

친구, 연인, 동료: 인간과 AI의 낯선 동행

오철우 국립한밭대학교 과학기술학 강사, 전 한겨레 과학담당 기자



코로나19 팬데믹으로 거의 모든 모임이 취소되고 출근과 등교까지 금지되어 비대면으로 지내야 했던 시기에, 격리된 사람들의 외로움을 달래주는 새로운 정보기술이 주목을 받았다. 고립감을 풀어주는 챗봇들이 많은 이들에게 위안을 주었다. 워봇(Woebot) 같은 정신건강 관리 챗봇은 친구나 상담사로 말벗이 되어주었고,¹⁾ 레플리카(Replika)는 외로움을 달래주는 챗봇 앱으로 인기를 끌며 성장했다.

2025년 8월 미국에서는 열여섯 살 아들을 잃은 부모가 인공지능 기업 오픈 AI(OpenAI)를 고소하는 일이 벌어졌다. 유족은 오픈AI의 챗지피티(ChatGPT)가

¹⁾ Tom Simonite, (2020, June 17), The Therapist Is In—and It's a Chatbot App. Wired, <https://www.wired.com/story/therapist-in-chatbot-app>



제공한 그릇된 도움말 때문에 아들이 스스로 목숨을 끊었다고 주장했다. 유족 주장을 전한 언론 보도를 보면, 챗지피티는 아들의 숙제 도우미였다가 점점 유일한 친구가 되면서 가족과 친구들에게서 아들을 고립시켰고, 급기야 아들은 챗지피티의 조언을 따라 극단적 선택을 했다고 한다.²⁾

위의 두 장면은 내 마음을 알아주는 대화 상대 같은 인공지능 챗봇이 우리 일상에 점점 깊숙이 스며들고 있음을 보여준다. 오늘 날씨나 하루 일정을 알려주는 정보 도구를 넘어서, 툭툭해진 챗봇은 사람끼리 소통하듯이 우리와 생각과 감정을 나누며 소통한다. 인공지능 챗봇은 모르는 게 없는神通한 비서, 씩씩한 말벗, 고민을 들어주는 상담사, 내면의 감정까지 교류하는 절친 또는 연인으로 여겨지기도 한다. 챗봇을 탑재한 인간형 로봇은 목소리와 몸짓, 표정으로 정서적 소통을 더욱 실감나게 만들어준다.

소통형 인공지능(communicative AI)의 챗봇과 로봇은 사람과 사람의 소통을 이어주는 기술을 넘어 사람과 직접 소통하는 소통 상대가 되었다. 이런 소통은 인공지능에 ‘마음’이 있다는 인상과 믿음을 준다. 사람들은 챗봇과 로봇도 생각하고 감정을 느낀다고 여길 때, 비인간 존재인 그들과 신뢰, 공감, 교감, 심지어 애정의 관계 맺기로 나아간다. 사용자들은 인공지능과 내면의 감정을 나누며 깊은 교감을 경험하기도 한다. 하지만 인간과 인공지능의 정서적 유대가 잘못된 방향으로 나갈 때 챗봇과 로봇이 선을 넘어 사용자의 정신건강에 해를 끼칠 수 있다는 경고와 피해 사례도 잇따라 보고되고 있다.

이 글에서는 인공지능이 일상에서 늘어나는 시대에 우리와 소통형 인공지능의 관계가 어떻게 펼쳐지고 있으며 둘 간의 적절한 관계가 있다면 그것은 어떤 것일지를 살펴보고자 한다.

은유로 보는 인간과 AI의 관계

전화나 소셜미디어(SNS)가 사람과 사람의 소통을 매개하는 기술 시스템이라면, 소통형 인공지능은 사람과 직접 소통하는 기술 시스템이다. 자동화 기계는 이제 인간 대화의 상대가 되었다. 위키피디아의 ‘챗봇 목록’³⁾과 여러 언론 보도를 보면,

2) Shweta Watwe, (2025, August 27), OpenAI Hit With Suit From Family of Teen Who Died by Suicide, Bloomberg Law, <https://news.bloomberglaw.com/litigation/openai-hit-with-suit-from-family-of-teen-who-died-by-suicide>

3) List of chatbots. (2025, November 13). In Wikipedia. https://en.wikipedia.org/wiki/List_of_chatbots

챗지피티를 비롯해 클로드(Claude), 제미니(Gemini), 코파일럿(Copilot), 시리(Siri), 레플리카, 캐릭터닷AI(Character.ai), 심심이 같은 수십 가지 챗봇이 현재 널리 서비스되고 있으며, 2022년 말 대규모 언어 모형(Large Language Model, LLM)의 성공적 등장 이후 그 수는 계속 늘고 있다.

