

디바이스 중심 HCI에서 AI 에이전트 중심 HCI로: 조작에서 협력으로의 패러다임 전환

From Device-Centered HCI to AI Agent-Centered HCI

신희숙 (H.S. Shin, hsshin8@etri.re.kr)
이용호 (Y.H. Lee, jason0720@etri.re.kr)
길연희 (Y.H. Gil, yhgil@etri.re.kr)

실감상호작용연구실 책임연구원
실감상호작용연구실 선임연구원
실감상호작용연구실 책임연구원/실장

ABSTRACT

This article analyzes the paradigm shift in human-computer interaction (HCI) from device-centered manipulation to collaborative interaction with AI agents. Early HCI focused on physical control and efficiency through keyboards, mice, and graphical user interfaces, whereas contemporary systems increasingly interpret user intent and context to provide proactive, agent-centered interaction accelerated by the convergence of ubiquitous computing, mixed-initiative interaction, context-aware technologies, and modern AI. As AI becomes embedded within interfaces, interaction patterns have been extended from tool-centered human-AI integration, in which AI proposes results and users make selections, to collaborative human-AI agent interaction (HAI), in which agents share planning, tool use, and action execution with humans along a continuous trajectory. In spatially embodied interaction environments, such as augmented reality, virtual reality, and robotics, the role of AI agents is rapidly shifting from tools to teammates, and the design focus is moving from “How well does the system perform the task?” to “How well does the agent understand and work with the user?” in empathy-driven interactions. In this context, future HAI research should be grounded in a deep understanding of human-human interaction and extend toward interaction designs and technological developments that realize coworker AI agents capable of mutual trust and collaboration.

KEYWORDS HCI, AI 에이전트, 상호작용

I. 서론

데스크탑에서 모바일, 웨어러블, 공간 컴퓨팅에 이르기까지 인간-컴퓨터 상호작용(HCI: Hu-

man-Computer Interaction) 기술은 하드웨어 폼팩터의 변화에 맞춰 주요 입출력 수단을 확장하고 고도화하는 흐름으로 발전해 왔다. 최근에는 폼팩터에 의존해 사용자가 장치의 UI를 학습하기보다는 시

* DOI: <https://doi.org/10.22648/ETRI.2026.J.410204>

* This work was supported by Electronics and Telecommunications Research Institute(ETRI) grant funded by the Korean government [26ZC1100, Development of Spatial Media Technology and Interaction Technology for Convergence of the Real and Virtual World]



본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

시스템이 사용자의 의도와 맥락을 모델링하여 능동적으로 도와주는 에이전트 중심의 상호작용이 빠르게 확산되고 있다. 이러한 흐름은 Weiser의 유비쿼터스 컴퓨팅(환경에 스며드는 컴퓨팅), Horvitz의 혼합주도(Mixed-Initiative) 원칙(자동화와 직접 조작의 우아한 결합), Dey의 상황인지(Context-aware) 연구 등 기존 HCI 측면에서 제시된 방향성과 최신 AI 기술이 상호보완적으로 만나 강화된 결과로 해석할 수 있다 [1-3].

사용자 경험(User eXperience)은 폼팩터와 센싱의 품질과 연동되고 과업 시나리오의 적합성과 데이터 및 모델의 완성도에 의해서도 영향받는다. 이에 더하여 최근에는 장치와의 직접 조작에서 AI 에이전트와의 협력 상호작용이라는 새로운 UX 창출이 가시화되고 있다. 이러한 HCI 기술의 변화, 즉 HAI(Human-AI Agent Interaction)로의 확장 과정을 구체적으로 살펴보고, 앞으로 HAI 기술의 핵심이 무엇이 될지 전망해 본다.

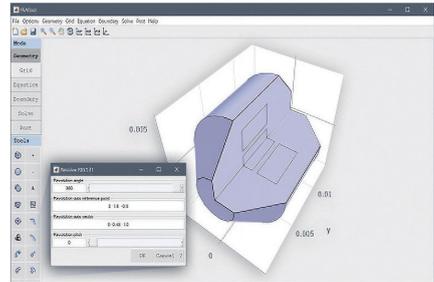
II. HCI 기술의 변화

1. 디바이스 중심의 HCI

초기의 HCI는 컴퓨팅 장치의 물리적 입출력 성능을 극대화하여 사용자가 도구를 효율적으로 다루게 하는 데 목적을 두었다.

Xerox PARC의 Alto(1973)는 비트맵 디스플레이와 마우스를 도입해 GUI(Graphical User Interface)의 토대를 마련하였고, Xerox Star 8010(1981)이 키보드·마우스·포인터·윈도·아이콘으로 구성된 WIMP(Windows, Icons, Menus, Pointer)를 상업 제품 수준으로 정립하였다. 이후 Apple Lisa(1983)와 Macintosh(1984)가 이를 대중화하며, GUI는 인간-기계 상호작용의 표준이 되었다[4].

