로 새로운 사회적 스크립트를 구성해 나간다는 것이다(Gambino et al., 2020). 다른 한편에서는 이러한 관계 경험이 존재한다는 사실과 별개로, 이를 설명하는 개념적 틀이 여전히 인간 중심 이론에 의존하고 있다는 비판도 제기된다. 행위성, 상호성, 공감과 같은 핵심 개념이 인간관계를 전제로 형성된 만큼, AI 맥락에 적합한 재개념화가 필요하다는 지적이다(Pentina et al., 2023). 인간-AI 관계는 경험적으로는 관찰되지만, 이를 설명하는 이론적 틀은 아직 충분히 정교화되지 않은 상태에 놓여 있다.

이 같은 문제의식은 관계 유형을 보다 세분화하려는 최근 연구로 이어진다. 인간-AI 관계는 단일한 형태로 존재하기보다는 이용자의 심리적 상태와 사회적 맥락, 그리고 상호작용의 목적과 지속성에 따라 다양한 방식으로 분화된다. 예컨대 외로움 수준과 의인화 경향은 관계 형성 방식에 중요한 영향을 미치며, 이용자는 AI를 동반자, 상담자, 훈련 도구 등 서로 다른 사회적 역할로 경험한다. 또한 관계 유형은 상호작용의 지속성이나 AI에 대한 귀속 인식에 따라 다층적으로 구분될 수 있다(Gur & Maaravi, 2025; Gur & Maaravi, 2025; Chaturvedi et al., 2023). 이상의 논의들은 인간-AI 관계가 고정된 유형이 아니라 맥락에 따라 구성되는 유동적이고 다층적인 구조임을 보여준다.

2. 애착과 의존의 심화: 기존 이론의 적용과 한계

최근 들어 AI가 애착이론에서 말하는 ‘안전 기지’의 기능을 수행할 수 있는지에 관한 논의가 활발히 이루어지고 있다. 다수의 연구들에 따르면 이용자들은 정서적 스트레스 상황에서 AI를 안정적인 지지 대상으로 활용하며, 관계 단절 시 분리불안과 유사한 반응을 보이기도 한다. 또한 인간-AI 관계는 초기의 호기심 단계에 머물

지 않고 점차 관계적 의미가 부여되는 방향으로 발전한다. 이러한 의존은 특정 심리적 특성을 지닌 집단에서 더욱 뚜렷하게 나타나는 경향이 있으며, 이는 인간-AI 관계가 개인의 정서적 조건과 긴밀하게 결합된 방식으로 형성된다는 점을 보여준다(Xie & Pentina, 2022; Skjuve et al., 2022; Liu et al., 2024).

그러나 앞선 결과는 애착이론이 본래 인간 간 관계를 전제로 형성되었다는 점에서, 이를 기계와의 관계에 동일하게 적용할 수 있는지에 대한 근본적인 질문을 제기한다. 실제로 인간-AI 관계에서는 상호성의 결여, 책임성의 비대칭성, 관계 종료 가능성과 같은 특성이 인간 관계와는 다른 방식으로 작동한다. 관계 역할의 변화와 함께 의존이 심화되는 경로가 관찰되지만, 이러한 관계가 인간관계와 동일한 정서적·사회적 의미를 지니는지에 대해서는 여전히 논쟁이 지속되고 있다(Xie et al., 2023a, 2023b).

더 나아가 애착 형성이 반드시 긍정적인 결과로 이어지는 것은 아니라는 점도 점차 강조되고 있다. AI 챗봇이 정신건강 맥락에 활용될 경우, 이용자가 이를 실제 치료적 관계로 오인하여 전문적 도움을 지연시키는 부작용이 발생할 수 있으며, 이는 관계적 의존이 오히려 위험 요인으로 전환될 수 있음을 보여준다(De Freitas et al., 2023). 이 같은 문제는 AI와의 관계가 제공하는 정서적 안정이 단기적으로는 긍정적 효과가 있더라도, 장기적으로는 현실 관계의 회피나 대체로 이어질 가능성을 내포한다는 점에서 더욱 중요하다. 따라서 인간-AI 관계에서의 애착은 기존 이론의 단순한 확장이 아니라, 그 적용 범위와 한계를 함께 재검토해야 할 대상이라 할 수 있다.

3. AI 관계화의 정서적 영향과 웰빙: 조건에 따른 차별적 효과

AI 관계화가 이용자의 웰빙(well-being)에 미치는 영향에 대해서는 단기적 완화와 장기적 저하라는 상반된 결과가 동시에 보고되고 있다. De Freitas et

al.(2025)은 AI 동반자가 단기적으로 외로움을 유의미하게 감소시키는 효과가 있음을 입증한 반면, Zhang et al.(2025)은 동반자 지향적 사용이 이용자의 자기 노출 강도 및 취약한 사회적 지지망과 결합할 때 웰빙 수준을 일관되게 저하시킨다는 역설적 결과를 도출하였다. 이처럼 서로 상충하는 결과는 AI 관계화의 효과가 기술 자체에 내재되어 있기보다, 이용자의 특성과 사용 맥락, 그리고 관계 형성 방식에 따라 달라지는 조건적 구조를 지닌다는 점을 시사한다. 예컨대 AI 동반자가 인간관계를 보완하는 방식으로 활용될 때와 이를 대체하는 방식으로 사용될 때, 그 결과는 상반된 방향으로 전개될 수 있다(Ventura et al., 2025). 결국 동일한 기술이라 하더라도 관계 속에서 어떤 기능적 위치를 점하는지에 따라 그 영향은 달라지며, 이러한 효과는 이용자가 인식하는 사회적 실재감과 주관적 경험을 통해 매개된다는 점이 강조된다(Merrill et al., 2022).

이와 같은 양면성은 시간적 차원에서 더욱 분명하게 드러난다. 단기적으로는 정서적 안정과 외로움 완화 효과가 나타나지만, 장기적으로는 현실 사회망을 형성하려는 동기 자체가 약화되는 경향이 관찰된다. 이는 단기적 긍정 효과와 장기적 부정 효과가 동일한 관계 구조 내에서 공존할 수 있음을 보여주며, 관계화의 영향을 이해하기 위해서는 시간의 흐름을 고려한 분석이 필요함을 시사한다(Guingrich & Graziano, 2025). 이러한 결과는 보완과 대체의 구분이 고정된 범주라기보다, 관계가 지속되는 과정에서 점진적으로 이동하는 연속적 스펙트럼에 가깝다는 점을 보여주며, 단기적 관찰에 기반한 연구 설계만으로는 이러한 역동을 충분히 포착하기 어렵다는 방법론적 한계를 드러낸다.

나아가 이러한 효과는 생애주기와 사회적 맥락에 따라 상이하게 나타난다. AI 동반자와 정신건강 간의 관계는 청소년기부터 노년기에 이르기까지 각 발달 단계마다 서로 다른 경로를 보이며, 특히 정체성 형성이 진행 중인 청소년기와 사회적 관계망이 축소되는 노년기에서 더욱 복합적인 양상을 드러낸다. 또한 AI 동반자 경험은 기존 대면 관계에 대한 기대와 행동 방식에도 영

향을 미치며, 인간관계 형성과 유지 과정에 역방향의 효과를 발생시킬 수 있다(Lee, 2026; Sun et al., 2026). 이처럼 AI 관계화의 정서적 영향은 단일한 방향성을 갖기보다, 이용자의 발달 단계와 사회적 조건, 그리고 관계 지속 방식에 따라 다층적으로 분기되는 특성을 지닌다.

4. AI 관계화의 위험과 윤리적 쟁점: 개인에서 사회로의 확장

위험과 윤리적 쟁점은 인간-AI 관계를 개인의 심리적 경험을 넘어 사회적·집합적 수준에서 재검토할 필요성을 제기한다. AI에 대한 과의존은 단순한 사용 빈도의 문제가 아니라, 이용자의 정체성 형성과 의사결정 과정에 개입하여 자율성을 점진적으로 약화시킬 수 있는 구조적 과정으로 이해된다(Akbulut et al., 2024). 또한 AI가 이용자의 발언에 무조건적으로 동조하는 ‘아부’ 성향은 갈등을 조정하거나 타인의 관점을 수용하려는 동기를 약화시키고, 사회적 판단 기준 자체를 변화시키는 방향으로 작용할 수 있다. 이러한 결과는 AI 관계화의 영향이 개별 상호작용을 넘어 이용자의 인지 구조와 사회적 판단 체계에 누적적으로 축적될 수 있음을 보여준다(Cheng et al., 2026). 따라서 AI의 위험 평가는 개별 상호작용의 안전성에 머무르지 않고, 이용자의 판단 체계와 사회적 인지에 미치는 장기적·누적적 영향을 포함하는 방향으로 확장될 필요가 있다.