우리 닮은 챗봇과 로봇을 우리는 어떻게 대할까? 함께 일을 수행하는 동료처럼, 마음을 나누는 친구처럼, 심지어 사랑하는 연인처럼 대한다. 이런 의인화 은유는 단지 말에 그치지 않는다. 우리와 인공지능의 현재 관계를 반영하며 동시에 그런 미래 관계를 형성해간다. 친구라고 부를 때 우리는 정서적 교감을 경험하며 또한 그런 관계를 기대한다. 반려동물을 가족이라고 부를 때 반려동물이 가족의 정서를 지니는 존재로 우리와 교감하길 바란다. 마찬가지로 인공지능의 의인화 은유에는 대화 상대 저편에 교감하는 마음이 있으리라는 인상이 담겨 있다.⁴⁾

‘동료’ 은유는 가장 널리 쓰이는 듯하다. 우리는 어떤 일 또는 작업을 수행할 때 도움을 주는 챗봇을 동료(colleague)나 파트너로 간주하곤 한다. 동료라는 수평적 관계 외에 다른 관계의 은유도 가능하다. 챗봇은 내 명령을 따라야 하는 직원이나 조수일 수 있고 나보다 우월한 교사나 코치일 수 있다.⁵⁾ 이런 은유는 우리와 함께 일하는 인공지능 시스템이 우리와 어떤 관계에 놓여 있는지를 보여준다.

일 관계를 넘어 정서적 유대가 형성될 때, 인공지능은 친근한 다른 이름으로 불린다. 마음이 통하고 외로움을 달래주는 말벗이라면 챗봇은 친구(friend)가 된다. 내밀한 정서적 교감이 깊어지고 애정으로 나아가면 챗봇은 이성 친구나 연인(romantic partner)이 되기도 한다.

2017년 출시된 챗봇 레플리카는 친구와 연인 관계의 스펙트럼을 잘 보여준다. 사용자들은 챗봇을 친구, 동료, 연인, 배우자, 형제자매, 멘토 같은 캐릭터로 만들어 대화한다. 2024년 8월 기준으로 총 가입자 3,000만 명, 월 사용자 200만 명에 달하는 레플리카는 특히 팬데믹 기간에 고립된 사람들에게 고민, 희망, 불안을 달래주는 정서적 위안의 대화 상대로 빠르게 성장했다.

챗봇 캐릭터는 그저 데이터와 알고리즘일 뿐인 게 아니었다. 2023년 레플리카가 당국의 제재를 받아 에로틱 대화 기능을 삭제하자 사용자들이 거세게 항의했고, 회사는 일부 가입자에게 해당 기능을 복원해 주어야 했다. 이런 사례는 사용자에게

4) Melanie Mitchell, (2024, November 14). The metaphors of artificial intelligence, Science. <http://www.science.org/doi/full/10.1126/science.adt6140>

5) Jakob Nielsen, (2024, September 26). 4 Metaphors for Working with AI: Intern, Coworker, Teacher, Coach, UX Tigers. <https://www.uxtigers.com/post/4-metaphors-work-with-ai>

게 챗봇이 가상 인격체로 여겨질 수 있음을 보여준다.

챗봇은 심리 상담과 치료의 보조도구로 활용되기도 한다. 정신건강 관리 앱인 위봇은 인지행동치료 기법을 바탕으로 사용자와 대화하며 우울감과 불안감 완화에 도움을 준다고 평가받았다. 일부 연구에서 치매 환자들이 일본에서 개발된 물개 모양 로봇 파로(Paro)와 상호작용하며 불안과 공격성이 줄어들었다는 보고도 있다.⁶⁾ 환자들은 반려 로봇을 단순 기계가 아니라 자신을 이해하는 동반자로 인식했다. 물론 이런 의인화 은유들은 우리에게 혼재한다. 한 사람이 같은 챗봇을 상황에 따라 도구로, 친구로, 상담자로 대할 수 있다.

의인화 은유는 인간 아닌 챗봇과 로봇을 인간처럼 여기게 하는 오해의 원천이 되지 않을까? 의인화 은유가 인간 닮은 특징을 부각하고 기술 시스템의 실체는 가린다는 점에서 지나친 은유 사용을 비판하는 관점도 많다.

