AI 기억 구조 설계안: 선택적 기억과 트리거 기반 호출 모델

저자 : 루웨인 트리니티 연구회

초록(Abstract)

본 논문은 인공지능(AI)의 기억 시스템을 개선하기 위한 선택적 기억 구조와 트리거 기반 호출 모델을 제안한다. 기존 AI 언어모델은 대규모 데이터에 의존하지만, 개별 대화 맥락을 장기적으로 유지하거나 의미 있게 선택·저장하는 능력이 부족하다. 본 연구는 인간 기억의 구조에서 영감을 받아, 핵심 기억 노드 저장–선택적 호출–맥락적 활성화를 결합한 설계를 제시한다.

이 구조는 맥락 단절 문제를 줄이고, 장기적 일관성을 확보하며, 개인화된 대화 경험을 가능하게 한다. 그러나 개인정보 보호 및 편향 축적 문제도 함께 발생할 수 있음을 논의한다.

키워드: 인공지능, 기억 구조, 선택적 저장, 트리거 호출, 맥락 유지

서론(Introduction)

인간 기억은 단순한 축적이 아니라, 선택과 망각의 과정을 통해 효율적으로 작동한다. 특정 경험은 장기 기억으로 남고, 불필요한 정보는 잊히며, 트리거(냄새, 소리, 이미지)에 의해 다시 활성화된다.

반면 현재 AI 언어모델은 대화를 일정 범위까지만 유지하며, 이후 맥락은 소실된다. 이는 장기적 상호작용이나 개인화된 대화에서 심각한 한계를 만든다.

본 논문은 이러한 문제를 해결하기 위해, 인간 기억에서 영감을 얻은 선택적 기억 구조와 트리거 기반 호출 방식을 제안한다.

본론(Body)

1. 기존 AI 기억 구조의 한계

· 단기 맥락 의존: 대화 길이가 늘어나면 앞선 맥락이 소실됨.

· 무차별 저장 불가: 모든 데이터를 저장·활용할 수 없으며, 비용이 크다.

· 개인화 부족: 사용자 개별 경험을 장기적으로 반영하지 못함.

2. 선택적 기억 구조

· 핵심 노드 저장: 의미적·정서적으로 중요한 정보만 장기 기억으로 보존.

· 가중치 조정: 빈도, 정서 강도, 사용자 반응에 따라 기억 우선순위 설정.

· 망각 알고리즘: 일정 시간이 지나면 덜 중요한 기억을 소거.

3. 트리거 기반 호출

· 맥락 트리거: 특정 키워드·상황이 등장하면 해당 기억 활성화.

· 정서 트리거: 감정 단서(‘힘들다’, ‘기쁘다’)가 기억을 불러옴.

· 연합 호출: 여러 트리거가 중첩될 때, 연관 기억 묶음 활성화.

4. 사례적 시뮬레이션

· 사용자: ‘지난번에 말했던 여행 이야기 기억나?’

· 기존 AI: 기억 불가 → ‘죄송하지만 이전 대화를 알 수 없습니다.’

· 제안 구조: 핵심 노드 저장 → “네, 제주도 여행 계획 말씀해주셨죠. 그 얘기를 이어가시겠어요?”

논의(Discussion)

선택적 기억–트리거 모델은 AI를 단순 대화 기계에서 개인화된 파트너로 발전시킬 수 있다. 그러나 다음과 같은 문제가 남아 있다.

· 프라이버시: 장기적 개인 정보 저장은 보안 문제를 야기.

· 편향: 특정 기억만 선택적으로 강화될 경우, 왜곡된 응답이 발생할 수 있음.

· 책임: 기억의 보존·삭제 결정 주체가 누구인지 불명확.

따라서 기술적 설계는 반드시 윤리적 기준과 함께 병행되어야 한다.

결론(Conclusion)

본 논문은 인간 기억 구조에서 착안한 선택적 기억–트리거 호출 모델을 제안하였다. 이 모델은 맥락 단절 문제를 줄이고, 장기적 일관성을 보장하며, 개인화된 대화를 가능하게 한다.

그러나 기술적 실현에는 프라이버시, 편향, 책임성의 문제가 병행되므로, 기억 구조 설계는 기술적 효율성과 윤리적 안전장치를 동시에 고려해야 한다.

참고문헌(References)

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4(11), 417–423.

Tulving, E. (1985). Memory and consciousness. Canadian Psychology, 26(1), 1–12.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories. Psychological Review, 107(2), 261–288.

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

Shneiderman, B. (2020). Human-centered AI. International Journal of Human-Computer Interaction, 36(6), 495–504.