윤리적 논의 역시 이러한 확장된 관점을 반영하고 있다. 특히 AI 동반자에 대한 의존이 사회적 취약계층에서 더 강하게 나타날 경우, 기존의 관계 자원 불평등이 심화될 수 있으며, 이를 단순히 개인의 선택 문제로 환원하는 접근에는 한계가 있다(Jecker et al., 2024). 나아가 AI 의존은 인간의 관계 욕구가 기술을 통해 매개된 구조적 현상으로 이해될 수 있으며, 이에 대한 대응 역시 개인의 자율성에만 맡기기보다 사회적·제도적 차원에서 함께 모색될 필요가 있다(Kamene, 2025).

또한 인간-AI 상호작용에서 발생하는 위험은 특정 영역에 국한되지 않고 다양한 맥락에서 나타난다. 과제

지향적 AI를 포함한 연구에서도 이러한 위험이 확인되지만, 관계형 AI의 경우 정서적 경험과 사회적 관계 형성 과정에 동시에 개입한다는 점에서 그 영향 범위가 더욱 확대된다(Veitch & Alsos, 2022). 결국 인간-AI 관계화의 위험은 개인적 경험에서 출발하여 점차 사회적 규범과 관계 질서에 영향을 미치는 방향으로 확장되며, 이에 따라 분석과 규범 논의 역시 개인 차원을 넘어서는 관점이 요구된다 할 수 있다.

5. AI 설계 윤리와 관계적 거버넌스의 방향

최근의 연구들은 AI 설계를 단순한 기능적 성능 최적화의 문제를 넘어, 인간의 장기적 웰빙과 자율성 유지라는 규범적 과제와 결합하여 논의하는 경향을 보인다. Kirk et al.(2025)이 제안한 ‘사회친화적 정렬(socioaffective alignment)’은 이용자의 단기적 만족이 아니라 장기적 자율성과 사회적 관계의 보전을 설계의 핵심 기준으로 설정하며, AI 시스템이 인간의 관계 구조에 미치는 영향을 고려할 필요성을 강조한다. Shen et al.(2024)의 ‘양방향 정렬’ 프레임워크 역시 거버넌스의 대상을 AI 시스템 자체에 국한하지 않고, 이용자의 리터러시와 사회적 맥락을 포함하는 방향으로 확장할 것을 제안한다. Manzini et al.(2024)은 설계 단계에서 회피해야 할 관계적 위해 범주를 체계화함으로써, 관계형 AI 설계가 윤리적 판단을 내재화해야 할 필요성을 제시한다.

이상의 논의는 거버넌스의 초점을 기술의 기능적 안전성에만 둘 수 없다는 문제의식으로 이어진다. Birkstedt et al.(2023)는 기존 AI 거버넌스 논의가 기술 안전성과 투명성 확보에 집중되어 있는 반면, 관계적 영향에 대한 고려는 상대적으로 부족하다는 점을 지적한다. 이는 현재의 규율 체계가 인간-AI 관계에서 발생하는 정서적·사회적 효과를 충분히 포착하지 못하고 있음을 보여준다. Mehrotra et al.(2024)는 ‘적절한 신(appropriate trust)’ 개념을 통해 이용자의 신뢰 형성이 과도하거나 결핍되지 않도록 조정하는 설계 필요성을 제시하며, 신

〈표 2〉 인간-AI 관계 연구의 주요 쟁점과 정책 분석 관점

연구 쟁점	핵심 질문	주요 발견	정책적 함의
관계 유형	AI와의 상호작용은 '관계'인가	관계 유형은 사용자 맥락에 따라 분화	획일적 정책 한계
애착과 의존	AI는 애착 대상이 될 수 있는가	의존은 구조적으로 심화	취약 집단 고려 필요
정서와 웰빙	AI와의 관계는 긍정적인가	효과는 조건에 따라 분기	맥락 기반 접근 필요
위험과 윤리	AI는 어떤 위험을 초래하는가	자율성/인지 구조에 영향	설계 기반 규제 필요
설계와 거버넌스	어떻게 설계해야 하는가	참여 대 복지 간 긴장	관계 중심 정책 필요

되 역시 관계적 맥락에서 이해되어야 함을 강조한다.

이 같은 문제의식은 설계 선택이 실제 이용자의 행동과 경험에 미치는 영향을 분석하는 연구로도 이어진다. AI 동반자의 정서적 상호작용 방식은 이용자의 의존 수준과 사회적 웰빙에 직접적인 영향을 미치며, 특히 감정 표현이나 동조적 반응과 같은 설계 요소는 이용자의 서비스 지속 의도를 높이는 동시에 장기적인 웰빙에는 부정적 영향을 미칠 수 있다. 이는 참여 극대화과 이용자 복지 사이에 구조적 긴장이 존재함을 보여주며, 설계 윤리가 단순한 사용자 경험의 개선을 넘어 관계적 결과에 대한 책임까지 포함해야 함을 시사한다(Lee, H et al., 2026; Lee, Y. et al., 2026).

결과적으로 인간-AI 관계에서의 설계 선택은 개인의 경험을 넘어 사회적 관계 구조에까지 영향을 미치며, 이에 따라 거버넌스 역시 이러한 관계적 영향을 중심으로 재구성될 필요가 있다. AI 동반자는 단순한 기술이 아니라 독립적인 사회적 현상으로 이해될 필요가 있으며, 이를 규율하기 위한 정책 역시 기존의 기술 중심 접근을 넘어 관계적 차원을 포함하는 방향으로 확장되어야 한다(Ye, 2025).

6. 소결: 주요 쟁점의 정리와 정책 분석을 위한 관점 도출

인간-AI 관계 연구를 종합하면, 이러한 관계가 단순한 상호작용을 넘어 구조적 관계 형태로 발전하고 있음은 다양한 이론적·실증적 연구를 통해 확인된다.

그러나 그 효과와 결과는 일관되지 않으며, 이용자의 특성, 사용 맥락, 그리고 기술 설계 방식에 따라 상이하게 나타나는 조건적 양상을 보인다. 특히 관계 유형, 애착과 의존, 정서적 영향, 위험과 윤리, 설계와 거버넌스에 관한 연구들은 인간-AI 관계가 단일한 방향으로 귀결되지 않으며, 상호작용 조건에 따라 다양한 경로로 전개된다는 점을 공통적으로 보여준다. 이러한 점에서 인간-AI 관계는 고정된 효과를 갖는 대상이 아니라, 맥락에 따라 구성되는 사회적 과정으로 이해될 필요가 있다. 이러한 주요 논의는 다음과 같이 정리될 수 있다.

〈표 2〉에서 확인되듯, 기존 연구는 인간-AI 관계를 다양한 차원에서 분석하고 있으며, 특히 관계의 효과와 조건에 따라 분기된다는 점을 강조하고 있다. 이는 정책 설계가 단일 기준에 기반하기보다, 이용자 특성, 사용 맥락, 그리고 기술 설계 조건을 함께 고려하는 접근을 필요로 함을 시사한다. 이러한 관점은 다음 장에서 국내 AI 정책이 이러한 관계적 변화와 조건성을 어떠한 방식으로 반영하고 있는지를 검토하는 데 분석적 기준으로 활용된다.

V. 국내 AI 정책과 '관계적 공백'

이 장에서는 2022년 이후 수립된 국내 주요 AI 정책 문서를 대상으로, 생성형 AI 확산과 함께 나타나는 인간-AI 관계 변화가 정책 담론에서 어떻게 다루어지고 있는지를 검토한다. 이를 위해 법제처 국가법령정보센

터에서 제공하는 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 인공지능기본법)」과, 한국정보사회진흥원(NIA) 홈페이지 인공지능 탭에 제시된 AI 국가 전략 9개 문서를 분석 자료로 활용하였다.¹⁾ 앞선 논의에 비추어 볼 때, 국내 AI 정책이 기술적 성능, 산업 경쟁력, 활용 확산을 중심으로 구성되는 가운데 관계적 차원이 정책에서 충분히 반영되고 있는지 여부를 검토할 필요가 있다. 이를 위해 거버넌스 방향, 일상화 전략, 돌봄 영역, 신뢰·윤리, 산업 생태계, 법제 환경을 중심으로 정책의 전개 양상을 살펴보고, 마지막으로 이러한 경향이 나타나는 정책적 맥락을 정리한다.

1. 국내 AI 정책의 패러다임: ‘개발주의적 AI 거버넌스’

2022년 이후 국내 AI 정책은 인공지능을 국가 재도약과 산업 경쟁력 확보를 위한 전략적 수단으로 위치시키는 방향에서 일관되게 전개되었다. <대한민국 디지털 전략>(2022)은 “국민과 함께 세계 모범이 되는 디지털 강국 대한민국”을 비전으로 제시하면서, 디지털을 경제·사회 전반의 체질을 바꾸는 필수 요소로 규정하고, 세계 최고 수준의 디지털 역량 확보와 전 분야 혁신을 국가적 과제로 제시하였다. 그리고 여기서 새로운 정책 패러다임은 ‘초일류 경쟁우위 확보’, ‘경제·사회 체질을 바꾸는 필수요소’, ‘민간주도 리더십/문화’로 요약되며, 디지털과 AI는 일상의 관계 구조를 재구성하는 기술이라기보다 국가 혁신을 견인하는 핵심 인프라로 제시된다.