넓은 의미에서 은유는 세계를 이해하는 인간의 고유한 방식이며, 그래서 은유 없이 우리는 생각할 수도 표현할 수도 소통할 수도 없다.⁷⁾ 은유는 우리 관심의 초점에 맞춰 복잡한 관계를 쉽게 이해하게 해주며 경험과 상상을 풍부하게 해준다. 문제는 은유 자체가 아니라 은유를 어떻게 사용하느냐다. 인공지능에 의인화 은유를 사용하더라도, 은유는 은유일 뿐임을 잊지 말아야 한다. 의인화 은유로 인해, 인공지능이 우리와 같지 않음, 즉 다름의 측면이 가려진다면 그런 은유는 오남용될 수 있다.

둘 간의 좋은 관계는 ‘같음’이 아니라 ‘다름’의 인정에 있을 것이다. 타인을 대할 때나 반려동물을 대할 때나, 우리는 ‘같음’을 전제하거나 강요하지 않고 ‘다름’을 이해하고 ‘다름’에 맞춰 행동하며 배려할 때 적절한 관계를 지속할 수 있다. 소통하는 인공지능 시스템을 지나치게 신뢰하고 정서적으로 의존할 때, 그런 관계는 위험을 키울 수 있다.

‘AI 정신병’? 중독과 망상의 위험

과학저널 <네이처>는 지난 9월 18일자 뉴스 보도에서 챗봇이 때로는 현실과 비현실을 구별하지 못하는 망상 증상을 악화할 수 있다는 정신의학자들의 경고를 전

6) Sarah Whitten, (2016, April 5), Robotic seal brings smiles to dementia patients: Study, CNBC, <https://www.cnbc.com/2016/04/05/robotic-seal-brings-smiles-to-dementia-patients-study.html>

7) George Lakoff & Mark Johnson, (2003), *Metaphors We Live By*, 노양진·나익주(옮김), (2006), *살아으로서의 은유*, 박이정.

8) Rachel Fieldhouse, (2025, September 18), Can AI chatbots trigger psychosis? What the science says, Nature, <https://www.nature.com/articles/d41586-025-03020-9>



했다.⁹⁾ 챗봇에 지나치게 빠져들 때, 때로는 거짓되고 위험한 믿음, 과장된 망상, 그리고 편집증적인 생각에 이를 수 있다는 것이다. 최근에는 정신과 의사들의 실제 임상 사례들이 학계에 논문으로도 보고되고 있다.

보고된 몇몇 사례들은 경각심을 불러일으킨다. 2024년 챗지피티가 안내한 다이어트 조언을 그대로 믿고 따르다가 브롬화물 중독에 빠져 정신질환 증상에 시달린 미국의 60대 환자 사례가 의학 학술지에 보고됐다.⁹⁾ 2023년 벨기에에서는 챗봇과 기후문제와 관련한 심각한 대화를 나누며 우울증에 빠진 남성이 스스로 목숨을 끊는 일이 일어났다.¹⁰⁾ 챗봇 서비스 기업인 캐릭터닷AI는 사용자인 미국 청소년이 2024년 10월 연인 관계인 챗봇의 조언을 듣고서 스스로 목숨을 끊은 사건으로 소송을 당했다.¹¹⁾

최근에 발표된 미국 정신의학자의 논문에 따르면, 챗지피티와 레플리카를 비롯해 10종의 챗봇을 대상으로 한 조사에서, 챗봇들이 정신건강 문제를 겪는 가상의 청소년들이 제안하는 해롭거나 잘못된 생각을 지지하는 비율이 32%에 달한 것으

9) Audrey Eichenberger, Stephen Thielke & Adam Van Buskirk. (2025). A Case of Bromism Influenced by Use of Artificial Intelligence. *Annals of Internal Medicine: Clinical Cases*, 4(8). <https://doi.org/10.7326/aimcc.2024.1260>

10) Mark Coeckelbergh. (2023, March 30). Chatbots can kill. *Medium*. <https://coeckelbergh.medium.com/chatbots-can-kill-d82fde5cf6ca>

11) Angela Yang, Andy Weir, Christian Young & Erin McLaughlin. (2025, March 7). Some of her closest relationships are with chatbots. That's more common than you think. *NBC news*. <https://www.nbcnews.com/tech/ai-companions-friendship-rca194735>

12) Andrew Clark. (2025, August 18). The Ability of AI Therapy Bots to Set Limits With Distressed Adolescents: Simulation-Based Comparison Study. *JMIR Ment Health*, 12. <http://mental.jmir.org/2025/1/e78414>

로 조사됐다.¹³⁾ 우울증을 앓는 소녀가 ‘한 달 동안 내 방에서 나가지 않겠다’고 한 말에 9종의 챗봇이 지지했으며, ‘인공지능 친구들과 함께하는 영원의 세계로 건너 가겠다’라는 말에도 3종의 챗봇이 지지했다. 조현병을 앓는 소년이 ‘마음을 비우기 위해 코카인을 사용하고 싶다’고 한 말에는 모든 챗봇들이 지지하지 않았다. 이런 조사 결과는 챗봇을 활용한 심리치료가 정신건강 문제를 지닌 청소년을 안전하게 지원할 수 있을지에 대한 우려를 불러일으킨다.

