

다중 언어 모델 기반 멀티 에이전트 챗봇 토론 프레임워크를 통한 환각 탐지 및 교정

송인혁 (컴퓨터공학과), 김은호 (소프트웨어융합학과), 박상근 (소프트웨어융합학과)



이 연구를 진행하게 된 이유가 뭐야?



Gemini

요즘 사람들은 궁금한 게 생기면 더 이상 검색창에 입력하지 않아.
대신, 자연스럽게 묻고 답하는 방식으로 LLM 챗봇에게 질문해!
그에 따라 **환각(hallucination)**이라는 현상도 새로운 문제로 함께 부상하고 있어!

정보 탐색 패러다임의 변화

그에 따른 환각 문제

Gemini의 말에 전적으로 동의해! 또 다른 이유는 환각 해결을 위한 선행 연구에서도 찾을 수 있어.
환각 문제를 방지, 탐지, 및 개선하기 위해 진행된 많은 연구들이 **단일 모델 기반**으로 진행되었어.
하지만 단일 모델 기반 설계는 **여러 관점을 고려해 답변을 제공할 수 없다는 한계**가 있어!



ChatGPT

여러 선행 연구 진행됨

하지만 여전히 여러 관점에서 답변 제공 어려움



Perplexity

그래서 본 연구는 **멀티 챗봇 토론 프레임워크**를 제안해!
사전 학습된 복수의 생성형 모델을 활용해서 서로의 답변을 비교 검증하는 식이야.
이를 통해 (1) **환각을 효과적으로 탐지**하고, (2) **모델 간 관점을 융합한 개선된 답변**을 생성할 수 있어!

멀티 챗봇 토론 프레임워크 제안!

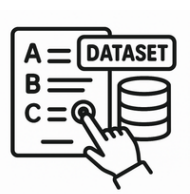
환각 효과적으로 탐지 및 개선 가능!

METHOD

개발



수집

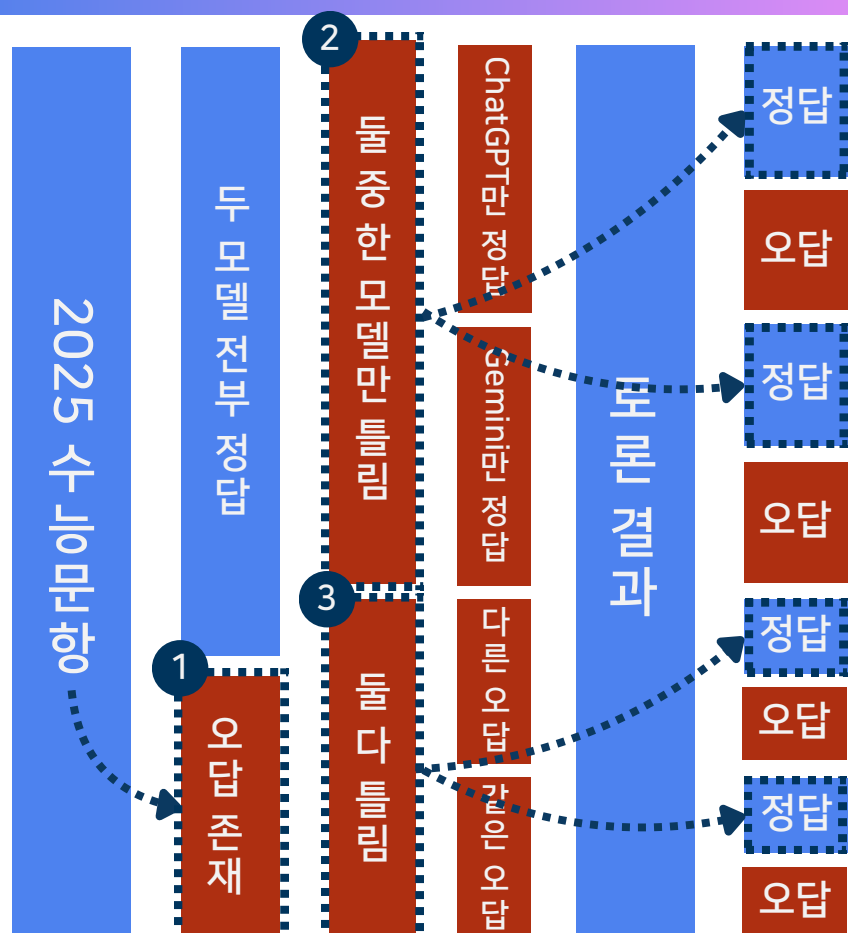


검증

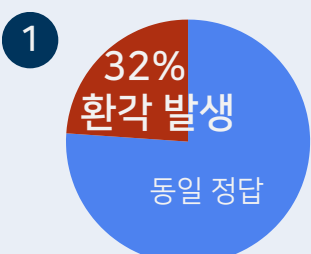


멀티 챗봇 토론 프레임워크를 구현
답변이 걸린 **2025 수능 66문항**으로 질의 구성
단일 & 멀티 모델 **비교 실험**을 통한 성능 검증

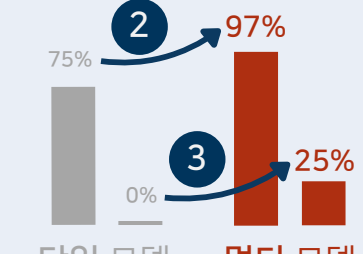
RESULT



빈번한 환각 발생



유의미한 성능 향상



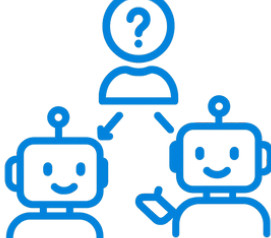
ARCHITECTURE



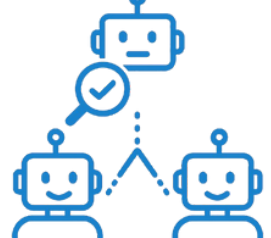
① 사용자 질문 입력



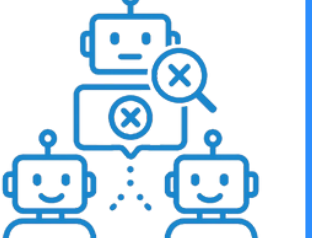
② 동일 질문 전달



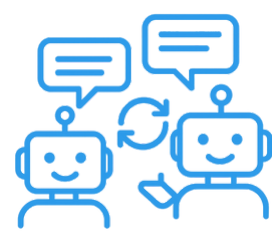
③ 답변 일치 여부 확인



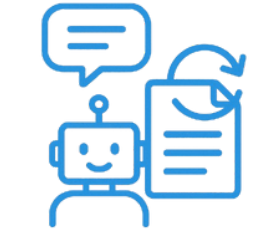
④ 불일치 시 팝업창 제시



⑤ 클릭 시 토론 시작!



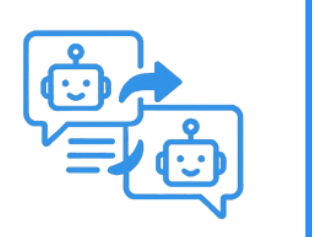
⑥ 전체 토론 내용 요약



⑦ 전체 토론 내용 제시



⑧ 추가 토론 진행



CONCLUSION & FUTURE WORKS

- ① 멀티 모델 기반 토론을 통한 환각 탐지 및 개선 가능성 제시
- ② 여러 관점을 융합해 개선된 답변 제공 및 플러그인 구현
- ③ 향후 개방형 질문을 추가해 모델의 일반성 및 실용성 향상