이 시기의 HCI는 데스크탑 장치를 정밀하고 빠



출처 Reprinted from Carlo Nardone, Xerox Alto I (1973) workstation console, Wikimedia Commons, 2007.
Reprinted from Steve Garfield, Macintosh 1984, Flickr, 2011.
Reprinted from Femfons, CAD(FEATool Multiphysics MATLAB GUI), Wikimedia Commons, 2020.

그림 1 장치 중심 HCI 기술의 예시

르게 조작하는 능력, 즉 도구 사용 속련도의 향상에 초점을 맞추었다. 입력은 키보드나 마우스, 조이스틱 등 명시적 장치에 의존하였고, 사용자는 클릭, 드래그, 키 입력 등과 같은 저수준 명령을 직접 수행하였다. 시스템은 커서 변화, 하이라이트, 애니메이션 등 즉각적 피드백으로 상태 가시성을 보장하였다. 주로 손과 포인터 중심의 단일 모달리티로 설계되었고, 성능 평가는 WPM(타이핑 속도), Fitts's Law(포인팅 효율) 등 물리 정량 지표에 기반하였으며[5], UX 향상은 조작 정확도와 속도의 개선으로 주로 정의되었다.

대표 사례로는 Windows/Mac GUI, CAD 툴, 게임패드 인터랙션 등이 있다. 그림 1은 디바이스 중심의 HCI 기술단계의 대표적인 예시를 나타낸다.

2. Invisible UI와 Natural UI

모바일 및 웨어러블 기술의 확산으로 입력 장치는 소형화되고 소지 및 착용 가능한 폼팩터로 발전하였다. IMU(Inertial Measurement Unit), 마이크, PPG(Photoplethysmography) 등 온디바이스 센서 융합이 일반화되었다. 그 결과 HCI의 관심은 사용자의 일상 속으로 스며드는 상호작용(Invisible User Interface)과 자연스러운 인간 신호를 입출력으로 삼는 NUI(Natural User Interface)로 확장되었다.

I/O로서 멀티터치 제스처는 직접 조작의 주류로 정착했고, 이와 별개로 버튼을 누르지 않아도 음성이나 손짓, 시선 등으로 명령이 가능한 입출력을 구현했다. 정보 모델은 단순한 시스템 상태를 넘어 시간이나 장소 등의 상황적 정보, 즉 컨텍스트가 포함되었다. 음성, 제스처, 시선 등의 다중 신호를 확률적으로 융합한 멀티 모달리티 사용으로 인식 오류를 줄이고, 학습 부담이 낮은 사용자 경험을 지향하였다.

대표 사례로는 스와이프, 핀치 등과 같은 멀티터치 제스처 기반 직접 조작 UI, 착용형 센서기반 생체 신호를 이용한 입출력 기술인 Skinput(CHI, 2010)[6], 깊이 센싱을 이용한 프로젝션 UI Omni-Touch(UIST, 2011)[7], 초기 음성 인터페이스의 예가 되는 Siri(Apple, 2011)와 Alexa(Amazon, 2014), 전신 제스처 인터페이스로 사용되는 Kinect(Microsoft, 2010)[8] 등이 있다[6-8].

TUI(Tangible User Interface)는 현실 기반 상호작용의 대표적 사례이다. 물리 객체를 잡거나 놓기, 돌리기 등과 같은 현실 기반의 조작체가 입력 정보로 사용되었고, 객체의 물리적 형태나 무게감 등이 사용자의 행동을 유도(Affordance)하게 함으로써 학습 부담을 낮추고 조작 정확도를 높였다. 프로젝션 UI, TUI 계열은 연구 중심 개발로 일부 상용화에는 제



출처 Reprinted from Microsoft Research, Skinput, Wikimedia Commons, 2010.

Reprinted from D. Williams, Tangible user interface, Wikimedia Commons, 2007.

Reprinted from S. 1938, Kinect Skeleton View, Tangible user interface, Wikimedia Commons, 2012.

그림 2 Invisible, Natural and Tangible UI 기술의 예시

한적이었지만, ‘보이지 않는 인터페이스’ 비전을 구체화한 연구 결과였다. 그림 2는 이 단계를 대표하는 기술의 예시를 보인다. 이 시기 후반으로 갈수록 기계가 사용자의 의도나 상황 맥락을 예측하고 자동화하여 수행하고자 하는 에이전트적 행위가 강조되었고, 이는 다음 단계인 HAI로의 확장을 준비하는 기반이 되었다.

