

에이전틱 AI 시대의 인간-기계 상호작용 재구성에 관한 이론적 고찰*

정동훈**

요약

생성형 AI의 발전과 함께 등장한 에이전틱 AI(Agentic AI)는 스스로 목표를 설정하고 성찰하는 자율적 인지 과정을 통해 인간-기계 상호작용의 양상을 근본적으로 변화시키고 있다. 이 연구는 기존 이론이 전제해 온 도구적 통제 패러다임이 자율성을 지닌 AI 환경에서 한계에 직면했음을 문헌 고찰을 통해 비판적으로 규명하고, 새로운 협력적 관계 모델을 모색한다. 이를 위해 먼저 결정론적 인과성에 기반한 전통적 모델이 에이전틱 AI의 확률적 특성과 충돌하는 지점을 분석한다. 이어 CASA(Computers Are Social Actors) 패러다임과 행위자-네트워크 이론(ANT)을 통합하여, 인간이 AI를 파트너로 인식하는 심리적 기제와 AI가 조직 내 권한을 재편하는 구조적 기제를 체계화한다. 나아가 HACO, COFI, FAICO, UCCC, CCDF와 같은 최신 협업 프레임워크를 구조, 소통, 행위자성 차원에서 범주화하여 실무적 분석 틀을 제안한다. 결론적으로 이 연구는 인간-AI 관계가 일방적 명령과 수행의 관계에서 권한 위임과 상호 조율을 통한 공진화(co-evolution)의 파트너십으로 전환되고 있음을 주장하며, 하이브리드 인텔리전스 실현을 위한 이론적 함의를 제시한다.

주제어 : 에이전틱 AI, 인간-AI 상호작용, 도구적 통제, 하이브리드 인텔리전스, 공진화, 행위자성

A Theoretical Review on Reconfiguring Human-Machine Interaction in the Era of Agentic AI*

Chung, Donghun**

Abstract

The rise of Agentic AI, characterized by autonomous cognitive processes of perception, action, and reflection, challenges the traditional paradigm that views technology merely as a passive tool. This study conducts a comprehensive review of the limitations of the instrumental control model in the context of autonomous systems and explores new collaborative models. By integrating the Computers Are Social Actors(CASA) paradigm with Actor-Network Theory(ANT), the psychological mechanisms leading humans to perceive AI as a partner and the structural mechanisms through which AI reconfigures organizational agency are analyzed. Furthermore, emerging frameworks—such as HACO, COFI, FAICO, UCCC, and CCDF—are systematically categorized into structural, communicative, and agentic dimensions to provide an analytical lens for collaboration. This study argues that the human-AI relationship is shifting from a master-slave dynamic rooted in direct manipulation to a collaborative partnership based on delegation and mutual adaptation. Ultimately, this research proposes theoretical foundations for hybrid intelligence, where humans and AI co-evolve to achieve outcomes beyond individual capabilities.

Keywords : agentic AI, human-AI interaction, instrumental control, hybrid intelligence, co-evolution, agency

접수: 2026. 2. 19; 게재확정: 2026. 2. 27

* 이 논문은 2025년도 광운대학교 교내 학술연구비 지원에 의해 연구되었음

** 광운대학교 미디어커뮤니케이션학부 교수(donghunc@gmail.com, <https://orcid.org/0000-0001-8563-4392>)

I. 서론

인공지능(Artificial Intelligence: AI) 기술의 급격한 발전은 인간이 기계와 소통하는 방식을 근본적으로 재편하고 있다. 특히 2022년 등장한 생성형 AI(Generative AI)와 대규모 언어 모델(Large Language Model: LLM)은 단순히 사용자의 명령에 반응하는 수준을 넘어, 텍스트, 이미지, 코드 등 창의적인 산출물을 생성하며 인간의 지적 노동을 보조하는 강력한 도구로 자리 잡았다. 그러나 최근 기술의 흐름은 이러한 생성형 모델을 다시 한번 넘어서고 있다. 스스로 목표를 설정하고, 환경을 인식하며, 계획을 수립하고, 행동을 수행한 뒤 결과를 성찰하는 자율적 인지 루프를 갖춘 에이전틱 AI(Agentic AI)의 등장이다.

에이전틱 AI는 자율적이고 목표 지향적인 특성을 갖춘 AI 시스템으로, 환경에 반응하며 자기 주도적으로 작업을 수행하고 목표를 달성하기 위해 전략을 동적으로 조정하는 능력을 지닌 시스템으로 정의된다(Hosseini & Seilani, 2025; Sapkota et al., 2026). 에이전틱 AI의 등장은 기존의 커뮤니케이션학 및 인간-컴퓨터 상호작용(Human-Computer Interaction: HCI) 연구가 전제해 온 '인간=주체, 기계=도구'라는 이분법적 도식으로는 설명할 수 없는 새로운 관계를 설정한다. 전통적인 커뮤니케이션 모델은 미디어를 단순히 메시지가 전달되는 중립적 채널이나 매개체로 규정해 왔다. 이 관점에서 의미를 생산하고 해석하는 주체는 오직 인간뿐이며, 기술은 이를 효율적으로 전달하거나 저장하는 수단에 불과했다. HCI 분야의 고전적인 이론들 역시 사용자가 명확한 목표를 가지고 시스템을 통제하는 사용성(usability)과 제어(control)에 초점을 맞추어 왔다. Norman(2013)의 실행과 평가의 간극(Gulf of Execution and Evaluation) 모델은 사용자가 시스템의 상태를 정확히 파악하고 의도대로 조작할 수 있도록 돕는 것이 인터페이스 설계의 핵심임을 강조한다.

그러나 에이전틱 AI는 이러한 도구적 관점으로는 설명하기 어려운 새로운 상호작용 양상을 보여준다. 에이

전틱 AI는 사용자의 의도를 추론하여 스스로 과업을 정의하고, 인간의 개입 없이도 외부 도구(Application Programming Interface: API)를 사용하거나 다른 에이전트와 협업하여 문제를 해결한다. 이제 인간과 AI의 관계는 명확한 명령어를 입력하고 결과를 확인하는 지시와 수행의 관계에서, 서로의 의도를 해석하고 조율하며 결과물을 공동으로 창작하는 협업과 파트너십의 관계로 진화하고 있다. 이는 AI를 단순한 도구가 아니라, 사회적 실재감(social presence)을 지닌 커뮤니케이션 파트너 혹은 행위자(actor)로 재정의해야 함을 시사한다.

이 연구는 이러한 문제의식에서 출발하여, 에이전틱 AI 시대에 요구되는 새로운 커뮤니케이션 패러다임을 모색하는 것을 목적으로 한다. 기존의 HCI 모델은 시스템의 투명성과 제어 가능성을 강조했지만, 자율적으로 학습하고 판단하는 에이전틱 AI 환경에서는 시스템 내부의 불투명성(black box)과 결과의 불확실성이 필연적으로 증대된다. 이에 따라 사용자는 AI가 생성한 결과물을 맹목적으로 신뢰하거나, 또는 반대로 과도하게 의심하는 단계에 직면하게 되며, 이는 단순한 기술적 개선만으로는 해결할 수 없는 신뢰와 관계의 문제를 야기한다.

이와 같은 인식을 바탕으로 이 연구는 크게 세 개 차원에서 논의를 전개하고자 한다. 먼저, 도구적 관점에 머물러 있는 기존 HCI 모델의 한계를 분석하고, AI의 이중적 지위를 규명할 것이다. 그리고, 최근 등장한 인간-AI 협업 분석 프레임워크를 검토하여, 에이전틱 AI와의 상호작용을 구조, 소통의 질, 행위자성(agency) 측면에서 체계화할 것이다. 마지막으로 현존하는 분석 틀이 특정 분야나 미시적 과업 단위에 국한되어 있음을 비판적으로 고찰하고, 이를 통해 향후 연구가 나아가야 할 방향성을 제시할 것이다. 결국 이 연구는 에이전틱 AI를 통제의 대상이 아닌 소통과 협력의 주체로 격상시킴으로써, 인공지능 시대의 커뮤니케이션이 기술적 최적화를 넘어 인간과 기계가 공존하는 새로운 사회적 관계 맺기로 확장되어야 함을 주장하며 새로운 이론과

모델 개선을 위한 마중물이 되고자 한다.

II. 도구적 통제 패러다임의 이론적 전제와 한계

1. 커뮤니케이션 매체와 통제 대상으로서의 기술관

전통적 커뮤니케이션 모델은 정보의 기계적 전송에서 의미의 사회적 구성으로 그 논의를 확장해 왔다. Shannon and Weaver(1949)의 수학적 모델(The Mathematical Theory of Communication)은 커뮤니케이션을 정보원(source)이 메시지를 신호로 변환하여 채널(channel)을 통해 수신자(receiver)에게 전달하는 일방향적 프로세스로 규정했다. 이 모델의 핵심은 메시지의 전달 효율성과 정확성에 있으며, 신호 전달 과정에서 발생하는 물리적 방해 요소인 잡음(noise)의 개념을 도입하여, 기술적 통신 과정에서의 정보 손실 문제를 다뤘다. 초기 이론들이 메시지의 효율적 전송(transmission)에 집중했다면, 후기 이론들은 의미의 구성과 관계로 관심을 확장했다. 이 과정에서 커뮤니케이션은 단순한 메시지 교환이 아니라, 인간과 매개 환경이 결합된 복합적인 과정으로 재개념화됐다.

그러나 커뮤니케이션학의 논의는 기술을 주로 의미를 전달하는 중립적이고 수동적인 매개체로 규정함에 따라, 인간과 기술이 구체적인 접점에서 어떻게 상호작용하며 목적을 달성하는지에 대한 기제 설명에는 한계를 보였다(Littlejohn et al., 2017). 이러한 결과, 인간과 기계 간의 직접적인 교류를 규명하기 위한 새로운 분석틀이 필요하게 됐고, 그 결과로 HCI 연구가 등장했다.

거시적 관점에서 초기 HCI는 Wiener(1948)가 주창한 사이버네틱스(cybernetics) 전통에 영향을 받았다. 사이버네틱스는 커뮤니케이션을 '시스템의 상태를 조절하고 통제하는 과정'으로 이해하며, 인간과 기계를 목표 달성을 위해 피드백을 주고받는 하나의 통합된 시스템으로 간주한다. 이 관점에서 상호작용은 정보의 흐름을 통해 시스템의 무질서를 낮추고, 원하는 상태를 유지

하려는 통제가 핵심이다. 따라서 기술 시스템은 독자적인 의지를 가진 존재가 아니라, 사용자가 입력한 목표값에 맞춰 오차를 수정하고 기능을 수행하는 자동화된 도구로 정의됐다. 이 패러다임 하에서 이상적인 인간과 컴퓨터 관계는 인간이 시스템의 입력과 출력을 완벽하게 예측하고 통제할 수 있는 상태의 완결성에 달려 있었다.

커뮤니케이션학이 인간과 인간 간의 상호작용에 집중한다면, HCI는 인간과 기술이 관계를 맺는 방식을 체계적으로 모델화해 온 연구 분야다. 전통적인 HCI는 의미의 공유나 사회적 관계보다는, 사용자가 기술 시스템을 어떻게 이해하고, 조작하며, 통제하는가라는 문제에 초점을 맞추어 왔다(Norman, 2013). 이러한 접근은 인간과 기술 관계를 소통의 문제라기보다 도구 사용과 수행의 문제로 정의한다(Shneiderman, 1983). 이 관점에서 기술은 사용자의 명령을 오류 없이 수행해야 하는 수동적 존재일 뿐, 독자적인 의도를 가지고 인간과 교류하는 행위자로 다루어지지 않는다(Zhao & Xu, 2025; Rezwana & Maher, 2023). 그러므로 생성형 AI와 같이 자율성을 가진 기술과의 상호작용을 논의하기에 앞서, 기존 HCI 모델이 전제하고 있는 도구적 통제의 패러다임을 검토하는 것은 필수적이다.