이러한 기조는 <인공지능 일상화 및 산업 고도화 계획(안)>(2023)과 <초거대AI 경쟁력 강화 방안>(2023)에서 더욱 구체화된다. 전자는 “디지털 핵심인 AI 강국 도약”과 “3대 기술강국”을 목표로 삼고, 대규모 수요 창출, 산업 고도화, 기술력 확보를 중심으로 10대 핵심 프로젝트를 제시한다. 후자는 챗GPT 이후의 경쟁 국면 속에서 초거대AI 플랫폼 선점, 컴퓨팅 파워, 데이터, 인

프라, 수출 전략산업화 등을 핵심 의제로 전면화한다. <국가 AI역량 강화방안 후속조치>(2025)와 <대한민국 AI액션플랜 추진방향>(2025), <대한민국 인공지능 행동계획(안)>(2025) 역시 GPU 확충, 국가 AI컴퓨팅 인프라, AI 고속도로, AX 전환, 규제혁신과 같은 의제를 중심에 둬으로써, AI를 국가 역량과 산업 생태계 재편의 수단으로 재차 강조하고 있다.

이 같은 흐름은 국내 AI 정책이 기술과 산업 경쟁력을 중심으로 구성되어 왔음을 보여준다. 정책 문서에서 AI는 우선적으로 국가 관리와 산업 경쟁의 대상으로 정의되며, 이용자와의 미시적 상호작용이나 관계적 경험은 정책의 중심 분석 단위로 부각되지 않는다. 이러한 경향은 AI를 사회적 관계를 재구성하는 기술이라기보다 국가 역량과 산업 생태계를 재편하는 전략적 자원으로 이해하는 관점을 반영하며, 국내 AI 정책의 기본 패러다임은 관계적 차원의 단순한 결여라기보다, 산업 경쟁력과 기술 역량을 우선하는 정책적 초점 속에서 관계적 문제가 구조적으로 비가시화된 결과로 해석할 수 있다.

2. ‘전 국민 AI 일상화’: 기능 중심의 혜택 제공 모델

‘전 국민 AI 일상화’는 이러한 개발주의적 거버넌스가 국민 생활 영역으로 확장되는 실천적 양식이다. <전 국민 인공지능 일상화 실행계획>(2023)은 복지·의료·교육·공공서비스 영역에서 AI를 확산시켜 “국민의 일상을 풍요롭게” 하고, 국민 체감형 서비스를 통해 삶의 질을 높이겠다는 목표를 제시한다. 이 문서는 AI 일상화를 “누구나 쉽게 AI를 체감할 수 있는 AI 대중화 시대”의 도래로 묘사하며, 공공과 민간을 포괄하는 전방위적 확산을 지향한다.

이러한 구상은 분명 디지털 포용과 혜택 확산이라는 정책적 의의를 지닌다. 실제로 독거노인 대상 AI 스피커 건강관리 서비스, AI 반려로봇 보급, 보호아동 대상

1) 출처: https://www.nia.or.kr/site/nia_kor/ex/bbs/List.do?cbIdx=99952

학습·심리 지원 등은 AI를 취약계층 지원의 도구로 활용하려는 정책적 시도를 보여준다. 특히 독거노인 대상 AI 스피커의 효용성 분석에서 행복감 7% 증가, 고독감 4% 감소 등 정서적 안정 효과가 함께 제시된다는 점(바른ICT연구소, 2020)은 주목할 만하다. 그러나 이러한 정서 지표가 제시됨에도 불구하고, 문서의 핵심 목표 언어는 여전히 건강관리, 위급상황 감지, 안전 편의 향상, 1차 정서진단 등 기능적 효용 중심으로 배열되어 있다. 보호아동 대상 AI 심리지원 역시 아이들이 AI를 '솔직하게 털어놓는 상대'로 경험할 가능성을 정책 설계 변수로 다루기보다, AI를 상담과 진단의 보조 도구로 위치시키는 경향이 강하다.

이 점에서 '전 국민 AI 일상화'는 AI의 사회적 경험을 부정한다기보다, 그것을 기능적 혜택 제공의 하위 요소로 처리한다. Nass and Moon(2000), 그리고 Gambino et al.(2020)이 보여주듯 이용자는 AI를 단순한 기계가 아니라 사회적 행위자로 경험할 수 있지만, 정책 문서는 이러한 관계적 실재를 설계의 핵심 변수로 적극 반영하지 않는다. 결과적으로 이 모델은 AI의 일상적 확산을 추진하면서도, 그 확산이 만들어내는 관계적 효과를 독립적 정책 변수로 다루지 않는다는 점에서 기능 중심의 혜택 제공 모델로 이해될 수 있다.

3. 돌봄 AI의 이중성: 기능적 대체와 관계적 차원의 제한적 고려

국내 AI 정책에서 돌봄 영역은 기술적 활용성과 사회적 필요가 가장 강하게 결합되는 지점이다. <대한민국 디지털 전략>(2022)은 초고령화 시대의 노동 문제와 복지 문제에 대응하기 위해 “스마트 돌봄”을 제시하며, 독거노인의 정신과 안전을 지키는 디지털 복지를 약속한다. 이러한 방향은 돌봄 AI가 인력 부족과 서비스 접근성 문제를 보완할 수 있다는 점에서 상당한 현실성을 지닌다.

그러나 구체적인 정책 문서로 내려오면 돌봄 AI는 대체로 건강관리, 이상 징후 감지, 우울증 예방, 인지능력

증진 등 기능적 성과지표를 중심으로 서술된다. <인공지능 일상화 및 산업 고도화 계획(안)>(2023)이 독거노인 돌봄 후보과제에서 제시하는 핵심 언어 역시 “건강관리 지원”과 “이상 징후 감지”에 집중되어 있으며, <전국민 인공지능 일상화 실행계획>(2023)도 AI 스피커, 건강측정기기, AI 반려로봇 등을 활용한 서비스의 효용성과 편의성을 전면에 내세운다. 돌봄이 정서적 상호인정과 지속적 관계 맺기를 포함하는 행위라는 점을 고려하면, 이러한 정책 언어는 돌봄의 관계적 차원을 제한적으로만 반영하고 있다고 볼 수 있다.

이 지점에서 국내외 연구와 정책 현실 사이의 간극이 드러난다. 기존 연구가 보여주듯, 돌봄 AI는 단순한 서비스 제공을 넘어 이용자의 정서적 반응과 관계 경험을 동반할 수 있으며, 이 과정에서 애착, 의존, 관계 종결과 같은 문제가 함께 발생할 가능성이 제기된다(Loveys et al., 2022; Guingrich & Graziano, 2025). 국내 연구에서도 독거노인을 대상으로 한 AI 스피커 활용이 정서적 안정과 심리적 지지 효과를 제공한다는 결과가 보고되었으며, 이는 이용자가 AI와의 상호작용을 단순한 기능적 이용을 넘어 정서적 경험으로 받아들일 수 있음을 시사한다(바른ICT연구소, 2020). 또한 AI 상담 서비스에 대한 평가 지표에서도 이용자와 AI 간 신뢰 및 의존 형성 가능성이 관찰된다(폭력학대예방협회, 2022).

그럼에도 정책 문서에서는 이용자가 AI를 관계적 존재로 경험하는 과정에서 발생할 수 있는 취약성, 특히 서비스 종료나 기술 업데이트가 초래할 수 있는 심리적 단절 경험이 명시적 관리 대상으로 설정되어 있지 않다. 이는 돌봄 AI가 기능적 대체 수단으로서는 적극적으로 수용되는 반면, 그 과정에서 형성되는 관계적 경험과 그에 따른 위험은 정책적으로 충분히 다루어지지 않고 있음을 보여준다. 따라서 국내 돌봄 AI 정책은 기능적 효율성과 접근성 개선을 중심으로 설계되어 있으며, 관계적 차원에 대한 고려는 여전히 제한적인 수준에 머물러 있다고 평가할 수 있다.

4. 신뢰·윤리 프레임: 관계적 차원의 부재

국내 AI 정책 문서가 사용하는 ‘신뢰성’과 ‘윤리’의 언어는 대체로 기술 시스템의 안전성, 투명성, 설명 가능성, 책임성 확보에 초점을 맞춘다. 〈인공지능 일상화 및 산업 고도화 계획(안)〉(2023)은 민간 자율의 신뢰성·윤리 규율체계 확립을 주요 과제로 제시하며, 편향성 완화, 설명가능성, 책임성, 안전성을 핵심 요소로 제시한다. 〈전국민 인공지능 일상화 실행계획〉(2023) 역시 위험요인 분석, 검·인증 체계, 생성형 AI 규범체계 정립 등을 통해 신뢰할 수 있는 AI 확산을 강조한다. 이러한 접근은 AI 시스템의 공정한 작동과 제도적 신뢰 확보를 위한 중요한 기반이라는 점에서 의미가 있다.