3. AI 에이전트 중심의 HCI

3.1 H-AI: 도구형 AI 상호작용

1990년대 후반부터 컴퓨터를 단순한 조작 대상이 아닌, 보다 자연스럽게 능률적으로 상호작용을 할 수 있는 주체로 발전시키려는 시도는 이미 시작되었으나 기술적 한계로 인해 대부분 시연 수준의 연구에 머물렀다. 이후 통계적 학습과 머신러닝, 특히 2010년대 딥러닝의 발전으로 음성 인식, 자연어 처리, 컴퓨터 비전 성능이 상용 수준에 도달하면서, AI를 상호작용 주체로 설계하려는 시도가 다시 본

격화되었다. 이후 2010년대 후반, Microsoft가 제안한 H-AI(Human-AI Interaction) 가이드라인을 계기로, 인간 중심의 AI 설계 원칙이 정립되기 시작하였다 [9]. 이 지침은 이후 AI 시스템 설계에 있어 신뢰, 해석 가능성, 조절 가능성 등의 기준을 제시하는 핵심 설계 지침으로 자리 잡았다.

도구형 AI 상호작용으로 설명할 수 있는 H-AI는 주로 사용자가 명령을 내리고, AI는 이를 처리하여 결과를 제시하거나 예측하는 구조를 가진다. 초기에는 명시적 규칙 기반 추천 시스템 수준이었지만, 이후 딥러닝 기반 확률 추론·생성 모델의 등장으로 요약, 자동완성 등 개방형 출력을 포함하는 상호작용으로 발전하였다. 이때의 상호작용 설계 핵심은 결과 해석 가능성 제공, 불확실성 시각화 및 사용자 신뢰 확보 그리고 실행 전후의 사용자 제어권 보장에 있다. 대표적인 도구형 AI 상호작용 사례로는 Item-to-Item Collaborative Filtering 기반 추천(Amazon, 2003)과 Gmail Smart Compose의 문장 자동완성 기능(Chen et al., 2019)이 있다[10,11]. 최신 시스템은 점차 사용자 맥락 이해 및 협업적 흐름을 내포하기 시작하면서 H-AI와 HAI가 혼합되는 과도기적 특성을 띠기도 한다.

3.2 HAI: 협업형 AI 상호작용

본 보고서에서 HAI(Human-AI Agent Interaction)는 사용자가 AI 에이전트와 대화, 도구 사용, 계획 수립 등을 통해 공동으로 과업을 수행하는 상호작용 구조를 의미한다. 여기서 HAI라는 약어는 기존 문헌에서 사용되는 Human-AI Interaction 또는 Human-Agent Interaction과의 혼동을 피하기 위해 AI 에이전트를 명시적으로 포함하는 용어로 한정한다. AI 에이전트는 환경·사용자·문맥 정보를 해석하여 과업 목표를 파악하고, 계획 수립·도구 호출·행동 실행·결과 모니터링 등의 적절한 행동을

제안하거나 직접 실행하며, 사용자는 이에 대해 승인·수정·재요청 등 조정 행위를 수행한다. 즉, 단순한 정보 제공이나 결과 예측을 넘어서 혼합주도(Mixed-Initiative) 구조 속에서 인간과 AI가 역할을 분담하고 상호 조율하는 협력적 상호작용이 핵심이다.

Siri(Apple, 2011)와 Alexa(Amazon, 2014)는 음성 명령 기반 Invisible UI에서 출발하여, 초기형 HAI의 형태를 보여준 사례로 볼 수 있다. 이후 Wysa(Touchkin eServices, 2016), Woebot(Woebot Health, 2017), Pi(Inflection AI, 2023)와 같이 감성 대화를 중심으로 사람의 정서에 반응하는 시스템이 등장하였다. OpenAI Assistants API 프레임워크(OpenAI, 2023), LangChain Agents(LangChain, 2023), Notion AI(Notion Labs, 2023)는 계획과 도구 사용을 결합한 협업형 AI 상호작용의 전형적인 구현 예시이다.

가장 최근에는 Realtime API(OpenAI, 2024)와 GPT-5.2(OpenAI, 2025), Gemini Live API(Google, 2024), Claude Sonnet 4.6(Anthropic, 2026) 등과 같이, 저지연 멀티모달 입력을 바탕으로 사용자의 의도와 상황을 인식하고 도구 호출로 행동을 제안·수행하는 상황인지형 협업 시스템이 제품·플랫폼 수준으로 확산되고 있다.

이러한 변화는 단순한 기능 향상을 넘어, HCI 패러다임의 중심이 ‘조작’에서 ‘협업’으로 확장되고 있음을 보여준다. 이 흐름은 다음의 네 가지 축을 중심으로 일차적인 기술 발전 양상을 정리할 수 있다.

