

2025년 4월 새싹 미니프로젝트

MultiAgent와 AutoGen으로 구현한 통합 검색 에이전트

2조 황의철, 김실헬, 박승현

01 개요

02 관련 연구

03 아키텍처

04 방법론

05 테스트 및 분석

06 결론

1. 정보 과부하 속, 바쁜 현대인을 위한 서비스의 필요성
2. 입력 한 번으로 정보를 요약해주는 챗봇의 필요성
3. '쇼츠 중독' 시대, 긴 콘텐츠 기피 현상
4. 사용자 친화적 서비스 설계



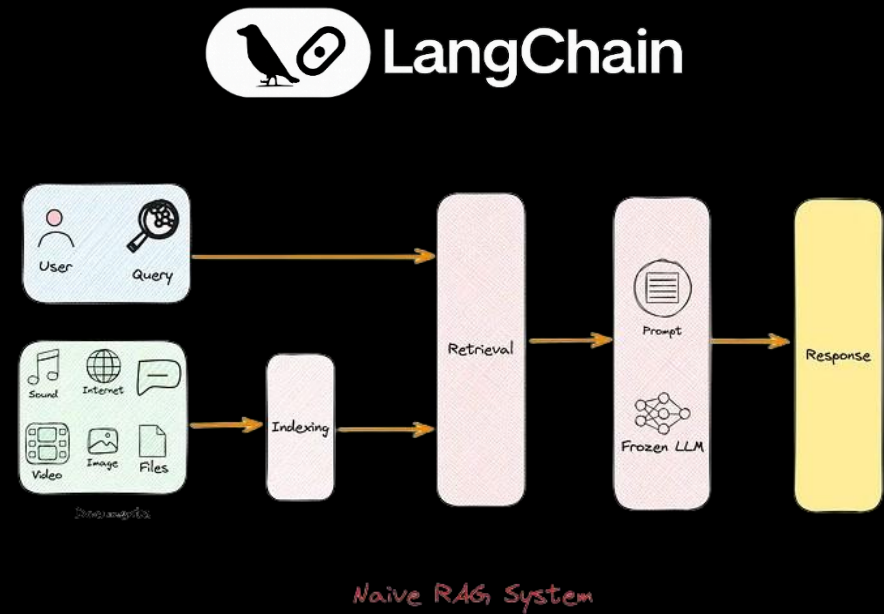
기존 문서 기반 검색 시스템의 한계

1. 정적 문서 기반 검색

- 사용자 질문이 아무리 다양해도, 항상 고정된 문서에서만 정보를 추출
- 실시간 이슈나 컨텍스트를 반영하기 어려움

2. 단일 응답 중심의 구조

- 기본 구조, 질의 → 검색 → 요약 → 응답
- 사용자의 입력에 따른 의도 파악이나 맥락 및 분기 처리 부족
- 예: “뉴스 기반으로 찾아줘” 와 같은 요청을 문서 기반 시스템은 유연하게 판단하기 어려움



멀티 에이전트 시스템의 도입

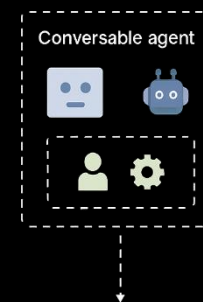
멀티 에이전트의 활용

- 각 에이전트는 자체적인 목표와 판단 능력을 가지고, 상호작용을 통해 더 복잡하고 다양한 작업을 수행
- 사용자 질의에 따라 에이전트를 선택적으로 실행하여 효율적이고 최적화 된 응답을 제공

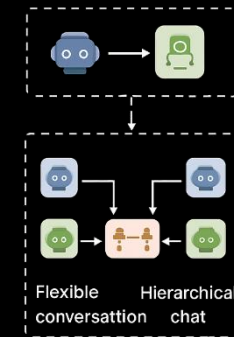
AutoGen 프레임워크 도입

- AutoGen은 Microsoft에서 제안한 LLM 기반 멀티에이전트 프레임워크
- 기존 멀티 에이전트 시스템의 개념을 활용하여 각 LLM 에이전트들이 대화 기반으로 협업하며 태스크별 다양한 작업을 수행 할 수 있도록 설계
- □ 참고 논문: [AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework](#) Microsoft Research, 2023. [arXiv:2309.11487](#)