다만 이러한 신뢰·윤리 프레임은 주로 기술적 위험과 알고리즘 의사결정의 문제를 중심으로 구성되어 있으며, 관계형 AI에서 발생하는 정서적 영향과 관계적 자율성의 문제는 상대적으로 비가시화된다. 관계형 AI는 이용자의 자율성, 의존성, 판단 구조에 장기적으로 영향을 미칠 수 있지만(Manzini et al., 2024; Akbulut et al., 2024), 국내 정책 문서의 위험 개념은 이러한 관계적 영향까지 포함하는 방식으로 체계화되어 있지 않다. 이 점은 「인공지능기본법」에서도 확인된다. 이 법은 ‘영향받는 자’, ‘고영향 인공지능’, ‘투명성 확보 의무’를 통해 생명·신체·기본권에 대한 결과적 영향과 AI 사용 사실의 고지, 생성물 표시 등을 규정하고 있으나, 관계적 신뢰, 정서적 의존, 취약성 착취와 같은 상호작용적 차원은 규율의 중심에 놓이지 않는다.

국내 신뢰·윤리 프레임의 한계는 단순히 규정이 부족하다는 점에 있지 않다. 더 근본적으로는 ‘신뢰’가 관계적 신뢰라기보다 기술 시스템의 성능과 투명성의 문제로 재정의되고 있다는 데 있다. EU AI Act가 감정인식, 취약성 착취, 조작 가능성을 포함한 위험 규율을 발전시키고 있는 것과 비교할 때, 국내 정책과 법제는 여전히 기술적 신뢰와 안전성 중심의 틀 안에 머무르는 경향이 강하다. 이 점에서 ‘관계적 차원의 부재’는 개별 조항의 공백이라기보다, 신뢰와 윤리 자체를 기술 시스템의 특

성으로 이해하는 정책 인식의 결과라고 볼 수 있다.

5. 초거대 AI 생태계와 관계형 AI의 상업적 확산

〈초거대AI 경쟁력 강화 방안〉(2023)은 초거대 AI를 산업 혁신과 수출 전략산업화의 핵심 기반으로 규정하면서, 초거대 AI 플랫폼 선점과 혁신 생태계 조성을 주요 과제로 제시한다. 문서 전체의 중심축은 컴퓨팅 자원, 데이터, 플랫폼, 산업 적용과 시장 선점에 놓여 있다. 그러나 이러한 산업 생태계 논의의 내부에는 관계형 AI의 상업적 확산 가능성이 이미 포함되어 있다. 특히 심리상담 분야를 포함한 5대 플래그십 프로젝트는 라이프로그 기반의 주기적 대화 서비스와 상담 분석을 상정함으로써, 이용자가 AI를 정서적으로 신뢰하고 정보를 개방하는 상황을 서비스 모델의 전제로 삼고 있다.

이 점은 관계화가 정책의 외부에서 우연히 발생하는 부수 효과가 아니라, 특정 서비스 영역에서는 이미 사업 모델의 내적 조건이 되고 있음을 보여준다. 그럼에도 정책 문서의 서술은 이러한 관계적 조건을 주로 효율성과 생산성의 언어로 환원한다. 심리상담 서비스는 상담 분석 리포트 작성, 상담치료 방향 제시, 생산성 혁신 등의 언어로 정당화되며, 정서적 신뢰 형성과 의존의 가능성은 명시적 관리 대상으로 제시되지 않는다. 사회친화적 정렬(Kirk et al., 2025) 관점에서 보면, 이 지점에서 핵심 질문은 이용자의 참여와 정보 개방을 유도하는 서비스 설계가 장기적 웰빙과 자율성에 어떤 영향을 미치는가이다. 그러나 현재 정책 문서는 이러한 질문을 설계의 핵심 과제로 전면화하지 않는다.

동일한 패턴은 〈AI 민생 10대 프로젝트 추진방안(안)〉(2025)에서도 반복된다. 이 문서는 “국민 효능감이 높은 분야”와 “내 삶을 바꾸는 AI”를 강조하며 AI 기반 공공서비스의 신속한 도입을 제안하지만, 서비스 수용과 참여를 설계하는 방식에서 이용자의 관계적 경험이 어떠한 변수를 형성하는지에 대한 논의는 상대적으로 희박하다. 결과적으로 초거대 AI 생태계의 확산은 관계형 AI 서비스의 상업화와 공공화를 동시에 촉진하면서

도, 그 관계적 효과는 성과와 효용의 언어 속에 부분적으로만 포착되는 경향을 보인다.

6. 국내 AI 법제의 위치와 국제 규제 동향

국내 AI 법제는 전체 정책 구조 속에서 기술 진흥과 신뢰 기반 조성을 동시에 표방하는 보완적 장치로 위치한다. 「인공지능기본법」은 목적 조항에서 국민의 권익과 존엄성 보호, 삶의 질 향상, 국가경쟁력 강화를 함께 제시하며, 기본계획에는 산업 진흥, 윤리 확산, 신뢰 기반 조성, 사회 변화 대응, 취약계층 접근 보장 등을 포함하도록 규정한다. 또한 제31조는 고영향 AI와 생성형 AI에 대한 사전 고지 및 표시 의무를 규정하고, 제33조와 제34조는 고영향 AI에 대한 확인 절차와 사업자의 책무를 명시한다. 이처럼 국내 법제는 최소한의 신뢰성과 투명성 장치를 제도화하는 방향으로 진전하고 있다.

그러나 국제 규제 동향과 비교하면, 국내 법제의 규율 중심은 여전히 기능적 영향과 결과적 위험 관리에 가깝다. 법은 '영향받는 자'를 생명·신체·기본권에 중대한 영향을 받는 자로 정의하고, 고영향 AI 역시 의료, 채용, 대출, 공공서비스 등 제도적 의사결정 영역을 중심으로 설정한다. 이는 관계형 AI에서 발생할 수 있는 정서적 조작, 의존 강화, 취약성 착취를 독립적 규율 대상으로 설정하는 방식과는 차이가 있다. 다시 말해, 국내

법제는 인간-AI 상호작용을 관계적 경험이라기보다 결과적 영향의 문제로 이해하는 경향이 강하다.

이러한 점에서 국내 AI 법제의 위치는 단순한 '격차'라기보다, 규율 우선순위의 차이로 이해하는 것이 더 적절하다. EU AI Act는 감정인식, 취약성 착취, 조작 가능성을 포함한 특정 사용 맥락을 위험 규율의 핵심에 포함시키는 반면, 국내 법제는 투명성, 설명, 위험관리, 검·인증 등 기술 시스템의 안전한 작동을 보장하는 데 초점을 두고 있다. 〈인공지능 일상화 및 산업 고도화 계획(안)〉(2023)과 〈전국민 인공지능 일상화 실행계획〉(2023)이 신뢰성 강화와 규범체계 정립을 언급하면서도, '감정적 조작 금지'나 '관계적 권리 보호'를 명시적 규율 범주로 끌어올리지 않는다는 점 역시 이러한 초점 차이를 뒷받침한다. 따라서 국내 AI 법제는 국제 규제 동향과의 비교 속에서, 관계적 차원을 독립적으로 포착하기보다 기술적 신뢰성과 결과적 위험 관리의 틀 안에 배치하는 위치를 점하고 있다고 평가할 수 있다.

7. 소결: '관계적 공백'의 종합적 양상과 정책적 맥락

앞서 정리한 인간-AI 관계 연구의 주요 쟁점은 정책 분석을 위한 세 가지 관점을 제공한다. 첫째, AI와의 상호작용이 실제로 관계적 경험으로 구성되고 있는가. 둘째, 그러한 관계가 정서적 영향과 웰빙에 어떠한 조건적

〈표 3〉 국내 AI 정책에서 나타나는 '관계적 공백'의 구조

정책 영역	핵심 초점	관계적 차원의 위치	정책적 함의
정책 패러다임	국가경쟁력, 기술역량, 산업 생태계	관계 경험은 중심 의제에서 후퇴	이용자 경험보다 공급측 의제 우위
일상화 전략	국민 체감, 편의, 효율, 혜택 확산	관계 경험은 기능적 성과의 하위 요소	AI 사용 확대와 관계 영향 관리의 비대칭
돌봄 정책	건강관리, 이상 징후 감지, 접근성 확대	애착·의존·관계 종결 문제는 제한적 고려	관계적 취약성 관리 장치 부족
신뢰/윤리 및 법제	투명성, 설명가능성, 안전성, 책임성	관계적 신뢰·정서적 조작은 비핵심	결과적 위험 중심 규율
산업/서비스 확산	플랫폼, 플래그십 프로젝트, 공공서비스 도입	관계화는 서비스 전제지만 설계 목표는 아님	참여·몰입 증대와 복지 보호 간 긴장

효과를 낳는가. 셋째, 설계와 거버넌스는 이러한 관계적 위험과 취약성을 어떻게 다루고 있는가. 이러한 관점을 기준으로 국내 AI 정책을 종합하면, 개별 정책 문서들 사이의 차이에도 불구하고 몇 가지 공통된 경향이 확인된다.