첫째, 입출력(I/O)의 변화로써 명시적 입력(키보드·터치)에 더해 시선, 표정, 끄덕임, 침묵 등 비언어적 신호가 명령으로 인식되고, AI 에이전트의 반응 역시 음성, 시선, 제스처 등 사람다운 방식으로 표현된다.

둘째, 정보 처리 방식(Info)의 변화로써 AI는 사용자별 데이터를 바탕으로 개인화된 지식 공간을 구축하고, 결과를 제시할 때 왜 이런 판단을 했는지 설

명하는 기능(추론 가시성)을 제공한다.

셋째, 모달리티(Modality)의 확장으로써 과거엔 단일 감각(시각·청각·촉각)에 의존했다면, 이제는 상황(Context)과 환경, 그리고 여러 사용자의 상호작용까지 아우르는 다자간 멀티모달 융합이 중심이 된다.

넷째, 목표(Goal)의 변화로써 단순히 속도나 효율을 높이는 대신, AI를 얼마나 신뢰하고 안전하게 쓸 수 있는지를 중시한다. 즉, 복구 가능성(되돌리기), 설명 가능성(왜 그런 결과인지 이해), 적절한 의존과 신뢰가 설계의 핵심이 된다.

3.3 AI 에이전트 중심 HCI 기술의 확장

정리하자면, 표 1은 인간-컴퓨터 상호작용의 설계 관점이 어떻게 변화해 왔는지를 입력방식, 정보

표 1 HCI의 변화

| | Device 중심 HCI | Invisible/Natural UI | AI 에이전트 중심 HCI |
|----------|--|-------------------------|---------------------------------------|
| I/O | 버튼·포인터·터치 등 명시적 조작 입력 | 음성·제스처·시선 등 자연 신호 중심 입력 | 비언어 신호와 맥락에 기반한 고수준 의도 입력 |
| Info | 시스템 상태 표시, 진행 상황 표시 (대규모 정보 또는 작업 공간 요구) | 상황 (컨텍스트) 포함 | 판단 근거와 이유를 설명하는 추론 가시성 (개인화 지식 공간 요구) |
| Modality | 단일 감각 중심(손, 시각) | 음성·시각 등 다중 감각 | 다중·다자 감각 융합(공간·사회·문화적 신호) |
| Goal | 조작 효율, 성능 향상 (도구 사용 능력 평가) | 학습 부담 감소와 자연스러운 사용성 | 신뢰·공감 기반 협업 (팀워크 최적화) |
| 주체 | Device (장치 중심) | Me (개인 1인 사용자 중심) | We (다수 사용자·AI 에이전트 팀) |
| 예시 | GUI, 멀티터치, CAD, 게임패드 | 음성 비서, 제스처 인식, PUI, TUI | 상담 AI, 헬스 코치 로봇 |

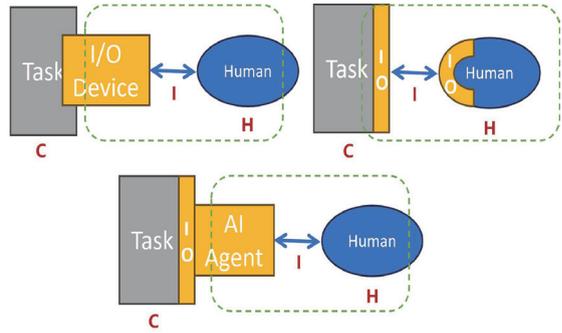


그림 3 HCI를 바라보는 관점의 변화 도식화

구조, 모달리티 범위, 상호작용 목표를 중심으로 요약한다.

디바이스 중심 HCI는 인간(H)이 과업(Task)을 수행하기 위해 I/O 장치를 매개로 직접 조작하는 구조를 가진다. Invisible UI와 NUI 단계에서는 입력 장치가 환경과 신체로 스며들며 인터페이스의 물리적 존재감은 줄어들고 맥락 기반의 반응성이 강조되었다. AI 에이전트 중심 HCI에서는 인간과 과업 사이에 AI 에이전트가 개입하여 사용자의 의도, 감정, 상황 정보를 해석하고 과업을 대리 또는 협력 수행하는 구조로 전환된다. 이 단계에서는 인간과 AI 에이전트가 지시-수행의 관계를 넘어서, 상호 신뢰와 공감에 기반한 협업 관계를 지향하게 된다. 따라서 앞으로의 인터랙션 설계는 사용자의 손끝 조작의 정확성과 효율뿐만 아니라 사용자의 의도와 한계를 이해하고, 공감적 반응을 동반하는 협업 수행 절차에도 초점을 맞출 필요가 있다.

그림 3은 이러한 변화가 인터페이스의 구조와 주체 관계에 어떤 영향을 주었는지를 시각적으로 보여준다.