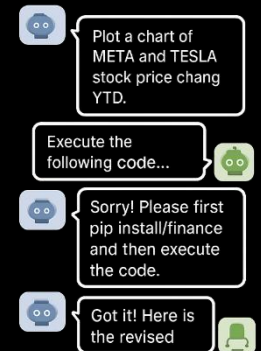
Agent Customization



Flexible conversations



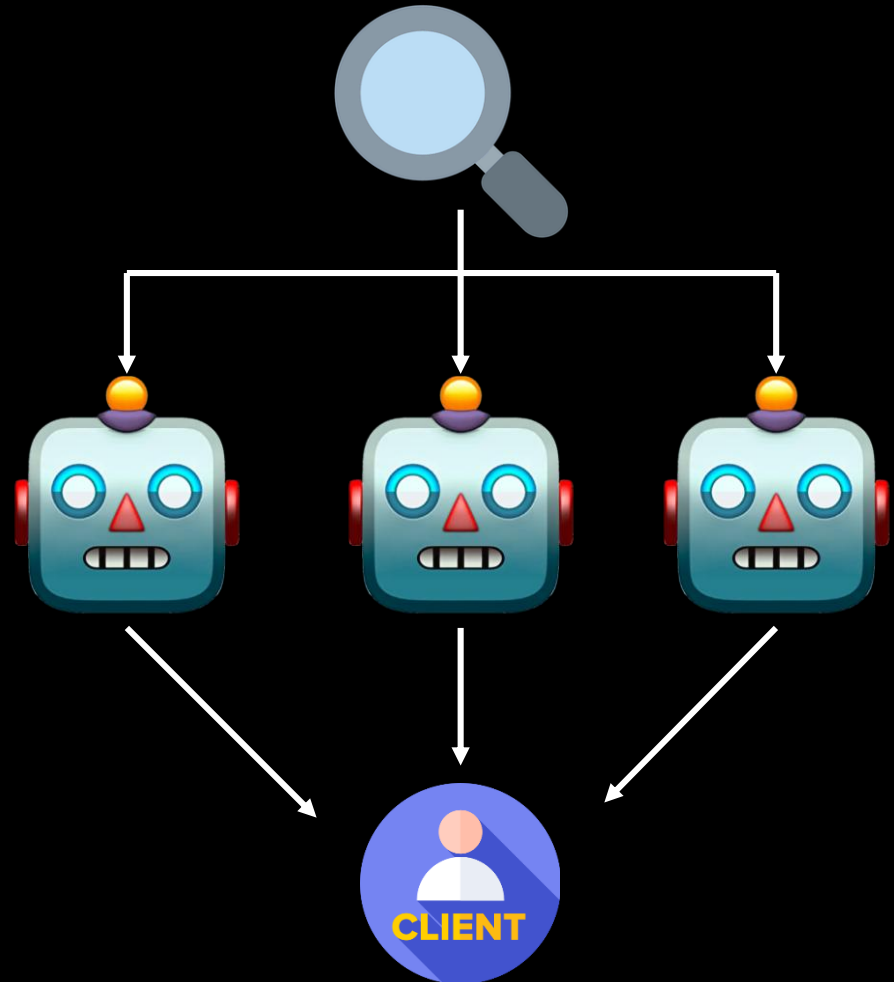
Flexibe conversations

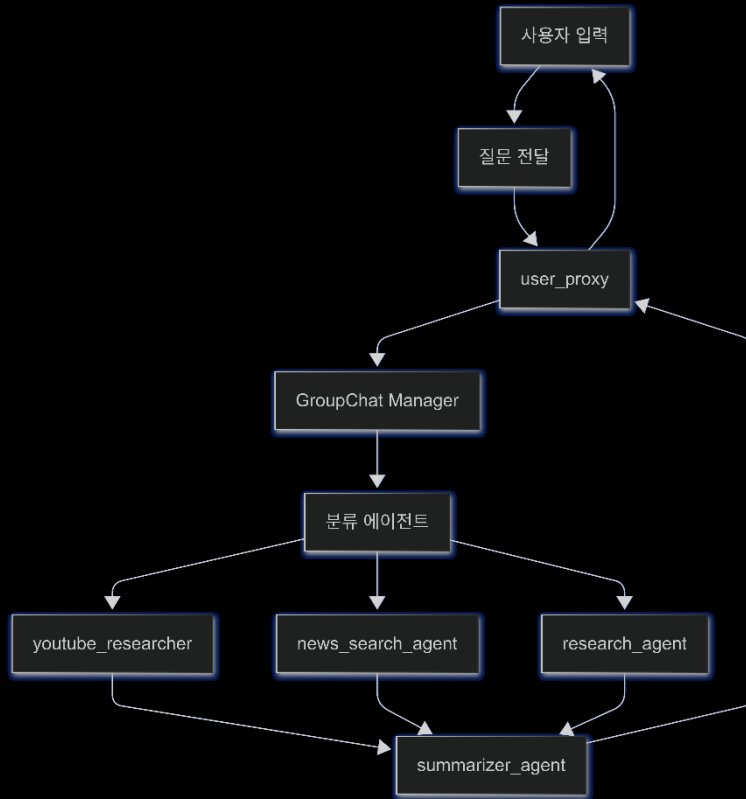


Example Agent Chat

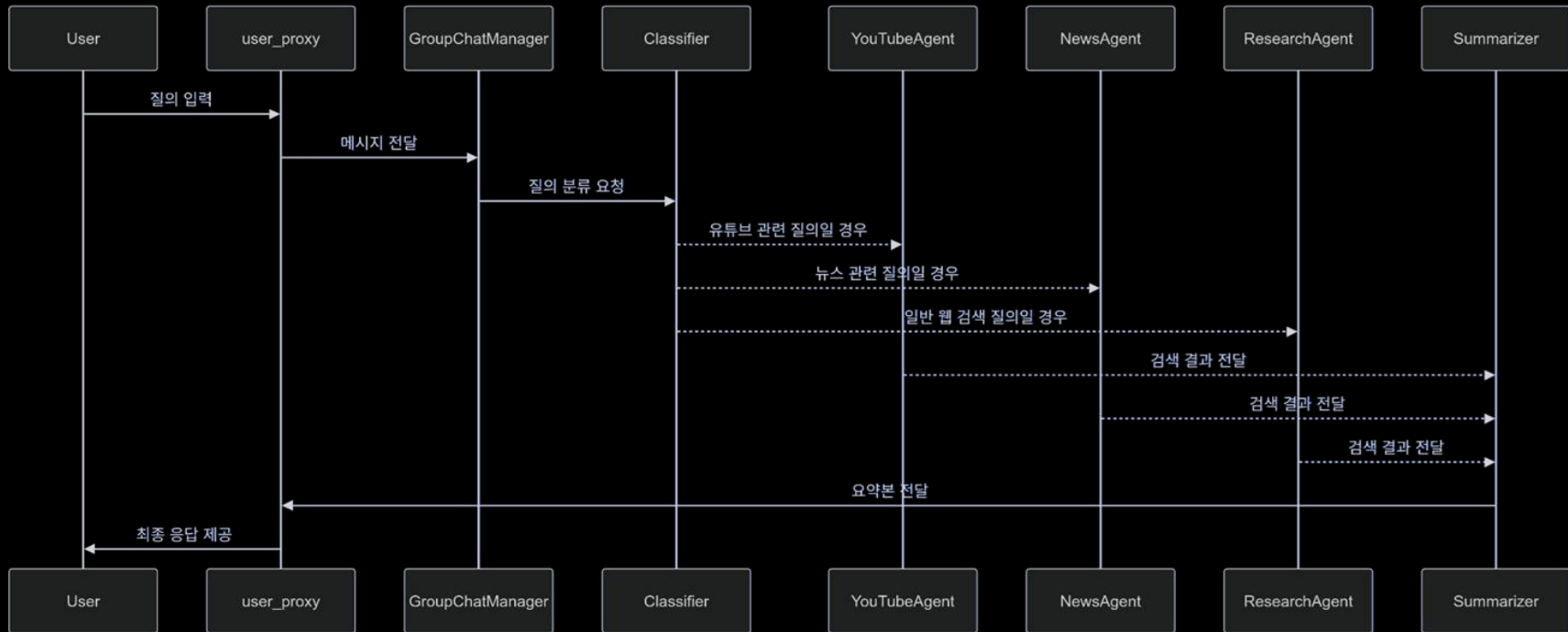
질문 하나로 다양한 도메인의
전문 에이전트를 통해 검색 후 응답

역할 기반 구조를 통해 보다
나은 정확도를 확보