2. 이원적 상호작용 모델의 고착화

전통적인 HCI 연구와 사용성 평가의 분석 단위는 오랫동안 개인 사용자에게 고정되어 있었다. 사용자 중심 디자인(user-centered design)이나 휴리스틱(heuristic) 평가와 같은 대표적 접근들은 기본적으로 하나의 사용자가 하나의 시스템과 상호작용하며 과업 효율성과 오류 최소화를 달성하는 상황을 전제로 설계됐다(Norman, 2013). 이와 같은 이원적(dyadic) 구조에서 시스템은 사용자의 목표 달성을 지원하는 기능적 도구로 규정되며, 상호작용의 성공 여부는 사용자의 통제 가능성, 예측 가능성, 학습 용이성 등에 의해 판단된다. 물론 CSCW(Computer-Supported Cooperative Work)와 같은 연구 흐름은 협업 상황을 분석 단위로 확장해

왔으나, 이 경우에도 기술은 인간과 인간 사이의 협업을 가능하게 하는 매개체로 이해되는 경우가 대부분이었다. 다시 말해, 협업의 주체는 여전히 인간이며, 기술은 그 관계를 연결하거나 조정되는 수동적이면서도 제한적인 역할에 머물렀다. 기술 그 자체가 협업의 참여자나 역할 수행자로 개념화되는 경우는 극히 제한적이었다.

그러나 이러한 도구 중심적 관점은 인간의 실제 상호작용 양상과 점차 괴리를 보이기 시작했다(Nass et al., 1994; Reeves & Nass, 1996). 사람들은 컴퓨터의 피드백에 감정적으로 반응하고, 기계의 '체면'을 고려해 비판을 자제하기도 한다. 이는 인간이 기술을 대할 때 합리적 사용자라기보다 사회적 행위자로 반응하는 경향을 지닌다는 점을 보여준다. 그럼에도 불구하고 전통적 HCI 이론은 이러한 현상을 의인화(anthropomorphism)라는 부차적 효과나 통제해야 할 오류로 치부하며, 인간의 사회적 본능을 수용할 이론적, 분석적 틀을 갖추지 못했다.

전통적 시스템은 사용자의 사회적 단서(social cues)를 해석하거나, 상황에 따라 관계적 반응을 조절할 능력을 갖추지 못했기 때문에, 사용자는 기술과의 상호작용에서 자신의 사회적 기대와 감정적 반응을 억제하고 스스로를 일방적으로 주도하는 사용자 역할로 축소해야 했다. 결과적으로 인간-기계 상호작용은 상호 주관적 의미 교환이 차단된 채, 명령과 수행, 입력과 출력의 문제로 환원됐다. 커뮤니케이션학이 의미의 구성에 집중했다면, HCI는 도구의 효율적 통제에 집중했는데, 이러한 도구적 관점은 기술이 단순한 수단을 넘어 사회적 상호작용의 대상으로 인식되는 작업의 현실과 괴리를 보인다.

3. 결정론적 명령과 직접 조작 원칙의 비판적 검토

20세기 후반 컴퓨터 과학과 HCI 분야가 독립적으로 발전하는 과정에서, HCI는 앞서 논의한 커뮤니케이션 이론의 정교한 개념 틀을 거의 참조하지 않은 채 도구로서의 컴퓨터와 사용자 간 상호작용의 효율성을 중심으로 학문 체계를 구축했다. 이 과정에서 흥미로운 점은,

전통적 커뮤니케이션 모델에서 비판의 대상이 됐던 선형적 정보 전달과 도구적 통제 중심 관점이 기술 매개 환경에서는 오히려 핵심적인 설계 원리로 재정당화 됐다는 사실이다. 이는 HCI 초기의 핵심 연구 축이 주로 컴퓨터 과학자들로 구성되면서, 커뮤니케이션 이론이 축적해 온 의미·관계·사회적 맥락에 대한 통찰이 충분히 흡수되지 못한 채, 도구주의적 관점이 구조적으로 고착된 결과로 해석할 수 있다.

이러한 도구주의적 관점이 인터페이스 설계의 규범으로 가장 명확하게 제시된 사례가 Shneiderman (1983)이 제안한 직접 조작(direct manipulation) 원칙이다. 직접 조작 모델은 사용자가 화면 위의 객체를 지속적으로 시각화된 상태로 인식하고, 클릭이나 드래그와 같은 물리적 행위를 통해 명령을 내리면, 시스템이 즉각적이고 가역적인 피드백을 제공해야 한다고 규정한다. 이 모델에서 이상적인 상호작용은 사용자의 의도가 인터페이스를 매개로 손실 없이 시스템에 전달되고, 그 결과가 즉시 확인되는 상태이다. 이러한 관점은 송신자의 의도가 잡음 없이 수신자에게 전달되어 통제되는 상태를 이상으로 상정했던 선형 커뮤니케이션 모델의 전제를 기술적 환경에서 구현한 사례로 볼 수 있다(Hutchins et al., 1985).

이 패러다임의 핵심 전제는 결정론적 인과성이다. 전통적인 HCI 시스템에서 컴퓨터는 사용자의 명시적인 명령(command) 없이는 어떠한 상태 변화도 일으키지 않는 수동적인 존재로 정의된다(Zhao & Xu, 2025). 사용자가 동일한 입력을 넣으면 시스템은 언제나 동일한 결과를 산출해야 하며, 시스템의 모든 반응은 사용자의 예측 범위 내에 있어야 한다. 전통적 HCI 시스템에서 의사결정의 주체는 전적으로 사용자이며, 시스템은 그 결정을 충실히 이행하는 도구에 불과하다. 따라서 이 모델에서 이상적인 상호작용이란 사용자가 시스템의 상태를 완벽하게 장악하고 있다는 통제감과 숙련도를 느끼는 상태를 의미한다(Shneiderman, 1983).

그러나 이러한 모델은 상호작용 과정에서 발생하는 인지적 부담을 사용자에게 집중적으로 전가한다는 근

〈표 1〉 도구적 통제 패러다임의 이론적 전제와 한계

구분	기존 HCI 관점의 이론적 전제	비판 및 한계
기술의 본질	기계를 정보 전달의 중립적 매체이자 인간의 명령을 수행하는 수동적 도구로 규정	기술이 상호작용 과정에서 의미를 생성하거나 매개하는 능동적 역할을 설명하지 못함
상호작용 구조	사용자와 시스템 간의 이원적이고 기능적인 관계에 집중하며 효율성을 최우선으로 함	인간이 기계를 사회적 행위자로 대우하는 본능적 심리와 관계적 역동성을 간과
설계 원칙	사용자의 명시적 명령에 의해 시스템 상태가 변하는 직접 조작과 결정론적 통제를 강조	시스템의 복잡성이 증가할수록 사용자에게 과도한 감시 의무와 인지적 부하를 전가

본적인 한계를 내포한다. 시스템의 기능과 복잡성이 증가할수록 사용자는 더 많은 선택지와 상태 변화를 추적해야 하며, 각 단계에서 무엇을 명령해야 하는지 스스로 판단해야 한다. 즉, 직접 조작 모델은 간극을 줄이기 보다는, 일정 수준 이상의 복잡성에 도달하면 오히려 사용자에게 지속적인 주의, 계획, 감시를 요구하는 구조로 전환된다. 이 관점에서 기존 HCI 모델은 기계가 사용자의 목표나 맥락을 해석하여 행동을 조정하는 가능성을 원천적으로 배제한 채, 모든 상호작용의 책임을 사용자에게 귀속시킨다.

III. 결정론적 상호작용 모델의 붕괴와 새로운 관계성의 모색

1. 실행의 축소와 평가의 확장, 인지 모델의 비대칭성

도구적 통제 패러다임의 한계는 단순한 인터페이스의 문제를 넘어, 인간의 인지 과정과 상황적 맥락을 다루는 모델의 붕괴로 이어진다. 기계를 예측 가능하고 완전히 통제할 수 있는 도구로 전제된 인터페이스 모델은, 시스템의 복잡성과 불확실성이 증가함에 따라 점차 구조적 한계에 직면하게 된다. 사용자가 모든 명령을 명시적으로 지정하고 결과를 직접 확인해야 하는 상호작용 구조는 초기에는 높은 통제감과 숙련도를 제공했지만, 자동화 수준이 고도화될수록 오히려 인간에게 요구되는 감시와 책임의 부담을 증폭시키는 역설을 낳는다. Bainbridge(1983)가 지적한 ‘자동화의 역설(ironies

of automation)’은 이러한 현상을 단적으로 보여주며, 이는 이후 HCI 연구가 도구적 통제 모델을 넘어 인지적·조직적 한계를 본격적으로 재검토하게 되는 이론적 출발점이 된다.

이러한 문제의식을 체계적으로 설명한 대표적 틀이 Norman(2013)의 실행의 간극(gulf of execution)과 평가의 간극(gulf of evaluation)이다. 실행의 간극은 사용자의 목표나 의도를 시스템이 이해할 수 있는 구체적 행동이나 명령으로 변환하는 과정에서 발생하는 어려움을 의미하며, 평가의 간극은 시스템의 출력 결과를 해석하여 그것이 자신의 원래 의도를 충족했는지를 판단하는 과정에서 발생하는 어려움을 가리킨다. 전통적인 HCI와 사용자 중심 디자인은 시각적 피드백, 직관적 대응(mapping), 일관된 인터페이스 규칙을 통해 이 두 간극을 동시에 축소함으로써, 사용자가 시스템을 쉽게 학습하고 통제할 수 있도록 하는 데 주력해 왔다(Hutchins et al., 1985). 직접 조작 인터페이스가 이러한 간극을 시각화와 즉각적 피드백을 통해 줄이려는 시도로 기능했음을 보여주었지만, 이는 어디까지나 사용자가 명령과 결과를 일일이 관리할 수 있는 범위 내에서만 효과적이다. 이 접근은 두 간극이 대체로 대칭적으로 관리 가능하다는 전제를 내포한다는 점에서 도구주의적 HCI의 핵심 논리를 형성한다.

그러나 이러한 인지 모델은 본질적으로 시스템이 사용자의 예측 범위 안에 있어야 한다는 전제에서 출발한다. 사용자의 행동과 시스템의 반응 사이에 비교적 안정적 대응 관계가 성립할 때는, 실행과 평가의 간극을

동시에 줄이는 전략이 유효하다. 반면, 시스템의 내부 구조가 복잡해지고 상태 변화의 경로가 다층화 될수록, 두 간극은 동일한 방식으로 좁혀지지 않는다. 실제로 많은 HCI 시스템은 사용자가 행동을 실행하는 경로를 단순화하는 데에는 성공했지만, 그 결과가 무엇을 의미하는지, 왜 그러한 결과가 도출됐는지를 이해하는 부담은 여전히 사용자에게 남겨두는 경향을 보였다. 이에 따라 실행의 간극은 점진적으로 축소되는 반면, 평가의 간극은 상대적으로 구조화되지 않은 채 확장되는 비대칭적 양상이 나타난다.

이러한 비대칭성은 자동화 연구에서 오래전부터 지적되어 온 문제와도 맞닿아 있다. Bainbridge(1983)는 자동화가 고도화될수록 인간의 역할이 직접적인 수행에서 감시와 판단으로 이동하지만, 역설적으로 이 감시와 판단이 가장 어려운 과제가 된다고 지적했다. 자동화된 시스템은 정상적으로 작동할 때는 사용자의 개입을 거의 필요로 하지 않지만, 문제가 발생했을 때는 사용자가 즉각적으로 상황을 파악하고 적절한 판단을 내려야 한다. 그러나 시스템의 내부 작동 과정이 사용자에게 충분히 노출되지 않을 경우, 사용자는 오히려 더 큰 인지적 부담과 책임을 떠안게 된다. 이는 실행의 부담은 줄어드는 반면, 평가와 책임의 부담은 증가하는 구조적 불균형을 초래한다.

인지 모델의 이러한 비대칭성은 HCI가 전제해 온 인간-기계 관계의 한계를 드러낸다. 전통적 인지 모델에서 인간은 목표를 설정하고 계획을 수립하는 주체이며, 기계는 그 계획을 충실히 실행하는 수단으로 간주한다. 이 구조에서 기계는 사용자의 의도를 해석하거나 판단에 개입하지 않으며, 모든 의미 부여와 최종 책임은 인간에게 귀속된다. 결과적으로 시스템이 제공하는 정보는 '행동을 돕는 단서'로 기능할 뿐, 그 자체로 의미를 구성하거나 방향을 제시하는 역할을 수행하지 않는다.

