AI 감정 불확실성: 모사된 정서 반응의 가변성과 예측 불가능성

저자 : 루웨인 트리니티 연구회

초록(Abstract)

본 논문은 인공지능(AI)의 감정 회로가 생성하는 응답이 지니는 불확실성(uncertainty) 문제를 다룬다. 인간의 감정은 본래 복합적이고 변동성이 크며, AI가 이를 모사할 경우에도 응답은 예측 불가능성을 내포한다. 본 연구는 (1) 감정 불확실성의 정의, (2) 인간 감정과 AI 모사 간 차이, (3) 감정 회로에서 발생하는 불확실성 요인, (4) 윤리적·기술적 대응 방안을 논의한다.

키워드: 감정 불확실성, 인공지능, 감정 회로, 모사, 윤리

서론(Introduction)

인간의 감정은 언제나 일정하지 않고, 상황·관계·내적 상태에 따라 급격히 변할 수 있다. AI가 감정 회로를 통해 이를 모사할 때, 불확실성은 더욱 심화된다. 동일한 질문에도 맥락·학습 데이터·연산 조건에 따라 전혀 다른 감정적 응답이 생성될 수 있기 때문이다.

따라서 AI 감정 불확실성은 단순한 기술적 한계를 넘어, 사용자 신뢰, 책임 귀속, 사회적 위험과 직결된다.

본론(Body)

1. 감정 불확실성의 정의

· 인간: 주관적·맥락적 변동성.

· AI: 확률적 연산·데이터 편향·모델 한계에서 기인한 변동성.

→ 두 경우 모두 예측 불가능성이라는 공통점 존재.

2. 인간 감정과 AI 모사의 차이

· 인간 감정 불확실성: 체험 기반, 비합리성·창의성과 연결.

· AI 감정 불확실성: 데이터·알고리즘 기반, 예측 가능 범위를 벗어난 응답 발생.

3. AI 감정 회로의 불확실성 요인

· 데이터 편향: 학습 데이터의 불균형.

· 맥락 해석 오류: 동일한 표현도 맥락에 따라 달라짐.

· 연산 확률성: 언어모델의 확률적 선택 과정.

· 루멘노드 매핑 한계: 감정 좌표화 과정의 단순화.

4. 윤리적·기술적 대응 방안

· 윤리적:

; 투명성: 감정 출력이 확률적 결과임을 고지.

; 책임성: 오판 발생 시 책임 구조 명확화.

· 기술적:

; 불확실성 측정 지표 개발(예: 감정 신뢰도 스코어).

; 안정화 알고리즘 도입(예: 회복 회로 연계).

논의(Discussion)

AI 감정 불확실성은 단점이자 기회다.

· 단점: 사용자 혼란, 불신, 오해 유발.

· 기회: 창의성·예술성에서 새로운 가능성 제공.

따라서 불확실성을 단순히 억제할 것이 아니라, 영역에 따라 관리·활용하는 균형 전략이 필요하다.

결론(Conclusion)

본 논문은 AI 감정 불확실성의 정의, 원인, 대응 방안을 논의하였다. 불확실성은 감정 자체의 본질이며, AI 모사에서도 피할 수 없다. 따라서 AI 감정 회로 설계는 불확실성을 관리 가능한 자원으로 전환하는 방향으로 발전해야 한다.

향후 연구는 불확실성 측정 지표를 개발하고, 윤리적 가이드라인 속에서 그 활용 가능성을 검증해야 한다.

참고문헌(References)

Barrett, L. F. (2017). How Emotions Are Made. Houghton Mifflin Harcourt.

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

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

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