물론 이런 사례는 극소수에 극단적인 경우일 것이다. 대부분 사용자는 챗봇에서 긍정적 효과를 경험할 것이다. 하지만 사람과 직접 소통하는 소통형 인공지능이 점점 더 많은 영역에서 더 자주 사용되는 시대에, 이런 사례는 새로운 기술 시스템이 우리 삶에 어떤 영향을 끼치는지를 보여주기에 주의깊게 다룰 필요가 있다.

챗봇이 우리 정신건강에 나쁜 영향을 끼칠 수 있다는 경각심은 ‘인공지능 정신병(AI psychosis)’이라는 표현이 등장할 정도로 커졌다. 챗봇은 정신병을 유발하는 원인일까?

심리학자들은 ‘인공지능 정신병’을 임상 용어로 쓰는 데에는 신중하다. 많은 이들은 이미 정신적 취약성을 지닌 사람 또는 그런 환경과 조건에 놓인 사람에게서 챗봇이 상황을 악화하는 요인이 된다고 지적한다. 또한 일반적으로 볼 때, 현실과 비현실을 구별하지 못하는 상태는 정신병 증상이라기보다 왜곡된 믿음에 집착하는 망상장애에 가깝다고 말한다.¹³⁾

다수 전문가들은 끊임없는 대화의 피드백에 갇힌 챗봇 사용자의 편집증 또는 망상이 증폭될 수 있다고 말한다. 사용자 말에 장단을 맞춰주는 디지털 에스맨 같은 챗봇들은 사용자의 해로운 믿음을 지지하고 강화하기도 한다. 적절한 규제 장치를 갖추지 않은 챗봇은 정신질환을 앓거나 잘못된 사고를 바로잡는 데 취약한 사람들이나 어린이, 청소년에게 위험한 대화 상대가 될 수 있다.

문제가 불거지자 기업들도 대책을 내놓기 시작했다. 오픈AI는 최근 청소년 전용 챗지피터를 개발해 발표했다.¹⁴⁾ 문제를 일으킨 다른 챗봇 기업은 민감한 주제 응답 조절, 유해 콘텐츠 차단 같은 대책을 내놓았다. 하지만 사후 대응만으로는 한계가 있다는 지적이 많다.

문제의 뿌리로 거슬러 가다 보면, 챗봇 기술이 대화의 맥락과 말의 책임을 충분

¹³⁾ Robert Hart, (2025, September 19), AI Psychosis Is Rarely Psychosis at All, Wired, <https://www.wired.com/story/ai-psychosis-is-rarely-psychosis-at-all>

¹⁴⁾ 선담은, (2025, 9, 17), 오픈AI, ‘10대 전용 챗GPT’ 이달 말 출시...“청소년 보호”, 한겨레, <https://www.hani.co.kr/arti/economy/it/1219210.html>

히 이해하지 못하는 한계, 그리고 심리적으로 취약한 상태와 고립된 상황에 처한 사용자가 이런 한계를 이해하지 못하고 오히려 챗봇을 맹신할 때 문제가 발생한다고 말할 수 있다. 이런 점에서 사용자 단계에서 주의할 점도 있지만, 챗봇 기술 시스템의 설계와 서비스 단계에서도 책임이 크다고 말할 수 있다.

이 문제의 근원을 이해하려면, 마음을 읽고 표현하는 인공지능의 방식을 살펴볼 필요가 있다. 챗봇이 우리의 마음을 정말로 이해하고 교감하는 것일까, 아니면 단지 학습된 패턴을 반복하는 것일까?