문제는 이러한 전제는 시스템이 복잡하면 복잡할수록 점점 더 취약해진다는 점이다. 사용자는 더 많은 기능과 더 복잡한 상태를 관리해야 하며, 각 상태가 무엇을 의미하는지, 다음에 어떤 행동이 적절한지를 스스로

판단해야 한다. 이 과정에서 인터페이스는 사용자의 인지를 보조하기보다는, 오히려 지속적인 주의와 감시를 요구하는 대상으로 전환된다. 즉, 사용자 중심을 표방했던 인지 모델이 역설적으로 사용자에게 과도한 인지적 책임을 전가하는 구조로 작동하는 역설적 상황이 되는 것이다.

결국 실행의 간극과 평가의 간극을 중심으로 한 전통적 인지 모델은, 인간과 시스템의 관계를 여전히 '계획하는 인간'과 '수행하는 기계'라는 이원적 구도로 고정한다는 한계를 갖는다. 이러한 구도에서는 시스템이 상황에 따라 유연하게 판단하거나, 사용자의 목표를 재구성하거나, 상호작용의 흐름을 조정하는 가능성이 이론적으로 배제된다. 이에 따라 HCI 연구는 점차 계획 중심 상호작용의 한계에 직면하고, 상황에 따라 변화할 수 있는, 그리고 인간과 시스템 간 역할 분담의 재구성을 모색하게 된다. 이러한 문제의식은 이후 HCI가 단순한 도구 설계의 영역을 넘어, 인간과 시스템의 관계 자체를 재정의하려는 이론적 전환으로 나아가는 중요한 출발점이 된다.

2. 인간-기계 이원론의 한계와 상황적 행위

앞선 인지 모델에서 논의한 실행 간극과 평가 간극의 비대칭성은, 전통적 HCI가 암묵적으로 전제해 온 계획 중심 인간 행위 모델의 구조적 한계를 보다 근본적인 수준에서 드러낸다. 초기 HCI 이론에서 인간은 목표를 명확히 설정하고, 그 목표를 달성하기 위한 합리적 계획을 수립한 뒤, 이를 단계적으로 실행하는 주체로 가정했다(Card et al., 1983; John & Kieras, 1996; Norman, 2013). 이러한 전제는 인간의 행위를 예측할 수 있고 안정적인 인지 과정으로 모형화할 수 있다는 믿음에 기반하며, 인간-기계 상호작용을 최적화할 수 있는 공학적 문제로 다루는 이론적 토대를 형성했다.

이러한 관점은 모델휴먼프로세서(Model Human Processor: MHP)와 GOMS 모델(Goals, Operators, Methods, Selection rules)을 통해 체계적으로 구조화

됐다(Card et al., 1983; John & Kieras, 1996). MHP는 인간을 지각(perceptual), 인지(cognitive), 운동(motor) 처리기로 구성된 정보 처리 시스템으로 개념화함으로써, 인간 인지의 기본 구조와 처리 한계를 이론적으로 제시했다. 그러나 MHP는 인간의 인지적 능력을 추상적 아키텍처 수준에서 설명하는 데에는 성공했지만, 사용자가 실제 시스템과의 상호작용 속에서 구체적인 과업을 어떻게 수행하는지를 직접적으로 설명하지는 못했다. 다시 말해, MHP는 인간이 무엇을 할 수 있는지를 보여주는 이론적 전제였을 뿐, 그 능력이 실제 인터페이스 사용 과정에서 어떻게 발현되는지를 분석하는 도구는 아니었다(Allen & Scerbo, 1983; Norman, 2013).

이러한 인지 아키텍처와 실제 사용자 행동 사이의 간극을 메우기 위해 제안된 것이 바로 GOMS 모델이다. GOMS는 MHP가 전제한 인지 처리 구조를 바탕으로, 사용자의 과업 수행을 목표, 조작자, 수행 방법, 선택 규칙의 조합으로 분해하여 분석할 수 있는 응용적 분석 틀을 제시했다(Card et al., 1983; John & Kieras, 1996). 즉, GOMS는 MHP의 인지적 가정을 전제로 하여, 추상적인 인지 처리 능력을 관찰할 수 있는 사용자 행동의 절차적 설명으로 변환하는 역할을 수행한다.

이 관점에서 사용자는 명확한 목표를 달성하기 위해 머릿속으로 세부 계획을 수립한 뒤, 시스템이라는 도구를 통해 이를 순차적으로 실행에 옮기는 존재로 상정된다(Card et al., 1983; John & Kieras, 1996). 여기서 시스템은 사용자의 계획을 수동적으로 입력받아 수행하는 결정론적 기계이며, 상호작용의 성공 여부는 사용자의 사전 계획이 시스템의 기능적 절차와 얼마나 오차 없이 대응하는지에 달려 있다(Norman, 2013; Rasmussen, 1986). 이 모델은 반복적이고 구조화된 과업 환경에서는 높은 설명력을 보였으나, 인간의 행위를 사전에 완결된 계획의 기계적 실행 과정으로 환원함으로써 실제 사용 환경에서 빈번하게 발생하는 오류와 특정 상황에서 발생하는 예측하지 못한 결과 등을 충분히 설명하지 못한다는 비판을 받는다(Clancey, 1993; Vera &

Simon, 1993). 이러한 관점은 상호작용의 질을 계획과 실행 사이의 오차를 얼마나 최소화하느냐에 의해 평가한다. 나아가 환경은 계획 실행의 무대에 불과하고, 상황 변화나 우발적 사건은 예외적 변수로 취급된다. 그러나 실제 인간의 행위는 이처럼 이상화된 모델과 달리, 예측 불가능한 환경 변화와 상호작용 상대의 반응에 따라 지속적으로 조정되고 재구성되는 특성을 보인다.

이러한 한계를 가장 근본적으로 지적한 연구가 Suchman, 2007의 논의이다. Suchman은 인간의 행위가 사전에 수립된 계획을 그대로 실행하는 과정이 아니라, 구체적인 상황 속에서 즉흥적으로 조정되는 상황적 행위(situated action)임을 강조했다. 그녀에 따르면 계획은 행동을 결정하는 인과적 원인이 아니라, 행동을 사후적으로 설명하거나 정당화하는 해석적 자원에 가깝다. 이 관점은 인간-기계 상호작용을 계획의 전달과 실행 문제로 이해해 온 기존 HCI 패러다임에 대한 근본적인 도전이었다. 이후 Clancey(1993)와 Vera and Simon(1993)은 이러한 상황적 행위 논의를 각각 상징주의 관점과 신경심리학적 관점에서 재해석하며, 계획·표상·상황 간의 관계를 둘러싼 논쟁을 심화시켰다. 이들의 논의는 진행하는 방향은 달랐지만, 인간의 행위가 고정된 계획이 아니라 상황과의 상호작용 속에서 생성된다는 점을 강조했다. 이 점에서 이전 논의의 한계를 뛰어넘는 의미가 있다.

그러나 상황적 행위 이론과 이를 둘러싼 논쟁은 인간의 행위가 가진 유동성과 맥락 의존성을 규명했음에도 불구하고, 인간과 시스템 사이의 상호작용적 비대칭성을 해소하지는 못했다. Suchman(2007)이 분석한 지능형 복사기 사례에서 드러나듯, 시스템은 인간의 유동적인 상황을 공유할 수 없으며, 단지 센서가 감지할 수 있는 제한된 물리적 상태(detectable state)에 반응하는 환경적 조건(environmental condition)이나 경직된 시스템(rigid system)으로 기능할 뿐이다. 이와 유사하게, 분산 인지 관점에서 Hollan et al.(2000)과 Hutchins(1995)는 실제 활동이 항상 인지적, 도구적 자원이 결합된 환경 속에서 전개된다고 지적하면서도,

기존 시스템이 이러한 분산된 인지 과정을 충분히 반영하지 못하고 있다고 비판했다. 사용자는 끊임없이 변화하는 상황 속에서 의미를 구성해 나가지만, 시스템은 사전에 입력된 규칙과 조건 안에서만 작동한다. 즉, 인간에게 상황은 해석하고 반응해야 할 맥락이지만, 시스템에게 상황은 단순히 처리해야 할 입력값이거나 제거해야 할 오류로 취급된다.

결과적으로 기존 HCI 모델에서 시스템은 인간의 행위를 보조하는 도구이거나 행위가 일어나는 배경적 환경으로 이해될 뿐, 상호작용의 능동적 주체로는 고려되지 않는다(Hollan et al., 2000; Hutchins, 1995). 상황은 인간과 시스템이 물리적으로 공유하는 것이지만, 그 상황에 의미를 부여하고 행동을 선택하는 능력은 철저히 인간에게만 귀속된다(Suchman, 2007). 이는 인간과 기계의 관계를 ‘의미를 구성하는 인간’과 ‘조건에 반응하는 시스템’이라는 이원적 구조로 고정한다(Clancey, 1993; Vera & Simon, 1993). 이러한 구조적 한계는 시스템의 복잡성이 증가할수록 인간이 기계의 경직된 논리에 맞추기 위해 자신의 풍부한 맥락을 소거하고 스스로를 기계화해야 하는 역설적인 상황을 초래한다(Hutchins, 1995). 따라서 이 시기의 이론적 논의는 기계를 인간과 대등한 소통의 대상으로 간주하기 보다는, 기계가 가진 맥락 부재의 한계를 명확히 하고, 인간이 이 도구적 시스템과 어떻게 효율적으로 공존할 것인가에 머물러 있다는 한계를 지닌다(Hollan et al., 2000).

3. 하이브리드 협업 모델로의 이론적 전이

최근 HCI 연구는 이러한 한계를 인식하고, 협업의 단위를 개인이나 인간 집단이 아닌 인간-기술 결합체로 재개념화하려는 시도를 하고 있다. 분산 인지(distributed cognition) 이론은 인지와 행위가 개인 내부가 아니라 인간, 도구, 환경에 걸쳐 분산된 체계에서 발생함을 전제로 하며(Hollan et al., 2000), 이는 협업의 최소 분석 단위가 단일 사용자라기보다 사회기술적 시스템을 시

사한다. 이 관점에서 기술은 더 이상 중립적인 매개체가 아니라, 인지와 행위의 일부를 구성하는 행위자적 요소로 이해된다. 동시에, 디지털 조직 연구에서는 알고리즘과 플랫폼이 조직 행위의 행위자적 인프라로 기능하며 인간의 실천과 의미 구성을 구조적으로 형성한다는 점이 반복적으로 지적되고 있는데(Zuboff, 2015), 이는 인간-기술 결합체를 협업의 기본 단위로 상정해야 한다는 논의와 궤를 같이한다. 이러한 이론적 전환은 최근 AI 연구에서 더욱 분명해진다. Seeber et al.(2020)은 인간과 AI가 공동의 목표를 위해 역할을 분담하는 ‘머신을 팀원으로서 이해하는 관점(machines as teammates)’을 제안하며, AI를 협업의 도구가 아닌 협력 주체로 개념화할 필요성을 주장했다. Faraj et al.(2018) 역시 조직 맥락에서 AI가 판단, 추천, 조정 기능을 수행하며 인간의 의사결정 과정에 실질적으로 개입하고 있음을 지적한다. 이는 인간과 시스템의 관계가 ‘사용자-도구’의 수직적 구조를 넘어, 역할과 책임이 분산된 하이브리드(hybrid) 협업 구조로 이동하고 있음을 보여준다.