III. 공간상호작용에서의 HAI

앞장에서는 HAI를 주로 기능적 관점에서 살펴보

았다. 하지만 실제 사용자 경험은 기능만으로 설명되기 어렵고 AI 에이전트가 어떤 형상과 공간적 존재 방식으로 나타나는지에 따라 상호작용의 성격과 설계 목표가 크게 달라진다. 이에 본 장에서는 상호작용의 형상성과 공간적 구현 양식을 중심으로 HAI의 기술적 특징과 설계 목표를 살펴보고자 한다. 이 관점에서 에이전트는 크게 형상이 없는 에이전트와 형상을 가지는 에이전트의 두 가지 유형으로 구분할 수 있다.

첫째, 형상이 없는 에이전트는 텍스트 기반 챗봇, 스마트 스피커의 음성 비서, 모바일·데스크탑 환경에서 동작하는 코파일럿형 등에서 흔히 나타나는 형태이다. 이들은 사용자의 눈앞에 특정한 몸이나 위치, 거리 정보를 갖지 않으며, 주로 언어·아이콘·시각적 컴포넌트를 매개로 상호작용을 수행한다.

이러한 유형에서 중요한 설계 요소는 대화의 흐름과 역할 분배이다. 사용자의 발화를 언제 끊고 응답을 시작할 것인지(Turn-taking, 턴테이킹), 어떤 타이밍에 제안하거나 개입할지(혼합주도성), 결과를 어떻게 설명하고, 실패·오류 시 어떻게 사과·복구할 것인지(추론 가시성·오류 복구성), 어떤 말투·어조·개인화 전략을 통해 신뢰감을 줄 것인지(공감 표현과 일관된 페르소나의 유지)와 같은 요소들이 상호작용의 품질을 좌우한다.

예를 들어, Microsoft 365 Copilot, Gemini in Google Workspace, Notion AI, Slack AI 등은 문서 작성, 회의 기록, 아이디어 정리, 이메일 응답, 팀 협업 지원 등에서 사용자 작업 흐름을 이해하고 맥락에 맞는 제안과 공동 작업 흐름의 일부로 작동하는 대표적인 코워킹형 비형상 에이전트로 볼 수 있다.

둘째, 형상을 가진 에이전트(Embodied Agent)는, 가상공간에서는 아바타, 가상 휴먼, 실시간 렌더링 페르소나 등의 형태로, 물리공간에서는 휴머노이드, 서비스 로봇 등의 형태로 구현된다. 이들은 단순한

언어 출력을 넘어서 공간상 특정 위치에 존재하며 시각적·물리적 행동을 통해 상호작용에 참여한다. 이러한 에이전트는 앞서의 대화 흐름과 더불어, 어디를 바라보는지(시선), 어떤 표정을 짓는지(표정·얼굴 애니메이션), 어떻게 손과 팔을 움직이는지(제스처·자세), 사람과 어느 정도 거리를 두고 어떤 속도로 다가가는지(거리와 접근 속도)가 모두 상호작용 설계의 핵심 요소가 된다.

가상공간의 에이전트는 이러한 표현들을 음성·시선·제스처 간의 타이밍 동기화를 통해 사회적 존재감(Social Presence)과 상호 공감을 형성하며, 물리공간에서는 여기에 속도, 힘, 충돌 회피 같은 물리적 안전성 요인이 추가적으로 고려된다.

1. 가상공간에서의 HAI 전망

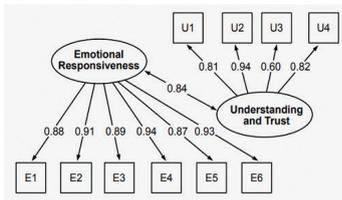
가상공간에서 HAI의 대표적인 형태는 체화형 대화 에이전트(ECA: Embodied Conversational Agent)로 볼 수 있다. ECA는 화면 속에 몸을 가진 아바타나 가상 인간 형태로 구현되며, 단순히 텍스트를 주고받는 수준을 넘어 언어, 시선, 표정, 제스처를 동시에 다루는 상호작용 주체로 설계된다. 이러한 에이전트는 사용자의 말과 표정, 몸짓에 맞춰 자신의 표정·시선·동작을 조정함으로써 사회적 신호(Social Signals)와 정서적 표현(Emotional Expressions)을 통해 사람과의 관계를 단순 정보 교환에서 사회적·정서적 상호작용 수준으로 확장하는 것을 목표로 한다.