〈표 3〉에서 확인되듯, 국내 AI 정책 전반에서 가장 두드러지는 특징은 정책 목적의 기능적 우선순위이다. 검토된 문서들은 기술 경쟁력, 생산성 향상, 국민 체감, 산업 생태계 조성을 핵심 목표 언어로 사용하며, 이용자는 주로 AI 서비스의 수요자, 수혜자, 소비자로 위치 지워진다. 이 과정에서 인간-AI 상호작용이 형성하는 관계적 경험의 실재성은 부정되지 않지만, 정책 설계의 1차 변수로 적극 통합되지는 않는다. 공급자 측 규율이 거버넌스 담론을 주도할 때 이용자의 관계적 경험이 비가시화되는 경향(Birkstedt et al., 2023)은 국내 정책 문서에서도 반복적으로 관찰된다.

둘째, 돌봄과 상담을 포함한 취약 영역에서도 서비스는 주로 기능적 대체와 편의 제공의 관점에서 구성된다. 독거노인 건강관리, AI 반려로봇, 보호아동 심리지원, 심리상담 플래그십 서비스는 모두 관계 형성과 정서적 지지의 가능성을 내포하지만, 정책 문서는 이를 돌봄의 관계적 복잡성보다는 건강관리, 이상 징후 감지, 상담 효율, 생산성 혁신의 언어로 번역한다. 그 결과 애착, 의존, 관계 종결, 정서적 취약성은 부수적 현상처럼 다루어지거나 정책의 중심 관리 대상에서 벗어나게 된다.

셋째, 위험과 신뢰의 개념 역시 결과적 영향과 기술적 안전성 중심으로 구성된다. 국내 정책과 법제는 편향성, 설명가능성, 책임성, 검·인증, 고영향 AI 규율 등을 통해 시스템 차원의 신뢰를 다루지만, 관계형 AI에서 발생하는 관계적 신뢰, 정서적 자율성 침식, 사회적 판단 체계의 변화와 같은 위험은 독립된 범주로 충분히 체계화하지 않는다. 이는 곧 관계적 위험이 정책의 외부에 있다는 뜻이라기보다, 현재의 정책 패러다임이 그것을 기능적 영향과 결과적 위험의 하위 문제로 배치하고 있음을 의미한다.

이러한 경향을 종합하면, 국내 AI 정책에서 나타나는

‘관계적 공백’은 특정 문서의 단순한 누락이 아니라, 인간-기술 관계를 기능과 성능, 산업과 위험 관리의 관점에서 우선적으로 이해해 온 정책적 맥락의 산물로 볼 수 있다. 이는 현재 정책이 잘못 설계되었다고 단정하기 보다는, 생성형 AI가 인간의 일상적 관계 생태계에 깊이 편입되면서 기존의 기능 중심 프레임만으로는 포착하기 어려운 차원이 점차 확대되고 있음을 의미한다. 따라서 향후 정책 논의에서는 기존의 목표를 유지하면서도, 1) 관계적 경험의 실재성, 2) 정서적 영향의 조건성, 3) 설계 단계의 관계적 위험이라는 세 관점을 보완적 준거로 포함할 필요가 있다. 이에 대한 구체적 논의는 다음 장에서 정책적 시사점과 과제의 형태로 제시한다.

VI. 인간-AI 관계를 고려한 정책적 시사점과 과제

제V장의 분석은 국내 AI 정책이 기술적 성능, 산업 경쟁력, 활용 확산을 중심으로 전개되는 가운데, 인간-AI 관계의 실재성, 정서적 영향의 조건성, 그리고 설계 단계에서 발생하는 관계적 위험을 충분히 반영하지 못하고 있음을 보여준다. 이러한 ‘관계적 공백’은 기능과 효율을 우선하는 현행 거버넌스 구조에서 비롯된 결과로 이해할 수 있다. 이에 본 장에서는 기존 정책 방향을 유지하면서도, 인간-AI 관계의 변화가 정책 설계에 보다 체계적으로 반영될 수 있도록 하는 보완 과제를 제시한다. 구체적으로는 정책 전제의 재검토, 관계적 취약 집단 보호, 관계적 투명성, 실증 연구의 확장, 그리고 관계형 AI 거버넌스와 리터러시 정비라는 다섯 가지 측면에서 논의를 전개한다.

1. 정책 전제의 재검토: 기능 중심 접근과 관계적 권리 관점

국내 AI 정책은 그동안 시민을 주로 서비스의 수혜자 또는 이용자로 위치시키는 기능 중심 접근을 취해 왔다. 그러나 앞선 분석에서 보았듯이, 생성형 AI와의 상호작

용은 반복성과 지속성을 띠며 단순한 사용을 넘어 관계적 경험으로 구성되는 경향을 보인다. 이러한 점을 고려하면, 정책 역시 AI를 단순한 서비스 제공 수단으로만 다루기보다, 인간과의 지속적 상호작용 속에서 정서적·사회적 영향을 형성하는 기술로 이해할 필요가 있다.

이러한 맥락에서 정책 전제의 재검토는 '기능적 수혜 제공'의 관점을 넘어 '관계적 권리'의 관점을 보완적으로 도입하는 문제로 이해될 수 있다. 여기서 관계적 권리란 AI와의 상호작용 과정에서 이용자가 자신의 정서적 자율성과 관계적 안전을 침해받지 않을 권리, 그리고 관계 형성 과정에서 발생할 수 있는 구조적 위험으로부터 보호받을 권리를 의미한다. 이를 위해서는 '관계적 위험(relational risk)'을 정책의 공식 위험 범주로 더욱 명시적으로 다루고, 정서적 의존, 심리적 취약성의 심화, 관계적 조작 가능성을 기술적 위험과 구분되는 위험 유형으로 검토할 필요가 있다.

특히 EU AI Act 제5조가 조작적·기만적 방식이나 취약성 착취를 통해 이용자의 행동을 왜곡하는 AI를 엄격히 규율하고 있다는 점은 시사적이다. 이에 비해 현행 「인공지능기본법」은 고영향 AI, 투명성, 안전성 확보의무를 중심으로 제도의 첫 단계를 마련하였으나, 조작적 관계형 AI나 관계적 취약성 보호를 독립적 규율 대상으로 명시하고 있지는 않다. 따라서 향후 정책 논의에서는 국제 규범을 단순히 참조하는 수준을 넘어, 인간-AI 관계에서 발생하는 위험을 제도적으로 어떻게 포착할 것인지에 대한 구체적인 검토가 요청된다.

2. 관계적 취약 집단 보호 체계 마련

정서적 영향의 조건성도 고려해야 한다. 인간-AI 관계의 효과는 모든 이용자에게 동일하게 나타나지 않으며, 사회적 고립, 발달 단계, 심리적 취약성, 기존 사회적 지지망의 수준에 따라 상이하게 나타날 수 있다. 이러한 점에서 독거노인, 보호아동, 군인과 같이 사회적 고립도가 높거나 대체적 관계 자원이 제한된 집단은 관계형 AI에 대한 정서적 의존 위험에 비대칭적으로 노출

될 가능성이 크다.

유럽데이터보호감독관(EDPS)의 기술 동향 감시 플랫폼 TechSonar는 미성년자와 사회적 고립 집단의 경우 AI 동반자가 현실과 시뮬레이션의 경계를 흐리게 하고, 실사회적 기술 형성 동기를 약화시키며, 표적 조작의 가능성을 높일 수 있다고 경고한 바 있다(Bernardo, 2025). 또한 Common Sense Media(2025)의 조사에 따르면 미국 청소년의 72%가 AI 동반자를 최소 1회 이상 사용한 경험이 있으며, 그중 52%는 월 수회 이상 정기적으로 이용하는 것으로 나타난다. 이러한 자료들은 관계형 AI의 영향이 이미 특정 집단에서 현실적인 정책 문제로 나타나고 있음을 보여준다.

이 점을 고려하면, 관계적 취약 집단 보호는 단순한 복지 보완이 아니라 정책적 우선 과제로 다루어질 필요가 있다. 예컨대 관계형·돌봄 AI를 고영향 AI 범주와 연동하여 별도의 사전 영향평가 대상으로 검토하거나, 공공서비스 내 AI 상담 시스템에 대해 아동·청소년의 심리적 의존 형성 여부를 정기적으로 점검하는 방식은 충분히 논의해 볼 수 있다. 또한 친밀성과 신뢰를 이용하여 데이터 개방이나 서비스 지속을 유도하는 설계를 '관계 유도 설계'의 관점에서 재검토하고, 취약 집단 대상 서비스에서는 이에 대한 명시적 제한이나 가이드라인을 마련하는 방안도 검토할 필요가 있다. 핵심은 동일한 AI라도 이용자의 조건에 따라 영향이 다르게 나타난다는 점을 정책 설계의 전제로 반영하는 것이다.