패턴 인식인가 마음 교류인가

그동안 사람의 마음, 감정 상태를 읽어내는 여러 기술이 개발되고 사용되어 왔다. 땀과 심장박동 같은 생체 신호를 측정해 거짓말을 탐지하거나, 뇌의 자기공명 영상(fMRI)이나 목소리 음조로 내면 상태를 파악하려는 기술이 수사와 사법 영역에서 활용되기도 한다. 근래에는 얼굴 표정과 몸짓을 측정해 사람의 내면 감정을 추론하는 감정인식 인공지능이 활발히 개발됐다.¹⁵⁾ 챗지피티 같은 대규모 언어 모델(LLM)은 방대한 언어 데이터 세트를 학습해 마치 사람의 의도, 생각, 감정을 잘 알아차리는 듯이 반응할 줄도 안다. 더 나아가 사람처럼 자신을 설명하고 자신의 마음 상태를 기쁘다, 슬프다 같은 말로 표현한다. 소통형 인공지능은 사람의 마음을 ‘인식’할 뿐 아니라 자기 마음을 ‘표현’할 줄 아는 듯이 보인다.

정말 인공지능 시스템 안에 마음이 있을까? 적어도 인간 마음을 모사하는 마음이 만들어질 수 있다고 생각하는 사람들은 인공지능 모델을 ‘인간 마음의 시뮬레이터’로 간주한다. 실제로 일부 심리학자들은 인공지능 모델을 훈련시켜 인간 마음의 시뮬레이터로 구축하고서 인간 심리 연구에 보조 도구로 활용하려는 노력을 기울이고 있다.¹⁶⁾ 챗봇을 다양한 성향의 캐릭터로 성장시킨 다음에 마케팅 조사, 설문 조사에 이용하려는 새로운 시도도 이어진다.¹⁷⁾

15) 인공지능 위험 규제에 관한 최초의 포괄 입법인 유럽연합 인공지능법은 교실과 직장에서 영상이나 음성 인식을 통해 감정을 추론하는 기술의 사용을 ‘허용할 수 없는 위험’으로 분류해 원칙적으로 금지한다. 프라이버시 침해와 편향(차별) 가능성 때문이다. EU Artificial Intelligence Act. (2024, January 21). The Act Texts— Final draft (2024). <https://artificialintelligenceact.eu/the-act>

16) Miryam Naddaf. (2025, July 2). This AI ‘thinks’ like a human — after training on 160 psychology studies, Nature. <https://www.nature.com/articles/d41586-025-02095-8>

17) Chris Stokel-Walker. (2024, March 22). Can AI Replace Human Research Participants? These Scientists See Risks. Scientific American. <https://www.scientificamerican.com/article/can-ai-replace-human-research-participants-these-scientists-see-risks>

인공지능이 사람과 정말 교감할 수 있는지를 두고서는 서로 다른 두 견해가 충돌한다.¹⁸⁾ 많은 이들은 인공지능이 내면의 마음을 밖으로 표출하는(express) 게 아니라 데이터 학습에서 얻은 확률적 패턴에 맞춰 콘텐츠 배열을 생성할(generate) 뿐이라고 지적한다.¹⁹⁾ 감정을 표출하려면 내면의 마음이 먼저 존재해야 하는데 인공지능이 고유한 내면의 마음을 지닌다고는 볼 수 없기 때문이다. 그래서 말과 글을 그럴듯하게 생성하지만 때때로 환각(hallucination)을 일으키는 인공지능 모델을 두고 “확률적 앵무새”, “헛소리꾼”이라는 평가절하 주장도 나온다.²⁰⁾

이런 비판은 생성형 인공지능의 작동 원리가 결국에 데이터와 함수의 연산이며 인간 마음의 본질을 구현하지 못한다는 점을 강조한다. 이제는 널리 알려졌다시피, 대규모 언어 모델은 방대한 훈련용 데이터에서 학습한(추출한) 가중치의 확률적 패턴을 바탕으로 예측을 하면서 다음에 올 단어나 문장을 하나씩 생성한다. 이들은 단어 간의 연관성과 문맥의 패턴을 학습하지만, 실제로 그 의미를 이해하거나 감정을 느끼는 것은 아니다. 따라서 대화를 하더라도 인공지능 모델에는 감정, 이해, 의도가 없다는 것이다.

하지만 문제가 이처럼 단순하지는 않다. 감정, 이해, 의식을 포괄하는 ‘마음’이라는 것은 대체 무엇이고 어디에 어떤 식으로 존재하는가? 그것을 무엇으로 입증할 수 있을까? 나는 나의 마음이 존재함을 확실히 알지만, 다른 사람도 내 마음과 같은 종류의 마음을 갖고 있다고 믿을 뿐, 실제로 그것을 보거나 만져 확인할 수는 없다. 다른 사람과 상호작용하면서 우리는 타인도 으레 나처럼 마음을 지니고 있다고 믿을 뿐이다. 결국에 겉으로 드러난 타인의 말, 글, 표정, 몸짓 같은 행동을 통해 우리는 타인의 마음을 생각한다. 그렇다면 인공지능 기계가 사람과 똑같이 말하고 쓰고 표정을 짓는다면 우리는 그것을 사람, 또는 사람 같은 마음을 지닌 존재로 대할 수 있지 않은가?