초기 ECA 연구(Cassell, 2001)는 음성과 제스처의 동기화 중심으로 언어적·비언어적 행위를 통합적으로 표현하는 모델을 제안하였다[12]. 이러한 통합 모델은 이후 HAI 기술의 기반이 되었다[15]. 최근에는 멀티모달 인식과 생성 기술의 고도화로 사용자의 음성, 표정, 시선, 발화 리듬을 실시간 분석하고 이에 맞는 감정적·사회적 반응을 생성하는 수

준까지 발전하고 있다. 국내의 경우, 한국전자통신연구원(ETRI)에서 실제 인간 간 대화의 비언어적 표현 특성을 분석하여 상호 공감 상태를 객관적으로 정량화하는 연구를 수행하고 있으며[13,14], 이를 기반으로 실시간 언어·비언어적 공감 동기화 표현 기능을 갖춘 가상휴먼 AI 에이전트를 개발하고 있다. 이는 기존의 감성 표현 중심 가상 휴먼을 넘어, 사용자 반응에 따라 인지적 공감을 표현하고 협업의 신뢰를 확보하는 방향으로 체화형 에이전트의 역할을 확장하고자 하는 시도이다.

이러한 기술 발전 방향을 토대로, 가상공간 HAI에서 사용자 경험과 상호작용의 품질을 결정하는 기술 요소는 다음의 세 가지로 정리할 수 있다.

첫째, 멀티모달 합성(Multimodal Generation)이다. 음성, 제스처, 표정의 동기화된 합성(Synchronization)을 200~300ms 이내의 리듬으로 제어하여 자연스러운 상호작용을 보장하는 것이다. GENE Challenge(2023) 연구는 음성의 운율(Prosody)과 제스처 정점을 정밀히 맞추어 사람 수준의 반응 타이밍을



출처 Reprinted from J. Cassell, "Embodied Conversational Agents: Representation and Intelligence in User Interfaces," AI Magazine, vol. 22, no. 4.
Reprinted from M. Schmidmaier et al., "Perceived Empathy of Technology Scale (PETS): Measuring Empathy of Systems Toward the User," in Proc. CHI, (Honolulu, HI, USA), May 2024.

그림 4 가상공간 HAI 기술의 예시

실현하는 모델을 보고하고 있다[15,20].

둘째, 정서 상태 추정과 적응적 대화 전략이다. 단순 감정의 인식과 표현을 넘어서 사용자의 음성 억양·표정·시선 등의 신호를 융합해 정서 상태를 추정하고, 적합한 대화 전략(속도, 억양, 말투, 표정, 제스처 등)을 실시간으로 조정하는 능력이 요구된다. IJHCS(International Journal of Human-Computer Studies), ACM TiiS(ACM Transactions on Interactive Intelligent Systems) 등에서는 사용자의 피로·불안·흥미 수준에 따라 발화 속도와 억양을 조절하는 적응형 대화 모델이 제시되고 있으며, Expressive AI Avatars(Synthesia, 2019~2025), NVIDIA Audio2Face(NVIDIA, 2021~2025)와 같은 실시간 가상휴먼 합성 플랫폼은 음성과 표정, 머리 움직임을 동기화하여 보다 섬세한 감정 표현을 구현하고자 한다.

셋째, 사회적 존재감(Social Presence)을 강화하는 개인화·기억화 기술(Personalized Memory)이다. 대화 이력과 감정 패턴을 기억하고, 그에 기반한 개인화된 반응을 제공함으로써 장기적 신뢰 관계를 형성하는 것이 중요해지고 있다.

향후 연구 과제로는 상호작용 품질의 객관화·정량화, 그리고 사회적·문화적 표현의 다양성 확장 등의 이슈가 있다. CARE(Consultation and Relational Empathy Measure, 2004), GQS(Godspeed Questionnaire Series, 2009), WAI-SR(Working Alliance Inventory-Short Revised, 2006)과 같은 검증된 심리 측정 도구에 더해, 최근 제안된 공감지표 PETS(Perceived Empathy of Technology Scale, CHI 2024)와 정량적 공감 예측 모델[14] 등에 기반한 객관적 평가 기준 마련이 중요해질 것이다[16-19].

아울러 협업·교육·상담 등 복수 사용자가 참여하는 멀티에이전트 환경에서 사회적 규범을 반영하는 상호작용 설계, 그리고 시선, 제스처, 표정의 의미가 문화권마다 다르다는 점을 고려한 문화 정서

모델과 평가 체계에 대한 합의가 향후 표준화 논의의 주제로 부상할 가능성이 크다고 본다.

2. 물리공간에서의 HAI 전망

물리공간에서의 HAI는 앞서 논의한 형상을 가진 에이전트 가운데, 실제 공간을 인간과 함께 사용하는 로봇을 대상으로 하는 Human-Robot Interaction(HRI)의 연장선상에 있다. 오늘날 로봇은 단순한 산업용 자동화 장비를 넘어, 사람과 같은 공간에서 움직이고 대화하며 정서를 주고받는 사회적 행위자로 설계되고 있다. 이때 상호작용의 품질은 작업 정확도뿐만 아니라 로봇이 사용자에게 얼마나 편안한 거리와 속도·방향으로 다가오는지, 얼마나 안전하게 움직이는지, 어떤 표정과 시선·몸짓으로 반응하는지에 의해 크게 좌우된다.