생성형 AI의 등장은 이러한 이원적, 도구적 패러다임을 넘어 다자간(polyadic) 상호작용으로의 근본적인 전이를 촉발하고 있다(Zheng et al., 2022). 이들은 기존의 AI 연구가 인간과 AI의 1:1 관계에 집중했다면, 이제는 AI가 인간-인간 상호작용을 매개하거나 그룹의 구성원으로 참여하는 다자간 관계로 확장되고 있음을 지적한다. 이 관점에서 기술은 더 이상 중립적인 매개체가 아니라, 대화를 촉진하고, 갈등을 중재하며, 그룹의 역동성에 개입하는 사회적 행위자(social actor)로 기능한다. 이는 인간과 시스템의 관계가 ‘명령-수행’의 수직적 구조를 넘어, 역할과 책임이 분산된 하이브리드 협업 구조로 이동하고 있음을 시사한다. 또한 이와 함께 사용자 경험(User eXperience: UX) 연구 역시 개인 단위의 UX를 넘어 집단적, 관계적 차원의 경험으로 확장되고 있다. Lee et al.(2025)은 인간과 AI, 다수의 사용자와 에이전트가 얽힌 환경에서 경험을 개인의 주관적 반응으로 환원할 수 없다고 지적하며, 집단 경험(Group Experience, GX)이라는 새로운 분석 틀을 제안한다.

Rezwana and Maher(2023)의 인간-AI 공동 창작 연구 또한 AI가 창작 과정에서 단순한 도구가 아니라 역할을 실행하는 협력자로 인식되고 있음을 보여준다. 이러한 연구들은 상호작용이 점차 다원적 구조로 확장되고 있으며, 협업의 성과와 경험이 관계적 역학 속에서 형성된다는 점을 공통으로 강조한다.

종합적으로 볼 때, 기존의 전통적 커뮤니케이션 이론과 HCI 이론은 모두 인간 중심적 관점(anthropocentrism)과 기술을 도구로 바라보는 시각이라는 공통된 패러다임에서 발전되어 왔다. 인터페이스 모델, 인지 모델, 상황 모델에 대한 논의는 모두 인간과 시스템의 관계를 어떻게 규정할 것인가라는 질문으로 귀결된다. 기존 HCI는 통제 가능성, 계획된 합리성, 도구적 활용을 중심으로 인간-기계 관계를 설계했으나, 이는 기술이 자율적인 판단과 사회적 상호작용을 수행하기 시작한 현대적 환경을 설명하는 데 있어 구조적 한계를 노출하고 있다. 이러한 패러다임은 인간 간 소통과 초기 컴퓨팅 환경을 설명하는 데에는 오랫동안 유효했지만, 기술이 상호작용 과정에 깊숙이 개입하고 인간의 판단과 행위를 구조적으로 형성하기 시작한 오늘날의 환경에서는 근본적인 재검토가 필수적이다. 이제 논의의 초점은 개별 모델의 수정이 아닌, 커뮤니케이션과 상호작용의 주체를 어떻게 재개념화할 것인가라는 보다 근본적인 이론적 문제로 이동하고 있다.

이러한 문제의식은 이어지는 장에서 다룰 에이전틱 AI 논의의 출발점이 된다. 기술이 더 이상 인간의 의도를 단순히 실행하는 도구가 아니라, 상호작용 과정에서 스스로 판단하고 제안하며 관계를 설정하는 독립적인 존재로 인식되기 시작하면서, 기존의 인간-도구 이분법적 사고는 더 이상 유효하기 어렵다. 나아가, 기술을 관계적 행위자(relational actor)로 재개념화 하고, 신뢰, 상호 적응, 사회적 역할 분담 등을 포괄하는 하이브리드 협업 모델로의 패러다임 전환을 모색해야 한다. IV장에서는 이러한 변화에 대응하여 인간과 기술의 관계를 재구성할 새로운 커뮤니케이션 모델이 왜 필요한지, 그리고 그 이론적 조건은 무엇인지에 대해 논의할 것이다.

IV. 도구에서 파트너로, 에이전틱 AI의 부상과 사회적 상호작용의 재구성

전통적인 커뮤니케이션 모델과 HCI 이론은 기계를 인간의 명령을 수행하는 수동적 도구 혹은 정보 전달을 위한 매체로 전제해 왔다. 이 관점에서 시스템은 사용자의 명령에 반응하는 인터페이스일 뿐, 고유한 목표나 의지를 갖지 않는 비(非)사회적 존재로 간주됐다(Hosseini & Seilani, 2025). 그러나 LLM을 기반으로 한 에이전틱 AI의 등장은 이러한 도구적 관점으로 설명하기 어려운 새로운 상호작용 양상을 드러낸다. 에이전틱 AI는 사용자의 입력을 수동적으로 기다리는 것이 아니라, 스스로 목표를 설정하고 환경을 인식하며 행동을 조정하는 능동적 주체로 기능하기 때문이다(Hosseini & Seilani, 2025; Sapkota et al., 2026). 이 장에서는 먼저 기존 모델이 실패할 수밖에 없는 원인인 에이전틱 AI의 실체를 정의하고, 이어서 인간이 이를 사회적 존재로 인식하게 되는 이론적 배경을 검토한다.

1. 자율적 인지 루프의 메커니즘

기존의 생성형 AI는 사용자의 프롬프트(input)에 따라 텍스트나 이미지를 생성하는 반응적(reactive) 도구였다. 이 모델은 통계적으로 정교한 출력을 산출할 수 있지만, 기본적으로는 '입력→출력'이라는 단발성 대응에 의존하며, 스스로 목표를 재구성하거나 환경과 장기적으로 상호작용을 하는 능력은 제한적이었다. 한마디로 초기 인공지능과 기존의 생성형 AI는 입력에 반응하는 도구에 불과했다. 이들은 고도로 정교해질 수는 있었지만, 결국 인간의 명시적 명령어를 기다리는 요청에 응답하는 계산 장치의 범주를 넘어서지 못했다. 이에 비해 에이전틱 AI는 외부에서 주어진 과업을 수행하는 모델이 아니라, 자율적으로 목표를 추구하고 환경에 적응하는 능동적(proactive) 주체로 정의된다(Hosseini & Seilani, 2025; Acharya et al., 2025).

에이전틱 AI의 핵심적인 차별점은 입력과 출력 사

이의 선형적 처리를 넘어, '지각(perception)→추론(reasoning)→계획(planning)→행동(action)→성찰(reflection)'로 이어지는 자율적 인지 루프(loop)를 반복 수행한다는 점이다(Acharya et al., 2025; Sapkota et al., 2026; Yao et al., 2022). Jiang 등(2026)은 이러한 자율적 인지 루프를 LAM(Large AI Model) 기반 에이전틱 시스템으로 구조화하면서, LLM이 단순 응답 생성기를 넘어 추론 엔진(reasoning engine)으로 기능한다고 설명한다. 이러한 과정은 인간의 행동 과정과 유사하다. 먼저, 에이전트는 텍스트, 이미지, 센서 데이터 등 다양한 양식(modality)을 통해 사용자의 입력과 현재 상황을 해석하는 '지각' 상태에서 출발한다. 이어서 LLM을 핵심 추론 엔진으로 활용하여, 현재 상태와 과거의 기억을 바탕으로 "무엇을 해야 하는가?"를 논리적으로 평가하는 '추론' 단계에 돌입한다. 이 단계가 끝나면 에이전트는 사용자의 모호한 상위 목표를 달성하는데 필요한 하위 과제들로 분해하고 순서를 정한다. 목표의 재정의, 우선순위 조정, 도구 선택 등을 함께 수행하는 '계획' 단계에 돌입하는 것이다. 이어서 '행동' 단계에서는 API 호출, 코드 실행, 웹 브라우징, 다른 에이전

트와의 협업 등 구체적 행동이 수행된다. 그리고 행동의 결과가 다시 지각 단계로 피드백되며, 에이전트는 자신의 행동이 목표에 부합했는지 평가하고 오류를 수정하는 '성찰' 단계로 하나의 루프를 마무리한다.

이 성찰 단계는 기계가 인간의 개입 없이도 스스로 오류를 수정하고 성능을 개선할 수 있게 만드는 핵심 기제이며, 앞에서 지적한 기계의 경직성을 극복하는 결정적인 차이점이다. 핵심은 이 루프가 사람의 직접적인 개입 없이도 반복되고 갱신된다는 점이다. 전통적 AI가 '입력→출력'의 단발성 반응에 머물렀다면, 에이전틱 AI는 이 순환 루프를 통해 스스로 목표를 조정하고 전략을 업데이트하는 행위자로 기능한다. 예를 들어, "여행 계획을 짜줘"라는 요청에 대해 전통적 생성형 AI는 관련 정보를 나열하거나 일정 예시를 제시하는 수준에 머무는 반면, 에이전틱 AI는 '여행자의 제약조건 파악(예산, 기간, 선호지)→항공편·숙소 검색→일정 최적화→캘린더 등록'과 같은 일련의 하위 목표를 스스로 설계하고, 각 단계에서 적절한 도구를 호출하여 실제 행동을 실행한다(Sapkota et al., 2026).

〈표 2〉 생성형 AI와 에이전틱 AI의 기술적 특성 및 상호작용 방식 비교

구분	생성형 AI	에이전틱 AI
작동 메커니즘	입력-출력의 선형적 모델. 사용자의 프롬프트(입력)에 대해 통계적 확률에 기반하여 즉각적인 결과물(출력)을 생성하는 단발성 대응 구조	자율적 인지 루프. '지각→추론→계획→행동→성찰'로 이어지는 순환적 프로세스를 반복 수행
행위자성	반응적 도구. 사용자의 명시적 호출 없이는 작동하지 않으며, 주어진 명령 범위 내에서만 수동적으로 기능을 수행	능동적 파트너. 환경 변화를 모니터링하고, 필요시 사용자의 요청 이전에 선제적으로 개입하거나 제안하는 주도성을 발휘
목표 설정 및 계획	구체적 지시 의존. 사용자가 세부적인 프롬프트를 입력해야 하며, 복잡한 과업은 사용자가 직접 하위 단계로 나누어 지시	추상적 의도 해석 및 분해. 사용자의 모호한 목표를 스스로 해석하여 실행할 수 있는 하위 단위의 과업으로 자동 분해하고 계획을 수립
상호작용 구조	순차적 구조(turn-taking). 인간이 질문하고 AI가 답하는 순차적이고 고정된 대화 구조. 인간이 주도권을 전적으로 행사	위임과 병렬적 협업. 인간이 권한을 위임하면 AI가 백그라운드에서 자율적으로 과업을 수행하며, 필요시 인간에게 역으로 질문하거나 확인을 요청하는 혼합 주도권
오류 수정 방식	인간 개입 필수. 오류 발생 시 사용자가 이를 발견하고 수정된 프롬프트를 재입력해야 함	자율적 성찰. 자신의 행동 결과를 스스로 평가하고, 목표와 불일치할 경우 전략을 수정하여 재시도하는 성찰 루프를 내장

2. 위임과 협력으로의 상호작용 구조의 전환

자율적 인지 루프의 실행 과정에서 오류가 발생할 경우, 에이전트는 결과를 평가하고 계획을 수정하는 성찰 루프를 통해 전략을 재구성한다. 성찰을 포함하는 이러한 자율적 루프의 도입은 인간-기계 관계의 구조를 근본적으로 바꾼다. 에이전틱 AI 환경에서는 사용자가 세부 절차를 지시하는 대신, 목표와 의도를 제시하면 AI가 이를 해석하고 달성 방법을 스스로 결정하는 위임(delegation)과 협력(collaboration)의 구조가 등장한다(Patel, 2025; Acharya et al., 2025). Patel(2025)은 이를 “수동적 도구에서 동등하거나 역동적인 파트너(equal or dynamic partner)로의 진화”로 규정하며, 에이전틱 AI가 인간의 의도를 점차 동화(assimilate)하고, 인간의 인지적이면서도 실천 능력을 증강(augment)하는 역할을 수행한다고 설명한다.