구성 요소	역할
User_Proxy	사용자의 메시지 관리 및 인터페이스로 전달하는 역할
Group Chat Manager	전체 에이전트 간 대화 흐름을 조율하고 응답 흐름 관리
Classifier Manager	질문 내용을 분석하여 적절한 에이전트로 분기
Youtube Researcher	키워드를 통해 유튜브 영상 검색 후 영상 자막을 통한 내용 분석 정리
News Search	뉴스 기사 중심으로 관련 정보 수집
Research	일반 웹 기반 자료 검색
Summarizer	각 검색 결과를 통합하고 요약하여 최종 응답 생성



“Few Shot Prompt” 기반 사용자 질의 분류

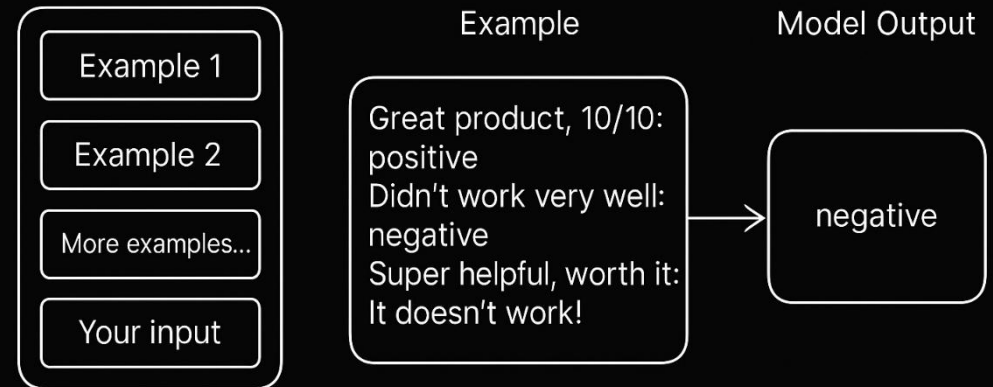
예시를 보여주며 분류 에이전트에게
질의에 따라 어떤 에이전트에게
전달할지를 사전에 학습 시킴

User: 오늘 뉴스 영상 내용 좀 요약해줘
→ 전달 키워드: "오늘 뉴스 영상"
→ 전달 대상: "youtube_research"

User: 오늘의 뉴스 내용 요약해줘
→ 전달 키워드: "오늘의 뉴스"
→ 전달 대상: "news_search_agent"



User: 잡채 요리방법 검색해서 요약해줘
→ 전달 키워드: "잡채 요리방법"
→ 전달 대상: "research_agent"