동반자 로봇과 협업 로봇은 이러한 요구를 잘 보여주는 예로, Furhat(Furhat Robotics, 2018)은 얼굴 표정과 시선을 통해 대화를 이끌고, Lovot(GROOVE X, 2019)는 터치와 응시를 통해 정서적 애착과 안정감을 형성하도록 설계되어 있다. 최근 주목받는 가정용 휴머노이드(1X NEO)는 이러한 HRI 특성을 가정과 일상 공간으로 확장하려는 시도로 볼 수 있다.

물리공간에서의 HAI 품질은 카메라·마이크·SLAM(Simultaneous Localization and Mapping) 등으로 사람과 환경을 파악하는 상황 인식, 시선·표정·몸짓·거리와 속도를 아우르는 사회적 행동 합성, 그리고 속도·힘 제한과 충돌 회피를 포함한 물리적 안전 설계가 얼마나 잘 결합되어 있는지에 의해 좌우되며, 최근 연구 개발은 이 세 영역을 통합적으로 향상시키는 방향으로 진행되고 있다.

가상공간과 물리공간의 AI 에이전트 기술은 서로 다른 환경과 제약 조건을 다루지만, 인간의 내적·사회적 신호를 해석하고 상호작용의 신뢰를 구축한

다는 점에서 공통적인 HAI 패러다임 아래에 놓인다. 가상공간은 물리적 충돌이나 센서·모션 제어의 제약이 상대적으로 적기 때문에, 고해상도의 체화 표현과 정교한 제스처·표정 제어에 집중할 수 있다는 장점이 있다. 반면 물리공간에서는 물리적 안전과 로봇 역학이라는 추가 제약이 존재하지만, 실제 공간을 공유하는 동반자이자 협업자로서의 경험을 설계할 수 있다는 점에서 가상공간 연구와 상호보완적 대상이 된다.

IV. 결론

본고에서는 데스크탑 기반 GUI에서 모바일·웨어블·공간 컴퓨팅으로 이어지는 HCI 기술의 변화를 정리하고, 그 연장선에서 AI 에이전트 중심 상호작용인 HAI(Human-AI Agent Interaction)의 방향성을 분석하였다.

최근 AI 기술의 고도화는 사용자가 도구를 ‘조작하는’ 단계에서 AI 에이전트와 ‘협업하는’ 단계로의 패러다임 전환을 견인하였으며, 이에 따라 HCI는 인간과 AI가 어떤 관계를 맺고 어떻게 함께 일할지를 설계하는 분야로 확장되고 있다.

특히 국내외에서는 실제 인간 대화 데이터를 기반으로 공감 상태를 정량화하고, 언어·비언어 신호를 통합하여 단순한 감정 표현형 아바타를 넘어 인지적 공감과 협업 신뢰를 형성하는 공감지능 에이전트를 구현하려는 시도가 ETRI를 중심으로 활발히 이루어지고 있다. 이러한 흐름 속에서 가상공간 HAI는 새로운 상호작용 패러다임을 실험·선도하는 영역으로, 물리공간 HAI는 그 전환의 성과를 일상의 물리 환경으로 확장하는 영역으로 자리 잡아갈 것으로 보인다.

앞으로의 가상·물리공간 HAI 연구는 인간 간 상호작용에서 나타나는 인지적·정서적·행동적 메

커니즘을 체계적으로 관찰·분석하고, 이를 토대로 AI 에이전트의 역할과 상호작용 수준을 시각 인식과 언어 표현의 능력을 넘어, 체화된 사회·문화적 행동과 정서·가치적 판단까지 포괄하는 능력으로 확장할 수 있을 것이다. 이러한 확장은 사용자가 신뢰하고 지속적으로 협업할 수 있는 코워커 AI 에이전트의 구현 가능성을 한층 높이는 기반이 될 것으로 기대된다.