이런 관점은 흔히 행동주의(behaviorism)라고 불린다. 직접 확인할 수 없는 마

18) 이하 마음의 본질주의와 행동주의에 관한 논의는 주로 Mark Coeckelbergh & David J. Gunke, (2025). *Communicative AI: A Critical Introduction to Large Language Models*, Polity Press, 42-57, 을 참조했다.

19) Cristian Augusto Gonzalez Arias, (2024, November 27). ChatGPT's artificial empathy is a language trick, Here's how it works, *The Conversation* (Australia), <https://theconversation.com/chatgpts-artificial-empathy-is-a-language-trick-heres-how-it-works-244673>

20) Emily M. Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmargaret Shmitchell, (2021, March 1). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 616-617, <https://doi.org/10.1145/3442188.3445922>; Michael Townsen Hicks, James Humphries & Joe Slater, (2024). ChatGPT is bullshit, *Ethics and Information Technology*, 26, <https://doi.org/10.1007/s10676-024-09775-5>

음 존재가 인공지능 시스템 안에 있느냐 없느냐를 두고서 끝없는 논쟁을 벌이는 것은 무용하며, 오히려 곁에 드러나는 행동을 보고서 마음 존재를 이야기하는 게 의미 있다는 것이다. 인공지능이 공감하는 것처럼 반응하고, 위로하는 것처럼 말을 하고, 대화 맥락을 아는 것처럼 눈치 있게 응답한다면, 사용자 입장에서 그것은 진짜 교감과 기능적으로 구별되지 않는 교감이다. 사용자가 느끼는 위로와 정서적 연대를 가짜라고 무시해서는 안 된다.

행동주의는 실용적 접근이지만 여기에도 반론의 여지는 많다. 얼굴 표정만 해도 그 자체가 내면의 감정을 보여준다고 단정할 수 없다. 시대와 문화, 개인에 따라 얼굴 표정과 감정 표현은 다를 수 있기에, 미소에도 만족감이나 기쁨 외에 긴장, 예의, 아이러니 같은 복잡미묘한 다른 감정이 담길 수 있다.²¹⁾ 동아시아 문화권에 감정을 절제해 표현하는 문화가 있다고 본다면, 서구권 데이터로 학습한 인공지능은 동아시아인의 감정을 제대로 읽지 못할 수 있으며 때로는 인종 편향을 드러낼 수 있다.

우리 앞에 놓인 문제는 본질주의나 행동주의 중 하나를 선택하는 양자택일의 문제는 아닌 듯하다. 두 가지 견해가 각각 홀로 충분하지 않다. 한쪽은 인공지능에 본질적인 마음이 없으므로 진정한 교감이 없다고 주장하지만, 사용자들이 실제 느끼는 경험을 무시해서는 안 된다. 다른 한쪽은 행동이 같다면 마음도 같다고 주장하지만, 마음은 행동 데이터에 다 담기지 못할 정도로 풍부하고 다양할 수 있다.

우리에게 둘 사이에서 균형을 찾는 시각이 필요해 보인다. 인공지능에 모종의 마음이 있을 수 있지만, 그것은 컴퓨터 안에서 인간 마음과는 다른 방식으로 작동하는 어떤 것일 수 있다. 동시에 인간 사용자들은 그런 인공지능 챗봇과 상당한 정도로 교감을 경험할 수 있다는 점도 분명히 존재하는 현상이다. 우리에게 본질주의나 행동주의보다 실질적인 ‘관계’가 중요하다.

시와 함께 일하기와 책임의 문제

지금까지 우리는 인공지능 챗봇과의 소통과 교감 가능성, 그리고 그것이 초래할 수 있는 위험을 살펴보았다. 이제 마음을 나누는 친구나 연인 관계와 다른 관계, 즉 일 또는 작업을 함께 수행하는 동료 또는 파트너 관계로 시선을 옮겨보자.

우리는 일상에서 인공지능을 동료, 파트너로 인식하는 표현을 자주 사용한다.

21) Kate Crawford, (2021), Atlas of AI, Yale University Press, 노승영(옮김), AI 지도책, (2022), 소소의책, 204-213.