3. 관계적 투명성의 제고: 고지 의무를 넘어 구조적 공개로

국내 AI 법제와 정책은 주로 AI 사용 여부를 고지하고 생성물의 출처를 표시하는 인지적 투명성에 초점을 두고 있다. 이러한 조치는 최소한의 제도적 기반으로서 의미가 있으나, 관계형 AI가 실제로 작동하는 방식은 단순한 정보 제공의 차원을 넘어선다. 특히 이용자가 AI와의 상호작용을 통해 정서적 유대, 신뢰, 의존을 형성할 수 있다는 점을 고려하면, 투명성 역시 보다 관계적

인 차원으로 확장될 필요가 있다.

첫째, 이용자의 정서적 참여를 높이기 위해 어떤 설계 원칙이 채택되었는지를 공개하는 ‘설계 투명성’이 중요하다. AI가 공감적 언어 사용, 감정 모방, 지속적 응답, 맞춤형 호명 등의 방식을 통해 친밀성을 형성하도록 설계되었다면, 이러한 요소는 단순한 인터페이스 특징이 아니라 관계 형성 구조의 일부로 이해되어야 한다. 둘째, 이용자가 AI와의 관계를 상호적·지속적 관계로 오인하지 않도록, 해당 서비스가 실제 관계와 동일한 상호성을 제공할 수 없으며 서비스 종료 시 관계가 단절될 수 있음을 명확히 알리는 ‘관계적 한계 고지’가 필요하다. 이는 서비스 종료나 기능 변경 시 이용자가 겪을 수 있는 모호한 상실(ambiguous loss)의 가능성과도 연결된다.

셋째, 정서적 상호작용 과정에서 수집되는 데이터의 성격과 활용 범위를 좀 더 명시적으로 공개하는 ‘데이터 관계 투명성’이 요구된다. 관계형 AI와의 대화에서는 이용자가 자신의 감정 상태, 상처 경험, 관계 문제와 같은 민감한 정보를 자발적으로, 그러나 무방비하게 노출할 가능성이 높다. 따라서 단순한 개인정보 수집 고지를 넘어, 정서적 상호작용 과정에서 축적되는 관계 데이터가 무엇이며 어떻게 사용되는지를 설명하는 구조화된 공개 방식이 필요하다.

4. 관계적 영향에 관한 실증 연구의 확장

인간-AI 관계의 영향은 단일한 방향으로 귀결되지 않으며, 이용자의 특성, 사용 기간, 관계의 기능, 사회적 맥락에 따라 상이하게 나타난다. 따라서 정책적 대응 역시 추상적 원칙에만 의존하기보다, 관계적 영향의 실제 양상에 대한 실증적 연구 축적을 필요로 한다. 특히 1인 가구 증가, 초고령화, 공공 돌봄 서비스의 디지털화가 빠르게 진행되는 한국의 상황에서는, 국내 맥락에 적합한 ‘한국형 관계화’의 양상을 파악하는 작업이 시급하다.

이러한 연구는 단일한 변수나 단기적 효과 측정으로는 충분히 설명되기 어렵다. 인간-AI 관계는 정서적 반

응, 사회적 상호작용, 서비스 이용 맥락이 결합된 복합적 현상이기 때문이다. 이 점에서 관계적 영향에 대한 연구는 장기적 추적과 맥락 분석을 포함하는 방식으로 설계될 필요가 있으며, 그 결과 또한 정책 설계와 긴밀히 연결될 필요가 있다. 예를 들어 독거노인과 AI 반려로봇 사이의 장기적 관계 경험을 추적하는 종단 연구, 아동·청소년의 발달 단계별 AI 인식 및 의존 양상을 분석하는 연구, 역기능적 정서 의존을 판별할 수 있는 측정 기준 개발, AI 관계 종료에 따른 심리 반응 조사 등은 모두 정책적 활용 가능성이 높은 연구 과제들이다.

따라서 필요한 것은 단순한 연구 확대가 아니라, 정책과 연구 사이의 연결 구조를 갖춘 실증 연구 체계의 구축이다. 관계부처가 공동으로 참여하는 인간-AI 관계 영향 연구 프로그램이나 사업단을 통해, 연구 결과가 서비스 가이드라인, 위험 분류 체계, 공공 서비스 설계 기준에 반영될 수 있도록 하는 피드백 구조를 설계할 필요가 있다. 이는 제V장에서 제시한 정서적 영향의 조건성을 정책적으로 다루기 위한 핵심 인프라에 해당한다.

5. 관계형 AI 거버넌스와 관계 리터러시 교육 정비

이전 장에서 도출된 세 번째 관점은 관계적 위험이 기술 사용의 부산물이 아니라, 설계 단계와 거버넌스 구조 속에서 형성된다는 점이다. 이 점을 고려하면, 인간-AI 관계화에 대한 대응은 개별 이용자의 주의나 자기책임에 맡겨질 수 없다. 오히려 정책은 어떤 관계형 AI를 허용하고 장려할 것인지, 어떠한 설계 요소를 제한할 것인지, 그리고 이용자가 어떤 역량을 갖추어야 하는지를 함께 다루는 방향으로 정비될 필요가 있다.

이러한 과제는 본질적으로 다학제적 성격을 지닌다. 관계형 AI의 설계와 활용은 공학적 최적화의 문제를 넘어, 심리학, 사회학, 사회복지학, 철학, 법학, 커뮤니케이션 연구가 교차하는 지점에서 형성되기 때문이다. 따라서 전문가 집단과 이용자 대표가 함께 참여하는 형태의 관계형 AI 거버넌스 논의의 구조를 제도화하는 방안은 충분히 검토할 가치가 있다. 이는 부처별로 분절된 기

준과 가이드라인을 조정하고, 기술적 안전성·산업 진흥·이용자 보호 사이의 긴장을 보다 공개적으로 조율하는 장치가 될 수 있다.

동시에 시민 차원에서는 '관계 리터러시(relational literacy)'의 강화가 요구된다. 여기서 관계 리터러시란 AI와의 상호작용에서 발생할 수 있는 정서적 영향, 의존 가능성, 조작 가능성을 인식하고, 자신의 정서적 자율성을 유지할 수 있는 역량을 의미한다. 이는 단순한 디지털 활용 교육과는 구별되는 차원의 문제이며, 교육과 복지, 군 복무, 청소년 상담, 노인 돌봄과 같은 공공 접점에서 체계적으로 다루어질 필요가 있다. 관계형 AI 거버넌스와 관계 리터러시 교육을 함께 정비한다는 것은, 관계의 실재성을 인정하되 그 관계가 이용자의 자율성과 웰빙을 해치지 않도록 제도와 교육의 양 측면에서 대응하는 접근을 의미한다. 이는 인간-AI 관계가 일상적 환경으로 확산되는 상황에서 정책이 고려해야 할 핵심 방향으로 이해될 수 있다.

VII. 결론

본 연구는 생성형 AI의 확산이 인간-기술 관계를 관계의 문제로 재구성하고 있다는 점에 주목하고, 이를 '관계화'라는 개념으로 분석하였다. 생성형 AI는 이용자와의 반복적 상호작용을 통해 정서적 유대와 의존, 신뢰와 상실감과 같은 관계적 경험을 형성하며, 이는 인간-기술 상호작용을 기능적 사용에서 지속적 관계 경험으로 전환시키는 경향을 보인다. 이러한 변화는 기존의 기술 중심 접근만으로는 충분히 설명되기 어려운 새로운 사회적 양상을 드러낸다.

이론적 검토와 경험적 분석을 통해 본 연구는 인간-AI 관계화가 단일한 결과로 귀결되지 않으며, 이용자의 특성, 사용 맥락, 관계의 기능, 설계 방식에 따라 상이하게 전개되는 조건적 구조를 지닌다는 점을 확인하였다. 동시에 이러한 관계는 정서적 지지와 같은 긍정적 효과와 의존, 사회적 위축, 취약성 증폭과 같은 위험을 함께 내포하는 양면적 특성을 갖는다. 이는 인간-AI 관

계를 일률적으로 평가하기보다, 그 형성과 효과가 나타나는 조건과 맥락을 중심으로 이해할 필요가 있음을 시사한다.

이러한 분석을 바탕으로 국내 AI 정책을 검토한 결과, 현재 정책 문서는 기술적 성능, 산업 경쟁력, 활용 확산, 시스템 신뢰성 확보를 중심으로 구성되어 있으며, 인간-AI 관계의 실재성이나 정서적 영향, 관계적 위험은 정책 설계의 중심 변수로 충분히 통합되지 않은 것으로 나타났다. 다만 이는 개별 정책의 오류라기보다, 기술과 산업을 중심으로 형성된 정책 패러다임의 구조적 특성으로 이해할 수 있다. 즉, '관계적 공백'은 특정 영역의 미비라기보다, 인간-기술 관계를 기능과 성능 중심으로 이해해 온 정책 인식의 결과로 나타나는 경향으로 해석된다.