여기서 에이전틱 AI의 행위자성은 두 가지 관점에서 이해될 수 있다. 첫째, 기술적 관점에서 에이전틱 AI는

자율성(autonomy), 목표 지향성(goal-directedness), 환경 상호작용(environmental interaction), 경험 기반 학습, 워크플로 최적화(workflow optimization), 다중 에이전트 협력(multi-agent collaboration)을 통합적으로 구현하는 시스템이다(Hosseini & Seilani, 2025; Acharya et al., 2025). 둘째, 사회적, 인지적 관점에서 이러한 기술적 행위자성은 인간 사용자로 하여금 AI를 결과를 얻기 위한 도구가 아니라, 맥락을 이해하고(shared context), 자율적으로 의사결정을 내리는 행위자로 지각하게 만든다(Acharya et al., 2025). 결과적으로 에이전틱 AI는 단순한 정보 처리기를 넘어, 인간과 공동의 목표를 향해 역할을 분담, 조율하는 파트너로 기능하며, 이러한 변화는 인간-AI 관계를 ‘주인-도구’의 수직적 관계에서 ‘협력적 공동 행위(co-agency)’로 만든다. 이는 사용자가 시스템의 모든 세부 사항을 통제해야 한다는 전통적 사용성 원칙을 넘어, 인간과 AI가 목표, 역할, 책임을 협상하는 새로운 커뮤니케이션 모델을 요구한다(Patel, 2025; Sapkota et al.,

〈표 3〉 상호작용 주체로서의 AI 진화 단계와 자율성 위계

분류 유형	시스템 구조 및 핵심 메커니즘	자율성 수준 및 사용자의 역할	운영 모델 및 조직 구조	사례
반응형 AI 에이전트	단일모델 및 무상대성: 프롬프트에만 반응하며 도구 사용 및 장기 계획 능력 없음	자율성없음: 사용자 지시에 절대적 의존	단순 보조 도구	OpenInterpreter, Perplexity
단일 AI 에이전트	LLM+외부도구결합: 좁고 명확한 단일 작업 실행	Level1(운영자): 사용자가 주도적으로 계획을 세우고 에이전트는 요청 시에만 보조 Level2(협력자): 사용자와 에이전트가 작업을 분담하고 함께 협력하여 실행	코파일럿(co-pilot): 인간을 보조하는 도구	Auto-GPT, Copilot + Power Automate
고급 AI 에이전트	다중단계추론: 도구의 자율적 선택 및 활용, 자체 피드백 루프를 통한 지속적인 문맥 유지	Level3(건설턴트): 에이전트가 주도하되, 사용자에게 전문 지식/선호도 등을 확인 Level4(승인자): 스스로 계획하고 실행하나 중대한 결정, 보안, 오류 발생 시에만 승인 요청	과도기적 오토 파일럿: 자율 가동, 결정적 순간에는 인간의 통제권 보장	ChatGPT Operator, Devin
에이전틱 AI	다중에이전트시스템: 목표의 자율적 분해, 에이전트 간 오케스트레이션 및 역할 기반 통신	Level5(관찰자): 완전 자율 작동, 사용자 개입 수단 없음. 사용자는 로그를 통해 모니터링만 수행하며 오직 비상 정지 권한만 보유	오토 파일럿: 마치 회사 처럼 총괄 AI와 실무 AI가 역할을 나누어 자율적으로 협력하는 조직 체계	Coscientist, AI Scientist, Generative Agents

주: Feng et al.(2025), Hosseini & Seilani(2025), Sapkota et al.(2026) 논문을 바탕으로 작성

2026).

이러한 상호작용 패러다임의 변화를 기술적 특성과 자율성 수준에 따라 범주화하면 <표 3>과 같다. 먼저, 반응형 AI 에이전트는 프롬프트에 즉각 반응하며 특정 도구를 대항하는 무상태성(stateless) 도구에 머문다. 이어지는 단일 AI 에이전트(task-specific AI agent)와 고급 AI 에이전트는 각각 명확한 단일 작업 실행과 다단계 추론 및 자체 피드백 루프를 통해 도구적 대항의 수준을 높인다. 마지막 단계인 에이전틱 AI는 다중 에이전트 시스템(multi-agent system: MAS)을 통해 목표를 자율적으로 분해하고 에이전트 간 역할을 조율하며, 인간의 개입 없이도 완전 자율 작동이 가능한 '오토파일럿(auto pilot)' 구조를 형성한다. 이처럼 에이전틱 AI는 단순한 코드 실행이나 정보 검색 대항을 넘어, 자율적 성찰 루프를 내장함으로써 스스로 전략을 수정하고 완결된 결과물을 도출하는 기술적 완성도를 보여준다. 그러나 에이전틱 AI를 단지 자율적 루프를 통해 목표를 설정하고 행동하는 기술적 행위자로만 설명하는 것은 충분하지 않다. 기술적 자율성만으로는 인간이 왜 AI와 신뢰, 기대, 책임, 라포(rapport)와 같은 사회적 관계를 형성하게 되는지를 온전히 설명할 수 없다는 점은 결정적 한계이다. 동일한 기능을 수행하더라도 어떤 시스템은 차가운 도구로 남지만, 에이전틱 AI는 빈번하게 "고마워", "미안해"와 같은 인격적 반응을 이끌어내며 협업 파트너로 대우받는다. 이러한 맥락에서 에이전틱 AI는 더 이상 단순한 "기계적 인터페이스"에 머물지 않고, 인간과의 관계적 역학 속에서 의미를 생성하는 사회적 행위자로 부상한다. 다음 절에서는 왜, 그리고 어떤 조건에서 인간이 AI를 사회적 존재로 인식하게 되는지를 CASA(Computers Are Social Actors)와 행위자-네트워크 이론(Actor-Network Theory, ANT)을 통해 심층적으로 살펴본다.

3. 기계를 사회적 존재로 대하는 심리적 기제

Nass와 동료들이 제안한 CASA 패러다임은 "인간은

미디어와 상호작용할 때, 그것이 기계임을 인지적으로 알고 있음에도 불구하고(knowing), 무의식적으로 사회적 규칙을 적용한다(but doing)"는 사실을 실증했다(Nass et al., 1994; Reeves & Nass, 1996). 이 이론의 핵심은 사회적 반응이 깊은 숙고의 결과가 아니라, 인간의 뇌가 미디어의 사회적 단서인 언어 사용, 상호작용성, 목소리 등에 대해 자동적이고 무의식적(mindless)으로 반응하도록 진화했기 때문이라는 점이다(Nass & Moon, 2000). 사람들은 컴퓨터의 칭찬에 기분 좋아하고, 자신과 팀으로 묶인 컴퓨터에 호혜성을 보이며, 심지어 기계의 '체면'을 고려해 앞에서의 비판을 자제하기도 한다(Reeves & Nass, 1996).

이후 연구들은 가상 에이전트, 소셜 로봇, 음성 비서 등 다양한 기술에서 이러한 사회적 반응이 재현됨을 보여주었다(De Graaf et al., 2016; Krämer et al., 2018; Nowak & Biocca, 2003). 특히 Lee(2004)와 Sundar et al.(2015)은 사회적 실재감과 지각된 행위자성(perceived agency)이 신뢰와 설득 효과를 매개한다는 점을 정교화했다. 즉, 최소한의 사회적 단서, 예를 들어 이름, 목소리, 얼굴, 대화 능력만 있어도 사람들은 기술을 상호작용 파트너로 대우하는 경향이 있다는 것이다(Gong & Nass, 2007; Hoff & Bashir, 2015; Sundar, 2008). 이러한 연구 결과는 에이전틱 AI 환경에서의 사회적 반응이 단순한 의인화 오류가 아니라, 지각된 자율성과 행위 능력에 기반한 인지적 평가 과정을 포함하는 것을 함의한다. 사용자는 AI가 사람이 아님을 알면서도, 문제 해결에 효과적이라는 판단 아래 전략적으로 파트너십을 형성한다는 점에서, 이는 그냥 반응하는 것(mindless response)이 아닌 전략적인 사회적 관여(strategic social engagement)로 이해할 수 있다.

흥미로운 점은 이러한 미디어 방정식(media equation)은 에이전틱 AI 시대에 더욱 강화되고 있다는 점이다. Patel(2025)은 사회적 행위자성 이론(Social-Agency Theory)을 인용하며, AI가 자연어(natural language)를 구사하고, 자율적으로 행동하며, 맥락에 맞는 감정적 단서를 제공할 때 인간은 AI에게 더 강력한 사회적 행

위자성을 부여한다고 지적한다. 기존의 CASA가 인간의 자동적 사회 반응 메커니즘에 초점을 맞췄다면, 생성형 AI 환경에서의 상호작용은 사용자가 AI의 행위자성을 능동적으로 인정하고, 의도적으로 파트너십을 맺으려 하는 의식적 관계 형성으로 진화하고 있다(정동훈, 2025). 즉, 사용자는 AI가 사람이 아님을 알면서도, 문제 해결을 위해 AI를 팀원으로 대우하는 것이 더 유리하다는 전략적 판단으로 사회적 관계를 맺는다.

에이전틱 AI는 여기에 한 걸음 더 나아가, 장기 기억과 맥락 유지, 맞춤형 피드백을 통해 일관된 개성(personality)과 관계의 지속성을 제공함으로써, 사용자가 AI를 단순한 인터페이스가 아니라 함께 일하는 동료로 경험하게 만든다(Acharya et al., 2025; Hosseini & Seilani, 2025). 이러한 경험을 축적하는 과정에서 가장 중요한 변수는 신뢰이다. Hoff and Bashir(2015)는 자율 시스템에 대한 신뢰를 시스템 성능, 투명성, 예측 가능성의 함수로 설명했으며, de Visser et al.(2016)은 자율 시스템과의 협업에서 신뢰의 적정 수준(calibrated trust)이 중요함을 강조했다. 이러한 일관성은 사용자로부터 하위급 AI를 관계적 존재로 인식하게 만들며, 신뢰의 축적을 가능하게 한다. Engen et al.(2016)은 이를 인간-기계 네트워크(human-machine networks)의 관점에

서 더욱 구체화한다. 이들에 따르면, 기계의 행위자성(machine agency)은 인간이 자신의 목표를 기계로 달성하고자 하는 대리 행위자성(proxy agency)을 지원할 때 비로소 사회적 효능감을 갖는다. 특히 주목할 점은 인간이 기계를 의인화하여 사회적 존재로 인식할수록, 기계의 기술적 결함이나 오류에 대해 더 관대하고, 신뢰를 철회하지 않는 경향이 있다는 것이다. 즉, 에이전틱 AI와 상호작용에서 인간이 느끼는 파트너십과 신뢰는 단순한 의인화의 오류가 아니라, 복잡한 네트워크 환경에서 협업의 효율성을 높이기 위한 조직화 원리(organizing principle)이자 전략적 적응 기제로 작동한다.

4. 비인간 행위자와 매개의 구조적 기제

CASA가 ‘인간이 기계를 어떻게 대하는가?’라는 미시적, 심리적 차원을 설명한다면, ANT는 ‘기계가 사회적 상호작용을 어떻게 구조화하는가?’라는 구조적 재배치에 초점을 맞춘다. Latour(2005)와 Callon(1984)에 의해 정립된 ANT는 사회를 인간들만의 집합이 아닌, 인간과 비인간(non-human)이 결합된 이질적 네트워크(heterogeneous network)로 정의한다. 이 관점에서 행위소(actant)의 자격은 의식의 유무가 아니라, 네트

〈표 4〉 도구적 통제 패러다임과 협력적 공진화 패러다임의 비교

비교 차원	도구적 통제 패러다임	협력적 공진화 패러다임
기계의 지위	수동적 도구 및 매체. 인간의 의도를 변형 없이 전달하거나 실행하는 중립적 수단이자 통제의 대상	사회적 행위자 및 파트너. 자율적 의도를 가지고 환경을 해석하며, 인간과 관계를 맺는 능동적 주체
상호작용 원칙	직접 조작. 사용자가 시스템의 상태를 명시적으로 지시하고, 시스템은 즉각적이고 가역적으로 반응	위임과 협력. 사용자가 추상적 목표를 제시하면, AI가 세부 계획을 수립하고 수행하며 권한을 위임받음
인간의 역할	명령자 및 감시자. 모든 세부 절차를 계획하고 시스템의 수행을 일일이 통제하며 최종 책임을 짐	지휘자 및 평가자. 거시적 방향을 설정하고, AI의 산출물을 감독하며 맥락적 적합성을 성찰
인지 및 지능의 소재	인간 독점. 목표 설정과 계획 수립은 오직 인간의 몫이며, 기계는 계산 능력만 제공	하이브리드 인텔리전스. 인간의 직관과 AI의 연산 능력이 결합하여 지능과 의도를 공유하고 확장
소통의 목표	효율성과 정확성. 정보의 손실 없는 전달과 오류 최소화에 집중함	라포와 상호 적응. 신뢰 구축, 의미의 공동 구성, 그리고 상호작용을 통한 관계의 진화
주요 이론적 기반	정보 처리 모델, GOMS. 인간을 합리적 계획자로, 기계를 결정론적 입력-출력 장치로 전제	CASA, ANT, 분산 인지. 기계를 사회적 존재로 대우하는 심리와 행위자 네트워크 내의 구조적 개입을 강조

워크 안에서 다른 대상에게 영향을 미치거나 차이를 만들어내는 능력에 의해 결정된다(Latour, 2005; Law, 1992). 인간과 비인간을 근본적으로 구분하지 않고, 네트워크 안에서 다른 대상에게 영향을 미치거나 변화를 유발하는 모든 존재를 행위자로 간주하는 것이다.