A Few Shot Prompt

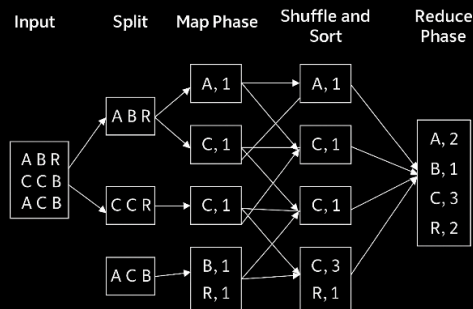




Tools

에이전트	사용 함수	추출 방식	처리 방식	비고
Search	tavily_search 	Tavily API 기반 웹 검색으로 데이터 추출	수집된 데이터 Summarizer 에이전트에게 인계	간단한 텍스트 기반 검색
News Search	news_search_Serper 	Serper API 기반 뉴스 도메인 특화 검색으로 기사 추출	수집된 데이터 Summarizer 에이전트에게 인계	실시간 뉴스 정보에 최적화
Youtube Search	youtube_script 	Youtube API를 통해 영상 검색 → 자막 스크립트 추출	LangChain Summarize 함수 사용 → Map-Reduce 방식 으로 요약하여 핵심 내용 도출	대량의 스크립트를 효율적으로 요약하기 위해 맵-리듀스 방식으로 분할하여 데 이터 리파인

Map – Reduce



청크 단위 요약(Map) 후, 같은 키워드를 기준으로 그룹화해 중복 요약
통합(Reduce) → 대용량 스크립트를 효율적으로 처리

1개 영상 당 평균 8,000 – 10,000 토큰 이상 발생하는
유튜브 스크립트를 정리하기 위한 효율적 방법

결과 요약

질의 유형	의도한 프롬프트	에이전트 분류 결과	품질	이슈 및 특이사항
"오늘 뉴스 요약해줘"	뉴스 요약 프롬프트	News Research	우수	기사 길이 적당함, 응답 안정적
"잡채 만드는 방법 찾아줘"	일반 검색 프롬프트	Research	양호	추출되는 정보 content 양의 부족으로 핵심 내용 요약 부족하는 경우 발생
"슈카월드 오늘 영상 요약 해줘"	유튜브 검색 + 자막 추출 프롬프트	Youtube Research	양호(자막 품질 이상 발생)	자동생성 자막 자체를 추출했기에 발음 오류, 문 장 단절 등의 이슈가 있었 으나 문맥 흐름에는 문제가 없어 정상 처리

문제 발생 및 해결방안

- 8,000토큰 이상의 과도한 텍스트 수집
 - 유튜브 자막, Tavily search 결과 content가 너무 많아 → 응답 오류 발생(토큰 초과), 질의 무시, 느린 응답 발생
- 토큰 수 제한
 - 수집 시점에서 2,000 – 3,000자 이내로 필터링
- Map-Reduce 방식 채택
 - 긴 텍스트를 청크로 분할 → 중복 제거 및 통합

한계

항목	설명
날짜 인식 오류	데이터 추출시 날짜 파싱이 가능한 부분이 없어 모델이 임의로 날짜를 생성하는 할루시네이션 발생
유튜브 자막 품질	자동 생성 자막 기반을 추출하기 때문에 어눌한 표현, 부정확한 문장 등이 요약 반영 됨
일반 검색 품질 저하	과도한 API 토큰 제한으로 인해 정보가 축약 되어 요약 결과가 부실해지는 가능성 발생
문서 기반 검색 미탑재	RAG 기능은 시간관계상 생략

개선 예정

항목	개선 방안
자막 품질 개선	Whisper API, 실제 동영상 자막 사전 다운로드 등 구비 예정
요약 성능 향상	토큰 제한을 고려하여 사전 추출 → Youtube Search와 비슷하게 청킹 후 요약
문서 기반 검색 추가	추후 Doc QA + PGVector 기반 RAG 시스템 확장 예정
날짜 추출 정확도	날짜 정규식 및 API가 아닌 직접 데이터를 추출하는 함수 설계 예정

Q & A