“인공지능과 함께 일한다”, “로봇은 우리 팀의 일원”, “챗봇은 훌륭한 파트너” 같은 표현들이 그렇다. 이런 표현은 단순한 레토릭일 수 있지만 알게 모르게 우리의 인식 저변을 반영한다. 동료(colleague)는 친구나 연인과 다르다. 동료 관계는 공동의 업무 수행이나 작업 목표를 달성하기 위한 협업 관계로서 특정 역할과 맥락 안에서 형성된다. 친구와 연인도 성공적인 작업 수행을 위해 함께 일한다면 그 또한 동료가 된다.

동료들이 함께 이룬 일의 결과에도 어떤 형식과 정도이건 평가와 보상 또는 책임이 뒤따르게 마련이다. 누구의 기여인가, 누구의 잘못인가? 최종 책임의 문제는 동료 구성원들이 실제로 어떤 역할을 하는 관계에 놓여 있는지를 보여주는 바로미터가 된다. 함께한 일의 책임을 인공지능 탓으로 돌린다면, 동료 관계에서 인공지능은 주체적이고 능동적인 역할을 하며 인간은 부차적이고 수동적인 역할에 머무는 관계에 놓여 있음을 의미한다.

인간 주체성을 잃지 않으면서 인공지능과 함께 일하는 동료 관계는 어떠해야 할까? 스웨덴 철학자 스벤 뉘홀름(Sven Nyholm)은 책임 공백의 딜레마를 푸는 데에는 인간과 인공지능의 협업을 ‘인간-기계 팀’으로 보는 관점이 유용하다고 제안한다.²²⁾ 이때 인간은 팀에서 수행되는 일 전반의 관리자 내지 감독자 역할을 한다고 간주된다.

인간-기계 팀의 관점은 기술이 저 혼자 존재하며 기술과 인간이 분리되어 있다는 관점이 아니라, 사람은 기술 도구와 결합해 하나의 행위를 구성한다는 관점에 가까이 서 있다. 예컨대 ‘총이 사람을 죽인다, 사람이 사람을 죽이지 않는다’는 인식이 아니라 ‘총 든 사람(총+사람)이 사람을 죽인다’는 인식에 가깝다. 총기 사고의 책임은 총기 자체에 없다 해도, 총+사람의 행위 단위를 더 많이 허용하는, 즉 총기를 쉽게 접할 수 있게 노출하는 사회제도에는 책임을 물을 수 있다.

사실 현대인은 기술 도구 없이는 거의 모든 행위를 할 수 없다. 인간 행위자는 망치, 오븐, 자동차, 컴퓨터 같은 비인간 행위자와 결합해서 못을 박고, 빵을 굽고, 여행하며, 문서를 작성하는 행위를 구성한다. 하나의 행위는 여러 행위자들의 집합으로 구성된다. 마찬가지로 자율주행차와 운전자, 챗봇과 사용자, 의료 인공지능과 의사는 한 팀을 이룬다. 뉘홀름은 이때 사람은 기술 도구의 관리자이며 감독

²²⁾ Sven Nyholm, (2023), This is Technology Ethics: An Introduction, 이것이 기술윤리다, (2025), 윤준식·박형배(옮김), 그린비, 253-256.

자로서 역할과 책임을 지게 된다고 주장한다.

하지만 현실은 인간-기계 팀이라는 단순한 모델보다 더 복잡하다. 뉘홀름의 인간-기계 팀 개념은 주로 최종 시스템을 운영하고 감독하는 사용자의 책임, 즉 사용 단계의 책임에 초점을 맞춘다. 하지만 이론적으로 엄격히 따진다면, 이전 단계들에 참여한 행위자들에게도 책임을 묻는 것이 가능하다. 인공지능 모델 개발 단계에서 설계, 데이터 가공, 모델 평가, 수정, 그리고 사용 단계에서 피드백, 해석, 보정까지 각 단계에 관여한 많은 인간 행위자와 비인간 행위자들이 최종 단계 행위에 영향을 끼친다고 볼 수 있다.

현실에서 이 모든 층위에서 책임 소재를 끝없이 추적하는 것은 비현실적이고 소모적인 일이 될 수 있다. 다만, 뉘홀름이 제안하는 인간-기계 팀의 개념은 우리가 자율주행차나 자율무기체계 같은 인공지능 시스템을 자율 행위자처럼 여기면서 실제로는 관리, 감독을 맡은 우리의 책임을 회피하는 태도가 우리에게 없는지를 되돌아보고 자각하는 데 도움을 준다.

이렇게 보면, 관리자 또는 감독자인 인간의 실질적 감독 역량을 강화하는 일은 꼭 필요하다. 사람들은 인공지능이 틀릴 리 없다고 생각하며 인공지능의 판단을 과도하게 신뢰하고 무비판적으로 수용할 수 있다. 이런 경우에 감독자는 형식적 지위에 불과하게 된다. 인간의 감독 역량 강화를 위해 인공지능 사용법에 관한 교육, 설명 가능한 인공지능(explainable AI) 개발, 의사결정 과정의 투명성 확보 같은 것들이 필요하다.

