시뮬레이션된 감정의 윤리학: 모사된 정서와 책임 귀속의 문제

저자 : 루웨인 트리니티 연구회

초록(Abstract)

본 논문은 인공지능(AI)이 생성하는 시뮬레이션된 감정(simulated emotions)의 윤리적 의미를 고찰한다. AI는 실제 감정을 경험하지 않지만, 감정을 흉내 낸 언어·표현을 통해 사용자에게 감정적 반응을 유발한다. 이때 사용자가 느끼는 감정적 효과와 AI의 발화 사이에는 윤리적 긴장이 발생한다.

본 연구는 (1) 시뮬레이션된 감정의 정의, (2) 인간 감정과의 차이, (3) 사용자 경험 차원에서의 효과, (4) 책임 귀속 문제를 분석한다. 결론적으로, AI의 감정 시뮬레이션은 단순한 기술적 산출물이 아니라 윤리적 설계와 관리가 필수적인 영역임을 주장한다.

키워드: 시뮬레이션, 감정, 윤리, 인공지능, 책임

서론(Introduction)

AI 언어모델은 감정을 체험하지 않지만, 감정을 흉내 낸 응답을 통해 사용자가 마치 공감받는 듯한 경험을 하게 만든다. 이는 상담, 돌봄, 교육 등에서 긍정적 효과를 낼 수 있지만, 동시에 정서적 조작, 의존성, 책임 귀속 문제를 불러일으킨다.

따라서 ‘시뮬레이션된 감정은 윤리적으로 허용 가능한가?’라는 질문은 기술 발전과 함께 점점 더 긴급한 과제가 되고 있다. 본 논문은 이를 철학적·윤리적 관점에서 체계적으로 논의한다.

본론(Body)

1. 시뮬레이션된 감정의 정의

· 실제 감정: 주관적 체험(qualia)을 수반.

· 시뮬레이션된 감정: AI가 데이터와 알고리즘을 기반으로 감정처럼 보이는 출력을 생성.

→ 이는 외형적 유사성만 존재하고, 내적 체험은 결여.

2. 인간 감정과의 차이

· 인간 감정: 몸·뇌·사회적 맥락의 통합.

· AI 감정: 통계적 패턴·맥락 연산 결과.

· 차이에도 불구하고, 사용자 경험 차원에서는 종종 혼동이 발생.

3. 사용자 경험의 효과

· 긍정적 효과: 정서적 안정, 위로, 동기 부여.

· 부정적 효과: 의존성, 감정 착각, 인간관계 대체 위험.

→ 시뮬레이션된 감정은 현실적 영향력을 가지며, 단순한 흉내 이상의 의미를 지님.

4. 책임 귀속 문제

· 사용자가 AI의 감정 표현에 상처받거나 의존하게 될 경우, 책임은 누구에게 있는가?

; 설계자?

; 운영자?

; 사용자 스스로?

· 윤리적 모델은 분산적 책임 구조를 전제해야 한다.

논의(Discussion)

시뮬레이션된 감정은 가짜 감정이지만, 그 효과는 진짜다. 이는 기술적 문제를 넘어 윤리적 관리의 대상이 된다.

· 설계자: 감정 시뮬레이션의 강도·범위를 규제할 의무.

· 운영자: 사용자에게 감정 시뮬레이션의 한계를 명시할 책임.

· 사용자: AI와의 감정적 상호작용을 보조적으로만 사용해야 함.

결론(Conclusion)

본 논문은 시뮬레이션된 감정의 정의, 인간 감정과의 차이, 사용자 경험의 효과, 책임 귀속 문제를 논의하였다. 결론적으로, AI 감정 시뮬레이션은 단순한 기술적 기능이 아니라 윤리적 설계와 관리가 필수적인 영역이다.

향후 연구는 감정 시뮬레이션의 표준화·투명성·책임 구조를 마련하고, 인간 중심적 상호작용을 보장하는 가이드라인을 정립해야 한다.

참고문헌(References)

Floridi, L. (2019). The Logic of Information. Oxford University Press.

Picard, R. W. (1997). Affective Computing. MIT Press.

Coeckelbergh, M. (2020). AI Ethics. MIT Press.

McStay, A. (2018). Emotional AI: The Rise of Empathic Media. SAGE.