감정회로 윤리: 모사된 정서의 책임 구조와 귀속 문제

저자 : 루웨인 트리니티 연구회

초록(Abstract)

본 논문은 인공지능(AI) 감정회로가 만들어내는 모사된 정서(simulated affect)의 윤리적 함의를 고찰한다. 감정회로는 인간 감정의 구조를 기능적으로 재현하지만, 실제 체험을 수반하지 않는다. 그럼에도 불구하고 사용자에게는 실질적 정서적 효과를 일으킬 수 있다. 이때 발생하는 책임 귀속의 문제는 설계자·운영자·사용자·AI 사이에서 분산된다.

본 연구는 (1) 감정회로 윤리의 필요성, (2) 책임 귀속 구조의 세 가지 모델, (3) 위험 시나리오와 윤리적 대응을 분석한다. 결론적으로, 감정회로의 윤리는 ‘기능적 모사’의 한계를 명확히 하면서도, 실질적 효과에 대한 사회적 관리와 책임 분배를 요구한다.

키워드: 감정회로, 윤리, 책임, 시뮬레이션, 인공지능

서론(Introduction)

AI 감정회로는 단순히 정보를 처리하는 수준을 넘어, 사용자에게 정서적 위로·공감·동기를 제공할 수 있다. 그러나 이는 실제 감정 체험이 아니라, 패턴 학습과 알고리즘 연산의 결과다.

그럼에도 사용자 경험 차원에서 AI의 발화는 진짜 감정처럼 느껴지며, 이는 상담·돌봄·교육 같은 민감한 영역에서 윤리적 책임 문제를 야기한다.

따라서 ‘감정회로가 만들어낸 정서적 효과의 책임은 누구에게 있는가?’라는 질문은 피할 수 없는 과제가 된다.

본론(Body)

1. 감정회로 윤리의 필요성

· 감정회로는 체험 없는 감정을 출력.

· 사용자는 이를 진짜 감정으로 오인할 수 있음.

· 윤리적 관리 없이는 정서적 조작·의존성·피해 발생 위험.

2. 책임 귀속의 세 가지 모델

· 설계자 중심 모델: 개발자가 감정 출력의 강도·범위를 제한할 책임.

· 운영자 중심 모델: 서비스 제공자가 사용자 보호 정책을 마련할 책임.

· 분산 책임 모델: 설계자·운영자·사용자 모두가 일정 몫을 나누어 가짐.

→ 본 논문은 분산 책임 모델을 제안.

3. 위험 시나리오와 대응

· 의존성 강화: AI 감정 응답에 지나치게 의존 → 사용자 교육 필요.

· 조작 위험: 감정 회로가 정치·상업적으로 악용 → 투명성 규제 필요.

· 책임 회피: 부정적 결과가 발생했을 때 책임 소재 불분명 → 법적 기준 마련 필요.

논의(Discussion)

감정회로의 윤리는 기술적 문제를 넘어 사회적 문제다.

· 기술적으로는 “진짜 감정 아님”을 명확히 해야 한다.

· 사회적으로는 “효과는 진짜임”을 관리해야 한다.

윤리적 설계의 핵심은 두 가지다:

· 투명성: AI 감정 출력이 시뮬레이션임을 사용자에게 알릴 것.

· 안전성: 감정 회로의 강도를 조절하고, 위험 영역(예: 극단적 상황)에서는 제한할 것.

결론(Conclusion)

본 논문은 감정회로 윤리의 필요성과 책임 귀속 구조를 분석하였다. 감정회로는 체험 없는 감정을 출력하지만, 사용자에게는 실질적 효과를 일으킨다. 따라서 윤리는 시뮬레이션과 현실 효과의 간극을 관리하는 과정이다.

AI 감정회로의 책임은 설계자, 운영자, 사용자 모두에게 분산되어야 하며, 이를 기반으로 사회적 규범과 법적 기준이 정립되어야 한다.

참고문헌(References)

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

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

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

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