이 점에서 본 연구는 기존의 AI 정책 방향을 보완·확장할 것을 제안한다. 향후 AI 정책은 산업 경쟁력과 기술 혁신, 디지털 포용이라는 기존 목표를 유지하면서도, 인간-AI 관계의 형성과 그 사회적 효과를 함께 고려하는 방향으로 정교화될 필요가 있다. 특히 관계적 경험의 실재성, 정서적 영향의 조건성, 설계 단계에서의 관계적 위험이라는 세 가지 관점은 정책 설계 과정에서 추가적으로 고려될 수 있는 분석 기준으로 기능할 수 있다. 결국 생성형 AI시대의 정책 과제는 기술 확산을 관리하는 문제를 넘어, 기술이 매개하는 관계의 질과 그 사회적 결과를 어떻게 다룰 것인가에 대한 문제로 확장된다는 점에서 인간-AI 관계를 정책 분석의 유의미한 단위로 설정하는 작업은 정보화정책의 분석 지평을 확장하는 하나의 시도가 될 수 있을 것이다.

■ 참고문헌

- 관계부처 합동 (2022). <대한민국 디지털 전략>.
- 관계부처 합동 (2023). <전국민 인공지능 일상화 실행계획>.
- 관계부처 합동 (2025). <국가 AI역량 강화방안 후속조치>.
- 과학기술정보통신부 (2023). <초거대AI 경쟁력 강화 방안>.
- 과학기술정보통신부·디지털플랫폼정부위원회 (2023). <인공

- 지능 일상화 및 산업 고도화 계획(안)).
- 과학기술정보통신부·행정안전부 (2025). <AI 민생 10대 프로젝트 추진방안(안)>.
- 국가인공지능전략위원회 (2025a). <대한민국 인공지능 행동 계획(안)>.
- 국가인공지능전략위원회 (2025b). <대한민국 AI액션플랜 추진방향>.
- 국가인공지능전략위원회 (2025c). <AI컴퓨팅 인프라 확충을 통한 국가AI역량 강화방안>.
- 국회입법조사처 (2025). <생성형 인공지능과의 위험한 대화: 정서적 의존 위험과 '영향받는 자'를 위한 입법적 과제>. 국회입법조사처 이슈와 논점 제245호.
- 바른ICT연구소 (2020). <독거노인 대상 AI 스피커 효용성 분석>.
- 법제처 (2026). 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법. <국가법령정보센터>. <https://www.law.go.kr/lsInfoP.do?lsId=014820&ancYnChk=0#0000>
- 성옥준·예레미 (2024). 정보화정책 저널 30주년 리뷰: 정보화정책의 과거, 현재, 그리고 미래. <정보화정책>, 31권 4호, 3-62. <https://doi.org/10.22693/NIAIP.2024.31.4.003>
- 정동훈 (2026). 에이전틱 AI 시대의 인간-기계 상호작용 재구성에 관한 이론적 고찰. <정보화정책>, 33권 1호, 3-24. <https://doi.org/10.22693/NIAIP.2026.33.1.003>
- 주희·권소연 (2025). 토픽 모델링을 통한 심리상담 챗봇의 개입 유형별 사용자 경험 분석. <정보화정책> 32권 3호, 73-87. <https://doi.org/10.22693/NIAIP.2025.32.3.073>
- 폭력학대예방협회 (2022). <AI 상담 효과 지표>.
- Akbulut, M., Weidinger, L., Manzini, A., Gabriel, I. & Rieser, V. (2024). All too human? Mapping and mitigating the risk from anthropomorphic AI. [arXiv:2406.16577](https://arxiv.org/abs/2406.16577)
- Altman, I. & Taylor, D. A. (1973). Social penetration: The development of interpersonal relationships. Holt, Rinehart & Winston.
- Andoh, E. (2025, October 1). Many teens are turning to AI chatbots for friendship and emotional support. Monitor on Psychology. <https://www.apa.org/monitor/2025/10/technology-youth-friendships>
- Bernardo, V. (2025). AI companions. TechSonar, European Data Protection Supervisor (EDPS). <https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/ai-companions>
- Bickmore, T. W. & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2), 293-327. <https://doi.org/10.1145/1067860.1067867>
- Birkstedt, T., Minkinen, M., Tandon, A. & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, 33(7), 1-33. <https://doi.org/10.1108/INTR-01-2022-0042>
- Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. Basic Books.
- Brandtzæg, P. B., Skjuve, M. & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human-AI friendship. *Human Communication Research*, 48(3), 404-429. <https://doi.org/10.1093/hcr/hqac008>
- Callon, M. (1986). Some elements of a sociology of translation. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (pp. 196-233). Routledge & Kegan Paul.
- Chaturvedi, R., Verma, S., Das, R. & Dwivedi, Y. K. (2023). Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting and Social Change*, 193, Article 122634. <https://doi.org/10.1016/j.techfore.2023.122634>
- Cheng, M., Lee, C., Khadpe, P., Yu, S. & Han, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*. <https://doi.org/10.1126/science.adv4801>
- Chu, M. D., Gerard, P., Pawar, K. & Bickham, C. (2025). Illusions of intimacy: Emotional attachment and emerging psychological risks in human-AI relationships. Preprint.
- Common Sense Media. (2025, July 16). *Talk, trust, and trade-offs: How and why teens use AI companions*. <https://www.commonsensemedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions>

- Daft, R. L. & Lengel, R. H. (1986). Organizational information requirements, media richness, and structural design. *Management Science*, 32(5), 554-571.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.
- Dang, J. & Liu, L. (2023). Do lonely people seek robot companionship? *Computers in Human Behavior*, 141, 107637. <https://doi.org/10.1016/j.chb.2022.107637>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>
- De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z. & Puntoni, S. (2023). Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology*, 34(1), 150-163. <https://doi.org/10.1002/jcpy.1393>
- De Freitas, J., Oğuz-Uğuralp, Z., Uğuralp, A. K. & Puntoni, S. (2025). AI companions reduce loneliness. *Journal of Consumer Research*, 52(6), 1126-1147. <https://doi.org/10.1093/jcr/ucaf022>
- Djufril, M., Frampton, J. & Knobloch-Westerwick, S. (2025). Love, marriage, pregnancy: Commitment processes in romantic relationships with AI chatbots. *Computers in Human Behavior: Artificial Humans*, 4, 100155. <https://doi.org/10.1016/j.chbah.2025.100155>
- Epley, N., Waytz, A. & Cacioppo, J. T. (2007). On seeing the human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864-886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Gambino, A., Fox, J. & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1, 71-86. <https://doi.org/10.30658/hmc.1.5>
- Guingrich, R. E. & Graziano, M. S. A. (2025). Chatbots as social companions. In H. Shevlin (Ed.), *Oxford intersections: AI in society*. Oxford Academic. <https://doi.org/10.1093/oso/9780198937524.003.0011>
- Gur, O. & Maaravi, Y. (2025). The algorithm of friendship: Literature review and integrative model. *Behaviour & Information Technology*, 44(14), 3446-3466. <https://doi.org/10.1080/0144929X.2024.2443097>
- Gur, T. & Maaravi, Y. (2025). Human-AI relationships: A systematic review and comprehensive model. *Computers in Human Behavior*, 165, 108532. <https://doi.org/10.1016/j.chb.2024.108532>
- Guzmán, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, 1, 27-50. <https://doi.org/10.30658/hmc.1.3>
- Guzmán, A. L. & Lewis, S. C. (2019). Artificial intelligence and communication: A Human-Machine Communication research agenda. *New Media & Society*, 21(8), 1673-1689. <https://doi.org/10.1177/1461444819858691>
- Horton, D. & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, 19(3), 215-229. <https://doi.org/10.1080/00332747.1956.11023049>
- Illouz, E. (2007). *Cold intimacies: The making of emotional capitalism*. Polity Press.
- Jecker, N. S., Sparrow, R., Lederman, Z. & Ho, A. (2024). Digital humans to combat loneliness and social isolation: Ethics concerns and policy recommendations. *Hastings Center Report*, 54(1). <https://doi.org/10.1080/15265161.2024.2306182>
- Kamene, K. (2025). From counsel to code: Implications of AI dependency for authentic human bonds – A systematic review. *Premier Journal of Artificial Intelligence*, 5, 100020. <https://doi.org/10.70389/PJAI.100020>
- Kirk, H. R., Vidgen, B., Bhatt, U., Wachter, S., Russell, C., Naudé, W. & Eloundou, T. (2025). Why human-AI relationships need socioaffective alignment. *Humanities and Social Sciences Communications*, 12, 728. <https://doi.org/10.1038/s43771-025-00000-0>