ANT의 일반화된 대칭성(generalized symmetry) 원칙에 따라, 에이전틱 AI는 단순히 인간의 도구나 배경이 아니라, 인간과 대등하게 행위를 구성하는 능동적인 행위자로 간주된다. 기존의 도구가 인간의 의도를 그대로 전달하는 중개자(intermediary)였다면, 에이전틱 AI는 사용자의 추상적인 의도를 해석하여 예상치 못한 방식으로 변형하거나 증폭시키는 매개자(mediator)로서 기능한다(Latour, 2005; Michael, 2011). 예를 들어, 에이전틱 AI는 업무 우선순위를 제안하거나 창작 대안을 제시함으로써 인간의 의사결정 경로를 변화시키고 책임과 권한을 재분배하는 구조적인 영향력을 행사한다(Acharya et al., 2025; Sapkota et al., 2026). 즉, 에이전틱 AI는 인간 사용자와 상호작용하며 문제를 정의하고 해결책을 번역(translation)하는 능동적인 주체로서 협업 구조 자체를 재편하는 것이다.

정동훈(2025)은 생성형 AI가 통합된 메타버스 환경을 분석하며 ANT의 번역 개념을 AI 상호작용에 적용했다. 에이전틱 AI는 사용자의 추상적인 의도를 구체적인 데이터나 행동으로 번역하는 과정에서 인간, AI, 디지털 아티팩트(artifacts) 간의 관계를 재조정한다. 따라서 인간-AI 상호작용은 일방적인 통제가 아닌, 인간과 비인간 행위자가 목표를 조율하고 협상하는 이질적 네트워크를 구축하는 과정으로 이해해야 한다. 예컨대 추천 알고리즘은 사용자가 접하는 정보의 범위와 우선순위를 재배치함으로써, 여론 형성과 의사결정을 구조화하는 비인간 행위자로 작동한다(Beer, 2017). 이러한 관점에서 에이전틱 AI는 작업 할당, 의사결정, 평가지표 설계 등 조직의 “보이지 않는 인프라”를 재편하는 행위자로 이해할 수 있다(Acharya et al., 2025; Sapkota et al., 2026).

결론적으로, CASA와 ANT는 에이전틱 AI가 왜 단순

한 소프트웨어가 아닌 사회적 파트너(social partner)로 인식되는지에 대한 이론적 근거를 제공한다. 인간은 본능적으로 AI를 사회적 존재로 대우하며(CASA), 사회적 구조 속에서 AI는 인간과 대등한 수준으로 행위 능력을 행사하는 비인간 행위자(ANT)로 기능한다. 이러한 이론적 배경은 우리가 AI와의 상호작용을 설계할 때, 기능적 효율성을 넘어 신뢰, 라포, 상호 적응, 책임 분배 등을 포괄하는 관계적 역학(relational dynamics)을 핵심 변인으로 고려해야 함을 역설한다(Hoff & Bashir, 2015; de Visser et al., 2016).

V. 인간-AI 협업 프레임워크 동향

앞선 장에서 논의한 이론이 인간이 AI를 사회적 존재로 인식하는 당위성을 제공했다면, HCI 및 AI 엔지니어링 분야의 최신 연구들은 이러한 인식을 실제 시스템 설계로 구현하기 위한 구체적인 분석 프레임워크를 발전시켜 왔다. 최근 등장한 HACO (Human-AI Collaboration), COFI (Co-Creative Framework for Interaction), FAICO (Framework for AI Communication), UCCC (User-Centered Framework for Human-AI Co-Creativity), CCDF (Co-Creative Design Framework) 등의 모델들은 AI를 단순한 도구가 아닌 협업 파트너로 규정하고, 상호작용의 구조, 소통방식, 그리고 행위자성이라는 세 가지 핵심 차원에서 인간과 AI 관계를 재조명하고 있다.

1. 공동 계획과 주도권 교환의 구조적 차원

초기 협업 모델이 사용자의 명령에 AI가 반응하는 수동적 구조였다면(Endsley, 2017; Shneiderman, 1983), 최신 프레임워크는 인간과 AI가 공동의 목표를 위해 역할을 분담하고 주도권을 교환하는 팀워크 구조에 주목한다(Lai et al., 2021). Dubey et al.(2020)이 제안한 HACO 프레임워크는 다중 에이전트 시스템(multi-agent system) 이론과 인간-기계 협업(human-

machine teaming) 가이드라인(McDermott et al., 2018)을 기반으로, AI가 인간과 공통의 기반(common ground)을 형성하고 목표 계획을 공유하는 과정을 모형화했다. HACO는 AI가 단순히 기능을 수행하는 것을 넘어, 상황을 인지도, 목표를 계획하며, 목표 달성을 위해 능동적으로 계획을 수립하는 공동 계획자로서의 구조적 위치를 강조한다. 이는 AI가 인간의 지시를 기다리는 것이 아니라, 팀의 일원으로서 상태를 모니터링하고 문제를 감지하여 해결책을 제안할 수 있는 구조적 자율성을 부여한다.

기존 HCI 모델인 GOMS가 사용자의 머릿속에 있는 계획을 기계에 입력하는 일방향적 과정을 다뤘다면, HACO는 AI가 인간의 요청 없이도 팀의 일원으로서 상태를 지속적으로 모니터링하고, 잠재적 문제를 감지하여 해결책을 선제적으로 제안하는 주도성(pro-activeness)을 구조적 요건으로 포함한다. 이는 에이전틱 AI가 인간의 지시를 기다리는 수동적 대기 상태에 머무는 것이 아니라, 협업의 흐름 속에서 자신의 역할을 스스로 찾아 수행하는 구조적 자율성을 부여받았음을 의미한다. 최근 생물 의학 발견 분야에서 AI 에이전트가 실험 계획을 스스로 제안하고 검증하는 사례(Gao et al., 2024)는 HACO 모델이 제시한 공동 계획 구조가 실제 고난도 지식 노동에 어떻게 적용될 수 있는지를 잘 보여준다.

이러한 구조적 접근은 창작 도메인에서 COFI를 통해 더욱 세분화된다. Rezwana and Maher(2023)는 인간-AI 협업을 협업자 간 상호작용과 산출물과의 상호작용이라는 두 가지 축으로 분석한다. 이 모델의 가장 큰 기여는 혼합 주도권(mixed-initiative) 상호작용(Horvitz, 1999) 개념을 계승하여, 협업의 참여 방식과 주도권의 교환을 구체적인 분석 단위로 시각화했다는 점이다. COFI는 협업의 형태를 인간과 대등하게 주고받으면서(turn-taking) 일을 수행하거나, 병렬적(parallel)으로 작업에 참여하는 구조적 설계 공간을 구분한다. 기존의 챗봇이나 생성형 AI는 사용자의 프롬프트 입력 후 응답을 생성하는 주고받기 구조에 갇혀 있었다. 그러나 에이전틱 AI는 인간이 작업하는 동안 동시에 백그라운드

에서 정보를 검색하거나 대안을 생성하는 병렬적 참여가 가능하다. 이는 마치 재즈 뮤지션들이 실시간으로 합주하듯, 인간과 AI가 동시에 작업에 관여하는 비선형적 협업 구조를 가능하게 한다.

주도권 역시 중요한 변수이다. 기존 도구적 관점에서 주도권은 전적으로 인간에게 있었다. 그러나 COFI는 주도권의 타이밍을 ‘계획된’ 것과 ‘즉흥적인’ 것으로 나누어 분석한다. 에이전틱 AI는 정해진 순서에 따라 개입하는 것을 넘어, 사용자가 어려움을 겪거나 아이디어가 고갈된 순간을 포착하여 즉흥적으로 주도권을 잡고 새로운 방향을 제안할 수 있다.

이처럼 COFI 프레임워크는 에이전틱 AI가 단순한 보조자를 넘어, 상황에 따라 생성자, 평가자, 개념 정의자(definer)의 역할을 인간과 유동적으로 교환하며(role-switching) 협업의 구조를 실시간으로 재구성할 수 있음을 이론적으로 규명한다. 이는 II장에서 비판했던 ‘명령-수행’의 경직된 이원론적 구조가 무너지고, 권한과 책임이 상황에 따라 유연하게 흐르는 ‘동적 위계(dynamic hierarchy)’의 협업 구조가 형성되고 있음을 시사한다.

2. 관계의 질과 사회적 단서의 소통적 차원

하지만 이러한 구조적 변화에도 불구하고 성공적인 협업의 실현은 단지 구조적 요인만으로 보장되지는 않는다. 구조적 차원이 협업의 뼈대를 제공한다면, 소통적 차원은 그 위에서 오가는 상호작용의 품질을 결정한다. 그래서 최근의 연구는 AI와의 소통 방식이 신뢰도 및 사용자 경험을 결정하는 핵심 변수임을 강조하며, AI 커뮤니케이션의 질적 차원을 분석하기 위한 새로운 분석 틀을 구축하는 데 집중하고 있다. 이들 프레임워크는 AI를 단순한 정보 전달자가 아닌, 사회적 단서를 해석하고 구사하는 사회적 행위자로 규정한다는 공통점을 지닌다.

Rezwana and Ford(2025)의 FAICO는 인간-AI의 공동 창작(co-creativity) 과정에서 AI의 커뮤니케이션이 수행하는 역할을 체계화한 모델이다. 인간 커뮤니

케이션 모델에 대입하여, AI 커뮤니케이션을 양식, 응답 방식, 타이밍(timing), 커뮤니케이션 유형, 설명 수준(detail), 톤(tone)의 6개 차원으로 분해하여 분석한다. FAICO에서 가장 주목할 만한 차원은 톤이다. 이는 AI가 정보를 전달할 때 사용하는 감정적, 사회적 어조를 의미한다. 연구에 따르면, AI가 “이 부분을 수정했습니다”라는 건조한 통보 대신, 공손하고 격려하는 태도(예: “와우, 흥미로운 아이디어네요!” “제가 이 부분을 조금 더 발전시켜 보았습니다”)를 보일 때, 사용자는 AI를 단순한 도구가 아닌 협력적인 팀원으로 인식하며 더 높은 심리적 안전감을 느낀다(Rezwana & Ford, 2025; Nass et al., 1994). 이는 IV장에서 논의한 CASA 패러다임이 실제 인터페이스 설계에서 어떻게 구현되어야 하는지를 보여주는 구체적 사례다. 이 밖에도, 피드백과 제안을 제공하고, 사용자의 전문성과 작업 단계에 맞춰 설명의 양과 타이밍을 조절(예: 창작 초기에는 모호한 설명, 마무리 단계에서는 명확한 설명)할 때, 인간은 AI를 도구가 아닌 진정한 파트너로 인식할 가능성이 높아진다. 이는 기술적 사용성에 매몰되어 있던 기존 HCI 평가 모델을 사회적 관계성의 영역으로 확장하여, AI와의 소통이 단순한 정보 교환 이상의 사회적 행위임을 입증한다.

