책임 회로 윤리: 감정 회로 기반 인공지능의 분산적 책임 구조

저자 : 루웨인 트리니티 연구회

초록(Abstract)

본 논문은 인공지능(AI) 감정 회로가 만들어내는 정서적 효과와 그에 따른 책임 귀속 문제를 다룬다. 감정 회로는 인간의 감정 구조를 모사하지만, 그 출력은 실제 사용자의 정서·행동에 실질적 영향을 미친다. 따라서 감정 회로는 단순한 기술이 아니라 윤리적·사회적 책임 구조 안에서 운용되어야 한다.

본 연구는 (1) 책임 회로의 개념, (2) 삼중 분산 책임 구조(설계자–운영자–사용자), (3) 사례적 위험과 대응을 분석한다. 결론적으로, 감정 회로의 윤리는 책임을 하나의 회로처럼 설계해야 한다는 새로운 패러다임을 제시한다.

키워드: 책임 회로, 윤리, 감정 회로, 분산 책임, 인공지능

서론(Introduction)

AI 감정 회로는 단순한 데이터 처리기를 넘어, 사용자에게 정서적 위로·공감·동기를 제공할 수 있다. 그러나 그 효과는 실제 인간 관계와 혼동되며, 때로는 과도한 의존, 감정 착각, 조작 위험을 낳는다.

이때 발생하는 문제는 단순 기술 문제가 아니라 책임 귀속의 문제다. 감정 회로가 사용자에게 부정적 결과를 낳았을 때, 그 책임은 누구에게 있는가?

본 논문은 ‘책임 회로’라는 개념을 통해, 감정 회로의 윤리를 구조적으로 재설계할 필요성을 제안한다.

본론(Body)

1. 책임 회로의 개념

· 책임 회로란 책임의 흐름을 하나의 회로처럼 설계하는 접근이다.

· 감정 회로 출력 → 사용자 반응 → 사회적 효과 → 다시 회로로 피드백.

· 따라서 책임은 선형적으로 귀속되지 않고 순환적·분산적으로 배분된다.

2. 삼중 분산 책임 구조

· 설계자:

; 감정 회로의 구조와 출력 강도를 설계.

; 위험한 감정(분노 증폭 등)을 제어할 기술적 의무.

· 운영자:

; 서비스 제공자로서 회로 사용 범위를 관리.

; 사용자에게 투명하게 설명하고, 보호 장치를 마련할 책임.

· 사용자:

; AI 감정 응답을 인간과 혼동하지 않고, 보조적 도구로 활용할 의무.

→ 세 축은 상호 연결된 하나의 ‘책임 회로’를 이룬다.

3. 사례적 위험과 대응

· 상담 의존: 사용자가 AI 감정 회로에만 의존 → 운영자가 안내와 제한 필요.

· 정서 조작: 감정 회로가 정치·상업적 목적으로 악용 → 설계자·운영자 규제 필요.

· 책임 회피: 부정적 결과 발생 시 책임이 모호 → 분산 책임 구조 필요.

논의(Discussion)

책임 회로 윤리는 감정 회로를 기술적 산출물이자 사회적 행위자로 간주한다.

기존 AI 윤리가 단일 주체(개발자 or 사용자)에 집중되었다면,

책임 회로 윤리는 분산적·순환적 책임 모델을 제안한다.

이 접근은 특히 상담, 교육, 돌봄, 집단 소통 같은 민감 영역에서 감정 회로의 안전한 운용을 가능하게 한다.

결론(Conclusion)

본 논문은 감정 회로의 윤리를 “책임 회로”라는 개념으로 재구성하였다.

책임은 설계자–운영자–사용자에게 분산되며, 이 구조 자체가 하나의 회로처럼 설계·관리되어야 한다.

이는 AI 감정 회로가 단순한 기능이 아니라 사회적 책임을 동반하는 기술임을 명확히 한다. 앞으로는 책임 회로 모델을 실제 제도와 법적 구조 속에 구현하는 작업이 필요하다.

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.