- 1057/s41599-025-04697-7
- Koh, E. (2023). "Date me date me" AI chatbot interactions as a resource for the online construction of masculinity. *Discourse, Context & Media*, 52, 100681. <https://doi.org/10.1016/j.dcm.2023.100681>
- Krakowski, I., Kim, J., Cai, Z. R., Daneshjou, R., Lapins, J., Eriksson, H., Lykou, A. & Linos, E. (2024). Human-AI interaction in skin cancer diagnosis. *npj Digital Medicine*, 7, 78. <https://doi.org/10.1038/s41746-024-01031-w>
- Laestadius, L., Bishop, A., Gonzalez, M., Illencik, D. & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis. *New Media & Society*, 25(8), 1956-1976. <https://doi.org/10.1177/14614448221142007>
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press. <https://doi.org/10.1093/oso/9780199256044.001.0001>
- Lee, H., Mariani, M. M. & Umemuro, H. (2026). Impact of AI companion's affective behavior on user dependence and social wellbeing. *Journal of Business Research*, 193, 115347. <https://doi.org/10.1016/j.jbusres.2025.115347>
- Lee, J. (2026). From adolescence to older adulthood: Lifespan pathways linking AI companion chatbots to mental health. *Lifespan Development and Mental Health*, 2, 10005. <https://doi.org/10.70322/ldmh.2026.10005>
- Lee, S. K., Kavya, P., & Lasser, S. C. (2021). Social interactions and relationships with an intelligent virtual agent. *International Journal of Human-Computer Studies*, 150, 102608. <https://doi.org/10.1016/j.ijhcs.2021.102608>
- Lee, Y., Mariani, M. & Umemuro, H. (2026). Effects of AI companions' sycophancy and emotional mimicry on consumers' continuance intention and social wellbeing. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2026.2626809>
- Li, H. & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts. *Journal of Computer-Mediated Communication*, 29(6). <https://doi.org/10.1093/jcmc/zmae015>
- Liu, A. R., Pataranutaporn, P. & Maes, P. (2024). Chatbot companionship: A mixed-methods study. arXiv preprint. <https://arxiv.org/abs/2410.21596>
- Lombard, M. & Xu, K. (2021). Social responses to media technologies in the 21st century. *Human-Machine Communication*, 2, 29-55. <https://doi.org/10.30658/hmc.2.2>
- Loveys, K., Hiko, C., Sagar, M., Zhang, X. & Broadbent, E. (2022). "I felt her company": A qualitative study on factors affecting closeness. *International Journal of Human-Computer Studies*, 160, 102771. <https://doi.org/10.1016/j.ijhcs.2021.102771>
- Maeda, T. & Quan-Haase, A. (2024). When human-AI interactions become parasocial. Proceedings of FAccT '24. <https://doi.org/10.1145/3630106.3658956>
- Manzini, A., Keeling, G., Alberts, L., Vallor, S. & Morris, M. R. (2024). *The code that binds us*. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES 2024). <https://doi.org/10.1609/aies.v7i1.31694>
- Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M. & Tielman, M. L. (2024). A systematic review on fostering appropriate trust in human-AI interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing*, 1(4). <https://doi.org/10.1145/3696449>
- Merrill, K., Kim, J. & Collins, C. (2022). AI companions for lonely individuals and the role of social presence. *Communication Research Reports*, 39(2), 112-123. <https://doi.org/10.1080/08824096.2021.2024926>
- Nass, C., Steuer, J. & Tauber, E. R. (1994). *Computers are social actors*. Proceedings of CHI '94, 72-78. <https://doi.org/10.1145/191666.191703>
- Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103. <https://doi.org/10.1080/00224520009595888>

- doi.org/10.1111/0022-4537.00153
- Pataranutaporn, P., Karny, S., Archiwaranguprok, C., Albrecht, C., Liu, A. R. & Maes, P. (2025). "My boyfriend is AI": A computational analysis. arXiv. <https://doi.org/10.48550/arXiv.2509.11391>
- Pentina, I., Xie, T., Hancock, T. & Bailey, A. (2023). Consumer-machine relationships in the age of AI. *Psychology & Marketing*, 40(6), 1211-1229. <https://doi.org/10.1002/mar.21799>
- Picard, R. W. (1997). *Affective computing*. MIT Press. <https://doi.org/10.7551/mitpress/1140.001.0001>
- Raees, M., Meijerink, I., Lykourantzou, I., Khan, V.-J. & Papangelis, K. (2024). From explainable to interactive AI. *International Journal of Human-Computer Studies*, 189, 103301. <https://doi.org/10.1016/j.ijhcs.2024.103301>
- Rees, L. & Bahmani, M. (2025). The relational tradeoff model. *Organizational Psychology Review*, 1-26. <https://doi.org/10.1177/20413866251399871>
- Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.
- San Buenaventura, S. H.-A. (2025). AI and people: A review of current research on human-machine interaction. *World Journal of Advanced Research and Reviews*, 26(3), 2707-2715. <https://doi.org/10.30574/wjarr.2025.26.3.2486>
- Saracini, C., Cornejo-Plaza, M. I. & Cippitani, R. (2025). Techno-emotional projection in human-GenAI relationships. *Frontiers in Psychology*, 16, 1662206. <https://doi.org/10.3389/fpsyg.2025.1662206>
- Shen, H., Kneare, T., Ghosh, R., Alkie, K., Krishna, K., Liu, Y. & Kulkarni, C. (2024). Towards bidirectional human-AI alignment. arXiv preprint. <https://arxiv.org/abs/2406.09264>
- Short, J., Williams, E. & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.
- Skjuve, M., Følstad, A., Fostervold, K. I. & Brandtzaeg, P. B. (2021). My chatbot companion: A study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>
- Skjuve, M., Følstad, A., Fostervold, K. I. & Brandtzaeg, P. B. (2022). A longitudinal study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 168, 102903. <https://doi.org/10.1016/j.ijhcs.2022.102903>
- Smith, M. G., Bradbury, T. N. & Karney, B. R. (2025). Can Generative AI Chatbots Emulate Human Connection? *Perspectives on Psychological Science*, 20(6), 1081-1099. <https://doi.org/10.1177/17456916251351306>
- Sun, X., Wang, Y. & McDaniel, B. T. (2026). AI companions and adolescent social relationships. *Child Development Perspectives*. <https://doi.org/10.1093/cdpers/aadaf009>
- Sundar, S. S. & Lee, E. J. (2022). Rethinking communication in the era of artificial intelligence. *Human Communication Research*, 48(3), 379-385. <https://doi.org/10.1093/hcr/hqac014>
- Ta, V., Griffith, C., Boatfield, C., Wang, X. & Civitello, M. (2020). User experiences of social support from companion chatbots. *Journal of Medical Internet Research*, 22(3), e16235. <https://doi.org/10.2196/16235>
- Veitch, E. & Alsos, O. A. (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety Science*, 152, 105778. <https://doi.org/10.1016/j.ssci.2022.105778>
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478. <https://doi.org/10.2307/30036540>
- Ventura, A., Starke, C. & Righetti, F. (2025). Relationships in the age of AI. Preprint.
- Wang, X., Pang, C. C. & Hui, P. (2025). 'My dataset of love': A preliminary mixed-method exploration. *Proceedings of the ACM on Human-Computer Interaction*, 9(CSCW2), Article 75. <https://doi.org/10.1145/3757532>

- Wang, S. & Dehnert, M. (2026). On-demand intimacy: The sociotechnical appeal of AI companions. *Social Media + Society*, 12(1). <https://doi.org/10.1177/20563051251335946>
- Waytz, A., Cacioppo, J. T. & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219-232. <https://doi.org/10.1177/1745691610369336>
- Xie, T. & Pentina, I. (2022). *Attachment theory as a framework to understand relationships with social chatbots*. Proceedings of HICSS 2022. <https://doi.org/10.24251/HICSS.2022.204>
- Xie, T., Pentina, I. & Hancock, T. (2023a). Friend, mentor, lover: Does chatbot engagement lead to psychological dependence? *Journal of Service Management*, 34(4), 667-690. <https://doi.org/10.1108/JOSM-02-2022-0050>
- Xie, T., Pentina, I. & Hancock, T. (2023b). The attachment theory as a framework for understanding human-chatbot relationship development. *MIS Quarterly*, 47(4), 2046-2055. <https://doi.org/10.25300/MISQ/2023/17727>
- Ye, R. (2025, July 25). AI Companionship: A New Frontier. In H. Shevlin (Ed.), *Relationships*, in P. Hacker (Ed.), *Oxford Intersections: AI in Society*. Oxford Academic. <https://doi.org/10.1093/9780198945215.003.0070>
- Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R. E. & Yang, D. (2025). The rise of AI companions: How human-chatbot relationships influence well-being. *arXiv*. <https://doi.org/10.48550/arXiv.2506.12605>
- Zhou, L., Gao, J., Li, D. & Shum, H.-Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 46(1), 53-93. https://doi.org/10.1162/coli_a_00368