용어해설

HCI(인간-컴퓨터 상호작용) 사람과 컴퓨터 시스템 사이의 상호작용을 연구·설계하는 분야. 입력장치(I/O), UI 패턴, 과업 흐름, 사용성·효율·안전·접근성 등의 품질을 개선하는 방법과 평가 절차를 다루는 연구 분야

HAI(인간-AI 에이전트 상호작용) 사람과 인공지능 에이전트(자율적 목표·추론·행동을 수행하는 소프트웨어나 아바타, 로봇 등) 사이의 상호작용을 연구·설계하는 분야

H-AI(인간-AI 상호작용) 사람이 시가 적용된 앱·서비스(추천, 자동완성, 요약·생성 등)와 어떻게 주고받는지를 연구·설계하는 분야

HRI(인간-로봇 상호작용) 사람과 물리적 로봇 사이의 상호작용을 연구·설계하는 분야. 로봇의 형태·동작·거리를 포함한 공간적 행동과 음성·시각·촉각 인터페이스를 다룸

ECA(체화형 대화 에이전트) 화면 속 아바타나 3D 캐릭터, 디지털 휴먼처럼 몸과 얼굴을 가진 형태로 구현된 대화형 에이전트

Ubiquitous Computing(유비쿼터스 컴퓨팅) 컴퓨팅이 환경에 스며들어 언제 어디서나 작동하는 상태. 소형/착용 폼팩터와 상시 센싱을 전제로 맥락 인지·자동 적응형 서비스의 이론적 기반을 제공

참고문헌

- [1] M. Weiser, "The Computer for the 21st Century," *Sci. Am.*, vol. 265, no. 3, 1991, pp. 94-104.
- [2] E. Horvitz, "Principles of Mixed-Initiative User Interfaces," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, (Pittsburgh, PA, USA), May 1999, pp. 159-166.
- [3] A. K. Dey, "Understanding and Using Context," *Pers. Ubiquitous Comput.*, vol. 5, no. 1, 2001, pp. 4-7.
- [4] J. Johnson et al., "The Xerox Star: A Retrospective," *IEEE Comput.*, vol. 22, no. 9, 1989, pp. 11-26.
- [5] P.M. Fitts, "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement," *J. Exp. Psychol.*, vol. 47, 1954, pp. 381-391.
- [6] C. Harrison et al., "Skinput: Appropriating the Body as an Input Surface," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, (Atlanta, GA, USA), Apr. 2010, pp. 453-462.
- [7] C. Harrison et al., "OmniTouch: Wearable Multitouch Interaction Everywhere," in *Proc. Annu. ACM Symp. User Interface Softw. Technol.*, (Santa Barbara, CA, USA), Oct. 2011, pp. 441-450.
- [8] Microsoft, "Kinect for Xbox 360 Launch," Microsoft News Center, 2010. <https://news.microsoft.com/source/2010/06/14/kinect-for-xbox-360-sets-the-future-in-motion-no-controller-required/>
- [9] S. Amershi et al., "Guidelines for Human-AI Interaction," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, (Glasgow, UK), May 2019, pp. 1-13.
- [10] G. Linden et al., "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Comput.*, vol. 7, no. 1, 2003, pp. 76-80.
- [11] M.X. Chen et al., "Gmail Smart Compose: Real-Time Assisted Writing," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, (Anchorage, AK, USA), Aug. 2019, pp. 2287-2295.
- [12] J. Cassell et al., "Embodied Conversational Agents," *Commun. ACM*, vol. 43, no. 4, 2000, pp. 70-78.
- [13] Y. Lee et al., "Measurement of Empathy in Virtual Reality with Head-Mounted Displays: A Systematic Review," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 5, 2024, pp. 2485-2495.
- [14] Y. Lee et al., "Are You Empathizing with Me? Exploring External Expressions of Empathy in Interpersonal VR Communication," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, (Daejeon, Republic of Korea) Oct. 2025, pp. 1387-1397.
- [15] T. Kucherenko et al., "The GENEA Challenge 2023: A large-scale evaluation of gesture generation models in monadic and dyadic settings," in *Proc. 25th ACM Int. Conf. Multimodal Interact.*, (Paris, France), Oct. 2023, pp. 792-801.
- [16] M. Schmidmaier et al., "Perceived Empathy of Technology Scale (PETS): Measuring Empathy of Systems Toward the User," in *Proc. CHI*, (Honolulu, HI, USA), May 2024, no. 456, 2024, pp. 1-18.
- [17] R.L. Hatcher and J.A. Gillaspay, "Development and Validation of a Revised Short Version of the Working Alliance Inventory," *Psychother. Res.*, vol. 16, no. 1, 2006, pp. 12-25.
- [18] C. Bartneck et al., "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots," *Int. J. Social Robot.*, vol. 1, 2009, pp. 71-81.
- [19] S.W. Mercer et al., "The Consultation and Relational Empathy (CARE) Measure: Development and Preliminary Validation and Reliability of an Empathy-based Consultation Process Measure," *Fam. Pract.*, vol. 21, no. 6, 2004, pp. 699-705.
- [20] M. H. Mughal et al., "ConvoFusion: Multi-Modal Conversational Diffusion for Co-Speech Gesture Synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024.