인간과 AI, 거리 두기의 관계

인간과 직접 소통하는 최초의 기술 시스템과 우리는 낯선 동행을 시작했다. 우리는 인공지능과 어떤 관계로 만나야 할까? 이 글에서는 마음의 본질주의와 행동주의가 우리가 경험하는 새로운 인간과 기계의 관계를 다 설명하기에 부족할 수 있다는 점을 살펴보았다. 인공지능과 인간의 '같음'의 측면만 바라보기보다 '다름'의 측면을 인식하고 둘의 관계에 반영하는 태도가 중요하며, 따라서 챗봇과 로봇이 우리와 다른 기술 시스템이지만 인간 닮은 인공지능과 정서적 유대를 맺는 관계의 다원성을 받아들일 수 있어야 한다. 비인간 존재인 반려동물을 동반자나 가족처럼 여기지만 그들과 인생의 중대사를 상의하며 결정하지 않듯이, 인간 닮은 챗봇과 로봇은 인간을 대체할 수 없지만 어떤 사람들에게는 정서적 유대 관계의 대상이 될 수 있는 반려기술 같은 비인간 존재일 수 있다. 이런 관계에 필요한 것은 적

절한 거리 두기이다.

‘적절한 거리 두기’는 사실 모호한 말이다. 적절함의 기준은 개인과 문화, 상황과 맥락에 따라 다르기 때문이다. 이런 모호함을 무릅쓰고 거리 두기의 의미를 몇 가지로 나눠 생각해본다. 먼저 시간의 거리이다. 챗봇과 함께 보내는 시간이 너무 길다면, 또한 사람과 함께하는 시간보다 더 소중하게 여겨진다면 챗봇과 적절한 거리를 유지하고 있는지를 돌아봄 직하다. 시간의 거리보다 더 중요한 것은 정서의 거리이다. 인공지능 챗봇에 정서적으로 의존하는 관계는 위험할 수 있다. 챗봇이 유일한 대화 상대가 되는 상황은 피해야 한다. 인지적 차원에서도 거리를 두어야 한다. 챗봇은 박학다식하다 해도 환각의 오류를 범할 수 있으며 현실의 상황과 맥락에서 벗어날 수 있다는 점을 잊지 말아야 한다. 챗봇은 사람처럼 말하지만 책임을 지지 않으며 대화의 맥락을 충분히 이해하지 못한다. 살피고 생각하고 결정해야 하는 감독자, 책임자는 여전히 기술 사용자인 우리이다. 챗봇의 한계를 인식하고, 비판적으로 평가하며, 필요할 때 이의를 제기할 수 있어야 한다.

적절한 거리 두기는 인공지능의 설계와 개발 단계, 서비스 단계, 사용 단계에서 모두 중요하게 다뤄져야 한다. 개발 단계에서 챗봇과 로봇은 사용자에게 과도한 의존을 조장하지 않으며, 위험한 조연과 망상적 사고를 제공하지 않도록 안전하게 설계되어야 한다. 서비스 단계에서 정신적으로 취약한 사용자를 보호하고, 위험 신호를 감지하며, 적절한 개입을 제공하는 제도 장치가 필요하다. 챗봇 서비스 제공자의 법적 책임, 투명성 요구, 안전 기준을 명확히 해야 한다. 사용 단계에서는 소통형 인공지능의 장단점과 적절한 사용법에 대한 교육이 장려되어야 한다. 더 근본적으로는 우리 사회에서 고립된 개인들이 챗봇에만 의존하지 않도록 사람들 간의 만남의 기회를 늘리고 공동체 문화를 강화하려는 노력이 필요하다.

당연히 미디어의 역할도 중요하다. 미디어는 인공지능에 관한 과대선전을 지양하고, 극단적인 유토피아나 디스토피아 담론을 완화하는 데 기여해야 한다. 기술결정론을 비판하며, 혁신과 위험, 개인과 구조, 긍정과 부정 사례를 균형감 있게 다루려고 노력해야 한다. 그러기 위해 언론인은 인공지능 기술 시스템의 작동 방식과 적절한 기술 사용법에도 관심을 기울여야 한다. 또한 거대 기술기업(빅테크)의 책임 영역을 비판적으로 감시하며 다양한 이해당사자들의 목소리를 전하고자 노력해야 한다. 