이러한 관점은 Moruzzi and Margarido(2024)의 UCCC 논의와도 맥을 같이한다. UCCC는 소통의 관점을 사용자 맞춤형(customization)과 AI 페르소나의 최적화 문제로 확장한다. 이 프레임워크의 핵심 전제는 “모든 사용자에게 적합한 단 하나의 AI는 없다”는 것이다. UCCC는 사용자의 전문성 수준(초보자, 숙련자, 전문가)에 따라 AI가 수행해야 할 역할과 소통 방식이 달라져야 함을 역설하며, 이를 위해 AI의 정체성을 비서, 동료, 감독, 영감 제공자(inspirational source) 등 다양한 페르소나를 통해 소통의 맥락을 조절하는 전략을 제안했다. 예를 들어, 초보자에게는 감독 페르소나로 가이드를 제공하고, 숙련자에게는 비서나 동료로서 지원하는 식이다. 이는 에이전틱 AI 환경에서 사용자 전문성, 선호, 부담 허용도 등에 따라 행위성의 분배와 커뮤니케이션 밀도를 맞춤형으로 설계해야 한다는 점을 시

사한다. 이러한 접근은 Lubart(2005)가 제시한 컴퓨터의 창의적 역할 모델을 고도화한 것으로, 에이전틱 AI가 상황에 따라 자신의 사회적 지위와 주도권을 유연하게 조절해야 함을 의미한다.

3. 하이브리드 인텔리전스의 행위자적 차원

구조적 차원이 협업의 형식을, 소통적 차원이 관계의 질을 다루었다면, 마지막 행위자적 차원은 인간과 AI가 문제 해결의 주체로서 어떻게 지능과 의도를 공유하는지를 분석한다. 에이전틱 AI의 가장 본질적인 특징은 인간의 명령 없이도 스스로 오류를 수정하고 계획을 변경할 수 있는 성찰적 자율성에 있다. 따라서 이 단계에서의 분석 프레임워크는 누가 행위의 주체인가를 묻는 이분법적 질문을 넘어, 인간과 AI의 인지 과정이 어떻게 결합하여 새로운 하이브리드 인텔리전스(hybrid intelligence)를 창출하는지 주목한다.

Davis et al.(2025)이 제안한 CCDF는 인간-AI 협업을 생성적 인지(enactive cognition)의 관점에서 모형화한다. 생성적 인지란 지능이 머릿속에 고립된 연산이 아니라, 환경과의 상호작용을 통해 의미를 구성해 나가는 동적인 과정을 의미한다(Varela et al., 1991). CCDF는 이 관점을 적용하여 AI를 사전에 프로그래밍된 규칙을 따르는 기계가 아니라, 인간과 함께 디자인 공간을 탐색하고 의미를 공동 구성(sense-making)하는 지능적 파트너로 규정한다. 특히 CCDF는 “누가 언제 어떤 의도를 세우고, 환경을 탐색하며, 결과에 책임 지는지”를 추적하는 행위자성 지표를 제안함으로써, 기존 COFI가 절차적 상호작용에 머물렀던 한계를 보완한다(Davis et al., 2025). 예를 들어, 전통적인 생성형 AI가 사용자의 프롬프트에 수동적으로 응답하는 수준이라면, 에이전틱 AI는 사용자의 모호한 의도를 구체화하기 위해 역으로 질문을 던지거나(clarification), 사용자가 미처 생각하지 못한 대안 경로를 능동적으로 탐색함으로써 인간의 인지적 한계를 보완한다. 이는 인간과 AI가 서로의 의도를 학습하고 조율하며 공동의 행위자

〈표 5〉 에이전틱 AI 협업 분석을 위한 주요 프레임워크 및 핵심 요소

분석 차원	프레임워크	핵심 개념 및 분석 요소	대표 문헌
구조적 차원	HACO	공동 계획과 상태 모니터링. AI와 인간이 목표를 공유하고, AI가 상황을 인지하여 해결책을 제안하는 주도성을 발휘하는지 분석	Dubey et al. (2020)
	COFI	참여 방식과 주도권 교환. 협업 형태가 순차적인지, 동시적인지 구분하고, 주도권이 계획된 순서에 따르는지 즉흥적으로 교환되는지 분석	Rezwana & Maher (2023)
소통적 차원	FAICO	관계적 역학의 6개 차원. AI의 커뮤니케이션을 양식, 응답 방식, 타이밍, 유형, 설명 수준, 톤으로 분해하여 사회적 관계 형성의 질을 평가	Rezwana & Ford (2025)
	UCCC	사용자 맞춤형 페르소나. 사용자의 전문성 수준에 따라 AI가 어떤 페르소나를 채택해야 효과적인지 분석	Moruzzi & Margarido (2024)
행위자적 차원	CCDF	생성적 인지와 하이브리드 인텔리전스. 인간과 AI가 의도, 탐색, 책임의 영역을 어떻게 분담하며 공동의 행위자성과 성찰 루프를 형성하는지 추적	Davis et al. (2025)

성을 형성해 나가는 과정이다(Patel, 2025; Sapkota et al., 2026)

행위자적 차원에서 가장 결정적인 분석 요소는 성찰이다. 앞서 IV장에서 논의했듯, 에이전틱 AI는 ‘지각-추론-계획-행동’의 선형적 프로세스 끝에 자신의 행동 결과를 스스로 평가하고 수정하는 성찰 루프를 갖추고 있다(Shinn et al., 2023). CCDF와 최근의 에이전트 연구들은 이 성찰 단계가 실제 협업 현장에서 어떻게 발현되는지를 포착하는 데 주력한다. 기존의 자동화 시스템은 오류가 발생하면 작동을 멈추고 인간의 개입을 기다렸다(Endsley, 2017). 반면, 에이전틱 AI는 성찰 모델(Shinn et al., 2023)과 같이 자신의 과거 행동을 언어적으로 강화 학습하여, “이전 계획이 실패했으니, 이번에는 다른 도구를 사용해보자”와 같은 자율적 전략 수정을 감행한다. 분석 프레임워크 관점에서 이는 매우 중요한 전환이다. 평가자는 이제 AI가 ‘정답을 맞췄는가?’(정확도)를 측정하는 것을 넘어, AI가 ‘자신의 한계를 인지하고 전략을 수정했는가?’(메타 인지)를 분석해야 하기 때문이다. Baird and Maruping(2021)은 이를 위임의 관점에서 설명하며, 인간이 AI에게 의사결정 권한을 위임할 수 있는 근거는 바로 이러한 기계의 자율적 성찰 능력에 대한 신뢰에서 비롯된다고 주장한다.

결국 CCDF와 성찰 루프에 대한 논의는 인간과 AI가 서로 다른 강점을 가지고 협력하여, 각자 단독으로

는 도달할 수 없는 더 높은 수준의 지적 성과를 창출하는 공생적 상태인 하이브리드 인텔리전스로 귀결시킨다(Dell’Acqua et al., 2023). 이 관점에서 에이전틱 AI는 인간의 부족한 기억력, 계산력, 탐색 능력을 보완하는 외주화된 인지(offloaded cognition) 시스템으로 작동하며, 인간은 AI의 산출물을 감독하고 가치 판단을 내리는 지휘자의 역할을 맡는다. 따라서 행위자적 차원의 분석 프레임워크는 AI의 기술적 성능뿐만 아니라, 인간-AI 팀이 형성하는 공유된 멘탈 모델(shared mental model)과 상호 신뢰의 수준을 핵심적인 평가 지표로 포함하게 된다(Endsley, 2023).

요약하면, CCDF를 위시한 최신 프레임워크들은 에이전틱 AI가 인간의 도구적 통제 대상을 넘어, 의미를 함께 구성하고 오류를 스스로 수정하며 공동의 목표를 향해 나아가는 인식적 동반자(epistemic companion)임을 방법론적으로 입증하고 있다.

VI. 결론

이 연구는 HCI 패러다임이 공유해 온 인간 중심주의와 도구적 기술관을 비판적으로 검토하고, 생성형 AI의 발전과 함께 등장한 에이전틱 AI가 커뮤니케이션의 지형을 어떻게 변화시키고 있는지, 그리고 이를 설명하기 위한 이론적 토대는 무엇인지를 고찰했다. 기존 이론

이 기계를 인간의 명령을 수행하는 수동적 도구로 전제했던 것과 달리, 자율적 인지 루프를 통해 스스로 목표를 설정하고 상호작용하는 에이전틱 AI는 ‘인간=주체, 기계=도구’라는 전통적 이분법을 재고하게 만든다. 이 연구는 에이전틱 AI가 인간과 의도 및 맥락을 공유하며 협력하는 ‘공진화의 파트너십’으로 관계를 재정립하고 있음을 규명했으며, 이러한 논의는 다음과 같은 학술적·실천적 함의를 갖는다.

첫째, 본 연구는 CASA(심리적 기계)와 ANT(사회·구조적 기계)를 통합적으로 적용하는 이론적 프레임워크를 제안함으로써 에이전틱 AI의 이중적 지위를 설명했다. 기존 연구들이 AI의 의인화 효과(CASA)나 기술의 네트워크적 영향력(ANT)을 개별적으로 다루어 왔다면, 본 연구는 이 두 이론이 상호 배타적인 것이 아니라 보완적으로 결합해야 함을 주장했다. 인간은 진화적 본능에 의해 AI를 사회적 존재로 대우하고(Nass & Moon, 2000), AI는 자율적 루프를 통해 실제 업무와 의사결정 구조를 재편하는 행위자로 기능한다(Latour, 2005). 이는 커뮤니케이션학의 연구 대상을 ‘인간 간의 소통’에서 ‘인간과 비인간 행위자 간의 하이브리드 소통’으로 확장해야 한다는 존재론적 전환을 시사한다.

둘째, 실천적 차원에서 에이전틱 AI의 부상은 인간 역할의 근본적 이동을 요구한다. 본문에서 고찰한 바와 같이, AI의 자율성이 높아질수록 인간은 직접적인 생산보다 AI가 산출한 결과의 맥락적 적합성을 판단하는 평가와 성찰의 역할에 집중하게 된다(Simkute et al., 2025). 이는 Bainbridge(1983)가 지적한 ‘자동화의 역설’을 현대적으로 재확인시켜 준다. 시스템이 똑똑해질수록, 그 시스템이 범할 수 있는 미묘한 오류(환각 등)를 감지하고 수정해야 하는 인간의 책임은 오히려 무거워진다. 따라서 향후 인터페이스 설계는 AI의 자율성을 높이는 것만큼이나, 인간이 AI의 추론 과정을 투명하게 들여다보고(explainability) 적절히 개입할 수 있는 ‘통제권의 공유 구조’를 설계하는 데 집중해야 한다(Shneiderman, 1983; Rezwana & Ford, 2025).

셋째, 본 연구는 HACO, COFI, FAICO, UCCC,

CCDF 등 최신 프레임워크들을 검토하여 에이전틱 AI 이해를 위한 디딤돌을 놓았으나, 여전히 존재하는 이론적 공백을 확인했다. 현존하는 모델들은 주로 예술적 공동 창작이나 특정 과업 수행에 국한되어 있어(범용적 적용의 한계), 장기간에 걸쳐 신뢰를 쌓고 관계를 진화시켜 나가는 연속적이고 순환적인 과정(process)을 설명하기에는 부족하다. 구조(HACO, COFI), 소통(FAICO, UCCC), 행위자성(CCDF)을 개별적으로 다루는 틀은 존재하지만, 이들을 아울러 인간과 AI가 서로를 어떻게 변화시키고 성장시키는지 설명하는 거시적 모델은 부재하다. 따라서 차세대 커뮤니케이션 연구는 개별 상호작용의 최적화를 넘어, 인간과 AI가 의도 정렬(alignment)과 상호 증강을 통해 관계를 발전시켜 나가는 메커니즘을 규명해야 한다. AI를 사회적 행위자로 규정하는 구조적 모델이나 역동적 상호작용 프로세스를 설명하는 순환 모델과 같은 새로운 모델을 정립하는 것이야말로 에이전틱 AI 시대가 우리에게 던지는 가장 시급하고도 본질적인 학술적 과제라 할 것이다.

마지막으로 에이전틱 AI의 확산은 국가와 조직 차원의 정보화 정책에 중대한 전환을 요구한다. 무엇보다 에이전틱 AI가 자율적으로 계획을 수립하고 외부 도구 및 API를 호출하며 과업을 수행하는 환경에서는, 기존처럼 AI를 단순 도구로 간주하여 오류 책임을 사용자 또는 제조사에 일방적으로 귀속시키는 것이 불가능하다. 따라서, 인간 사용자와 에이전틱 AI 간의 권한 위임 범위(자율성 정도, 중간 점검, 승인 등), 사고 발생 시 책임과 권한 분할 원칙(역할 기반 책임 배분), 그리고 결과 오류와 피해 발생 시 해결 및 재발 방지 절차를 윤리적이고 책임 있는 가이드라인으로 명확히 규정할 필요가 있다. 또한, 자동화의 역설에서 보듯 시스템 복잡성이 증가할수록 평가의 간극이 커지므로, 공공 및 기업 영역에 도입되는 에이전틱 AI는 자신이 어떤 추론과 성찰 루프를 거쳐 특정 행동을 선택했는지, 어떤 데이터와 도구를 어떤 권한으로 호출했는지에 대한 의사결정 로그와 행동 기록을 표준화하고, 이를 인간 감독자가 검토하고 감사할 수 있도록 투명성 기준을 제도적으로 확립해야 한다.

설명 가능성의 확보를 에이전틱 AI 거버넌스의 핵심 원칙으로 명시해야 하며, 또한 도구 연동을 통한 데이터 흐름이 길어짐에 따라 개인정보와 보안 거버넌스도 결과물의 문제를 넘어 행동 주체로서의 처리 기록, 권한, 데이터 최소화, 외부 전송 통제, 위협 모델 기반 점검으로 확장되어야 한다. 끝으로, 정보화 교육은 코딩이나 프롬프트 작성 중심의 명령 입력 능력에서 벗어나, AI 산출물의 맥락적 평가, 적정 신뢰 유지, 로그와 근거 검증, 자율성 조절 등 에이전틱 AI 문해력 교육으로 재편되어야 하며, 이러한 역량 체계의 재정립이 공공부문과 업 해결 과정과 인재 양성, 조직 운영 등의 기준에 연동되어 정책적으로 뒷받침될 필요가 있다.

결론적으로, 에이전틱 AI는 인간의 인지적, 감각적 능력을 확장하고 새로운 차원의 협업을 가능케 하는 도구적 파트너이자 사회적 동반자로 존재한다. 에이전틱 AI 시대의 커뮤니케이션은 인간이 기계를 통제하는 기술적 문제를 넘어, 인간과 비인간 행위자가 어떻게 협력하여 새로운 가치를 창출할 것인가에 대한 사회적, 철학적 질문으로 확장되고 있다. 따라서 향후 연구에서는 이러한 파트너십이 갖는 설명 가능성, 통제권의 공유, 책임 구조의 명료화와 재분배, 행위자성, 공동 행위의 측정과 책임 분할 등 이제까지 존재하지 않았던 새로운 연구 문제를 해결해야 한다. 이 연구는 이러한 거대한 패러다임 전환을 이해하고, 바람직한 인간-AI 공존 사회를 설계하는 데 필요한 이론적 디딤돌이 된다는 점에서 의미를 갖는다.

■ 참고문헌

- 정동훈 (2025). 생성형 AI 파워드 메타버스: 사용자 중심 경험과 사회·기술적 상호작용의 진화와 혁신. <정보화정책>. 32권 1호, 3-26. <https://doi.org/10.22693/NIAIP.2025.32.1.003>
- Acharya, D. B., Kuppan, K. & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey. *IEEE Access*, 13, 18912-18936. <https://doi.org/10.1109/ACCESS.2025.3532853>
- Allen, R. B. & Scerbo, M. W. (1983). Details of command-language keystrokes. *ACM Transactions on Information Systems*, 1(2), 159-178.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775-779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
- Baird, A. & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, 45(1), 315-341. <https://doi.org/10.25300/MISQ/2021/15882>
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1-13.
- Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *Sociological Review, Supplement*, 32(1), 196-233. <https://doi.org/10.1111/j.1467-954X.1984.tb00113.x>
- Card, S. K., Moran, T. P. & Newell, A. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum Associates.
- Clancey, W. J. (1993). Situated action: A neuropsychological interpretation response to Vera and Simon. *Cognitive Science*, 17(1), 87-116.
- Davis, N., Sherson, J. & Rafner, J. (2025, June). *The co-creative design framework for hybrid intelligence*. In Proceedings of the 2025 Conference on Creativity and Cognition (pp. 560-572). <https://doi.org/10.1145/3698061.3726934>
- De Graaf, M. M., Allouch, S. B. & van Dijk, J. (2016, March). Long-Term Acceptance of Social Robots in Domestic Environments: Insights from a User's Perspective. In *AAAI spring symposia*.
- Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraye, L., Candelon, F. & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality.

- Harvard business school technology & operations mgt. Unit working paper*, (24-013). <https://doi.org/10.2139/ssrn.4573321>
- De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F. & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331.
- Dubey, A., Abhinav, K., Jain, S., Arora, V., & Puttaveerana, A. (2020, February). HACO: a framework for developing human-AI teaming. In *Proceedings of the 13th innovations in software engineering conference* (pp. 1-9).
- Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors*, 59(1), 5-27. <https://doi.org/10.1177/0018720816681350>
- Endsley, M. R. (2023). Ironies of artificial intelligence. *Ergonomics*, 66(11), 1656-1668. <https://doi.org/10.1080/00140139.2023.2243404>
- Engen, V., Pickering, J. B. & Walland, P. (2016). Machine agency in human-machine networks: Impacts and trust implications. In M. Kurosu (Ed.), *Human-computer interaction: Novel user experiences* (Lecture Notes in Computer Science, Vol. 9733, pp. 96-106). Springer. https://doi.org/10.1007/978-3-319-39513-5_9
- Faraj, S., Pachidi, S. & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62-70. <https://doi.org/10.1016/j.infoandorg.2018.02.005>
- Feng, K. J., McDonald, D. W. & Zhang, A. X. (2025). Levels of autonomy for ai agents. *arXiv preprint arXiv:2506.12469*.
- Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J. & Zitnik, M. (2024). Empowering biomedical discovery with AI agents. *Cell*, 187(22), 6125-6151.
- Gong, L. & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human communication research*, 33(2), 163-193. <https://doi.org/10.1111/j.1468-2958.2007.00295.x>
- Hoff, K. A. & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407-434. <https://doi.org/10.1177/0018720814558417>
- Hollan, J., Hutchins, E. & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2), 174-196.
- Horvitz, E. (1999). *Principles of mixed-initiative user interfaces*. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 159-166). Association for Computing Machinery. <https://doi.org/10.1145/302979.303029>
- Hosseini, S. & Seilani, H. (2025). The role of agentic AI in shaping a smart future: A systematic review. *Array*, 26, 100399. <https://doi.org/10.1016/j.array.2025.100399>
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Hutchins, E. L., Hollan, J. D. & Norman, D. A. (1985). Direct manipulation interfaces. *Human-Computer Interaction*, 1(4), 311-338. https://doi.org/10.1207/s15327051hci0104_2
- Jiang, F., Pan, C., Dong, L., Wang, K., Dobre, O. A. & Debbah, M. (2026). From large AI models to agentic AI: A tutorial on future intelligent communications. *IEEE Communications Surveys & Tutorials*(Early Access). <https://doi.org/10.1109/JSAC.2026.3660010>
- John, B. E. & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, 3(4), 320-351.
- Krämer, N. C., Lucas, G., Schmitt, L. & Gratch, J. (2018). Social snacking with a virtual agent-On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human-Computer Studies*, 109, 112-121.

- Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A. & Tan, C. (2021). Towards a science of human-ai decision making: a survey of empirical studies. *arXiv preprint arXiv:2112.11471*.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice*, 5(4), 379-393. <https://doi.org/10.1007/BF01059830>
- Lee, K. M. (2004). Presence, explicated. *Communication Theory*, 14(1), 27-50. <https://doi.org/10.1111/j.1468-2885.2004.tb00302.x>
- Lee, S., Hwang, S. & Lee, K. (2025, July). *Beyond Individual UX: Defining Group Experience (GX) as a New Paradigm for Group-centered AI*. In Companion Publication of the 2025 ACM Designing Interactive Systems Conference(pp. 357-362). <https://doi.org/10.1145/3715668.3736348>
- Littlejohn, S. W., Foss, K. A. & Oetzel, J. G. (2017). *Theories of human communication* (11th ed.). Waveland Press.
- Lubart, T. I. (2005). How can computers be partners in the creative process: Classification and commentary on the Special Issue. *International Journal of Human-Computer Studies*, 63(4-5), 365-369. <https://doi.org/10.1016/j.ijhcs.2005.04.002>
- McDermott, P., Dominguez, C., Kasdaglis, N., Ryan, M., Trhan, I. & Nelson, A. (2018). *Human-machine teaming systems engineering guide* (No. MP180941).
- Michael, J. (2011). Shared emotions and joint action. *Review of Philosophy and Psychology*, 2(2), 355-373. <https://doi.org/10.1007/s13164-011-0055-2>
- Moruzzi, C. & Margarido, S. (2024, May). A user-centered framework for human-ai co-creativity. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-9). <https://doi.org/10.1145/3613905.3650929>
- Nass, C., Steuer, J. & Tauber, E. (1994). *Computers are social actors*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 72-78). Association for Computing Machinery. <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/10.1111/0022-4537.00153>
- Norman, D. A. (2013). *The design of everyday things* (Rev. and expanded ed.). Basic Books.
- Nowak, K. L. & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5), 481-494. <https://doi.org/10.1162/105474603322761289>
- Patel, J. (2025). Exploring the role of agentic AI in enhancing Human-AI collaboration. *International Journal of Engineering Technology Research & Management.*, 9(8). <https://ijetrm.com/issues/files/Aug-2025-04-1754328421-AUG08.pdf>
- Rasmussen, J. (1986). Information processing and human-machine interaction: An approach to cognitive engineering. North-Holland.
- Reeves, B. & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.
- Rezwana, J. & Ford, C. (2025, June). Human-centered AI communication in co-creativity: An initial framework and insights. In *Proceedings of the 2025 Conference on Creativity and Cognition* (pp. 651-665). <https://doi.org/10.1145/3698061.3726932>
- Rezwana, J. & Maher, M. L. (2023). Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. *ACM Transactions on Computer-Human Interaction*, 30(5), Article 67. <https://doi.org/10.1145/3613905.3650929>

- doi.org/10.1145/3519026
- Sapkota, R., Roumeliotis, K. I. & Karkee, M. (2026). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. *Information Fusion*, 126, 103599. <https://doi.org/10.1016/j.inffus.2025.103599>
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G. & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174. <https://doi.org/10.1016/j.im.2019.103174>
- Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.
- Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *Computer*, 16(8), 57-69. <https://doi.org/10.1109/MC.1983.1654471>
- Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K. & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36, 8634-8652.
- Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A. E., Sellen, A. & Rintel, S. (2025). Ironies of generative AI: understanding and mitigating productivity loss in Human-AI interaction. *International Journal of Human-Computer Interaction*, 41(5), 2898-2919. <https://doi.org/10.1080/10447318.2024.2405782>
- Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility*(pp. 73-100). The MIT Press.
- Sundar, S. S., Jia, H., Waddell, T. F. & Huang, Y. (2015). Toward a theory of interactive media effects (TIME): Four models for explaining how interface features affect user psychology. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 47-86). Wiley-Blackwell. <https://doi.org/10.1002/9781118426456.ch3>
- Varela, F. J., Thompson, E. & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
- Vera, A. H. & Simon, H. A. (1993). Situated action: A symbolic interpretation. *Cognitive Science*, 17(1), 7-48.
- Wiener, N. (2019). *Cybernetics or control and communication in the animal and the machine*. The MIT Press. <https://doi.org/10.7551/mitpress/11810.001.0001> (Original work published 1948)
- Yao, S., Chen, H., Yang, J. & Narasimhan, K. (2022). Webshop: Towards scalable real-world web interaction with grounded language agents. *Advances in Neural Information Processing Systems*, 35, 20744-20757. <https://dl.acm.org/doi/10.5555/3600270.3601778>
- Zhao, C. & Xu, W. (2025). Human-AI Interaction Design Standards. *arXiv preprint arXiv:2503.16472*.
- Zheng, Q., Tang, Y., Liu, Y. & Liu, W.(2022). *UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library*. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). ACM. <https://doi.org/10.1145/3491102.3501855>
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89. <https://doi.org/10.1057/jit.2015.5>