

SSAFY 2024 겨울 계절학기 최종 프로젝트 발표

직무기술서 기반 지원기업 맞춤 연계 QA 챗봇 서비스

목차

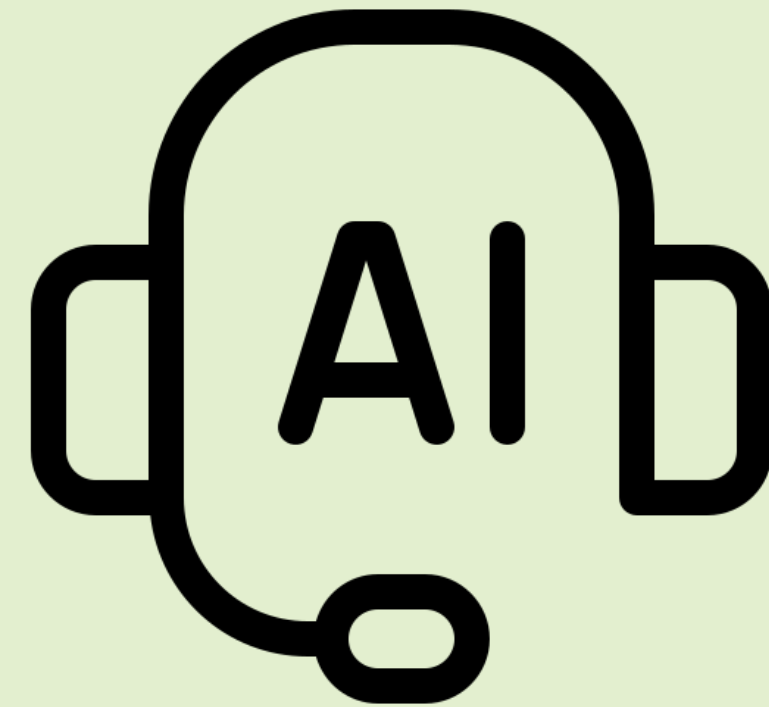
- 01 프로젝트 개요
- 02 QA 데이터셋 구축
- 03 서비스 구축 및 배포
- 04 프로젝트 개선사항
- 05 영상 시연

01. 프로젝트 개요



직무기술서의 중요도는 나날이 강조되나
이에 대한 탐색 및 분석엔 여전히 많은 취업 준비생의 노력이 투입됨

01. 프로젝트 개요



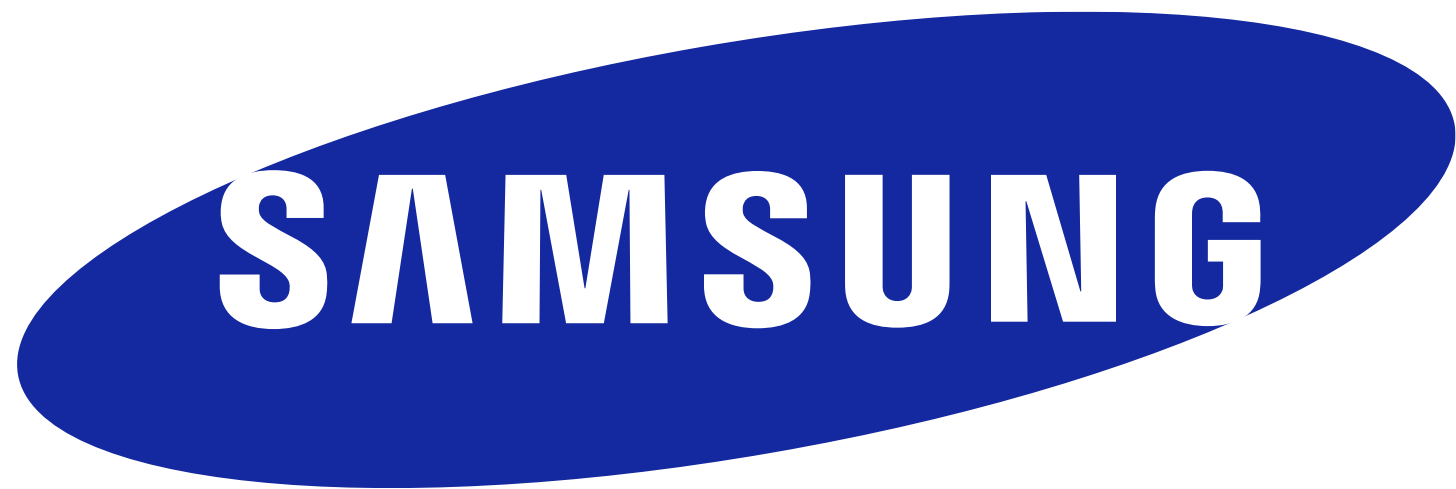
직무기술서를 바탕으로 학습시킨 LLM을 통해, 취업준비생에게 필요한 부분을
좀 더 쉽게 접근할 수 있도록 한 챗봇 서비스

01. 프로젝트 개요



- | **서비스 목적** - LLM 기반 챗봇을 통한 직무기술서 접근성 강화 서비스
- | **타겟 소비자** - 목표 기업, 직무가 존재하지만, 관련 정보가 부족한 취업준비생
- | **서비스 설명** - 직무기술서 정보를 바탕으로 LLM 챗봇 학습.
 1. 사용자가 원하는 기업 직무에 관련된 접근성 향상.
 2. 사용자가 보유한 역량, 경험을 바탕으로, 이를 우대하는 기업, 직무를 제안.

02. QA 데이터셋 구축

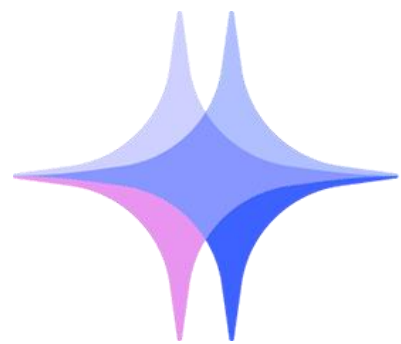


| 수집 데이터 - 삼성그룹 14개사의 2024년 채용공고, 직무기술서 PDF 파일

| 수집 절차

- 수집 기간: 2024/12/26 ~ 2024/12/28
- 수집 방법: 공식 채용 사이트를 통한 직무기술서, 채용 공고 PDF 파일 수집

02. QA 데이터셋 구축



upstage
Solar LLM

QA 데이터 구축

1. 수집 PDF 데이터의 Markdown 형태 변환.
2. Upstage Solar LLM을 통해 QA Pair 데이터셋 생성
(chunk = 1024, overlap = 10)
3. 4가지 질문 분류 체계 구축 (기업분석, 직무분석, 필요역량, 우대역량)

데이터셋 평가

1. DeepEval을 활용해 취합된 데이터셋에 대한 자동화 평가
2. Matplotlib를 사용한 평가 메트릭 시각화 진행

02. QA 데이터셋 구축

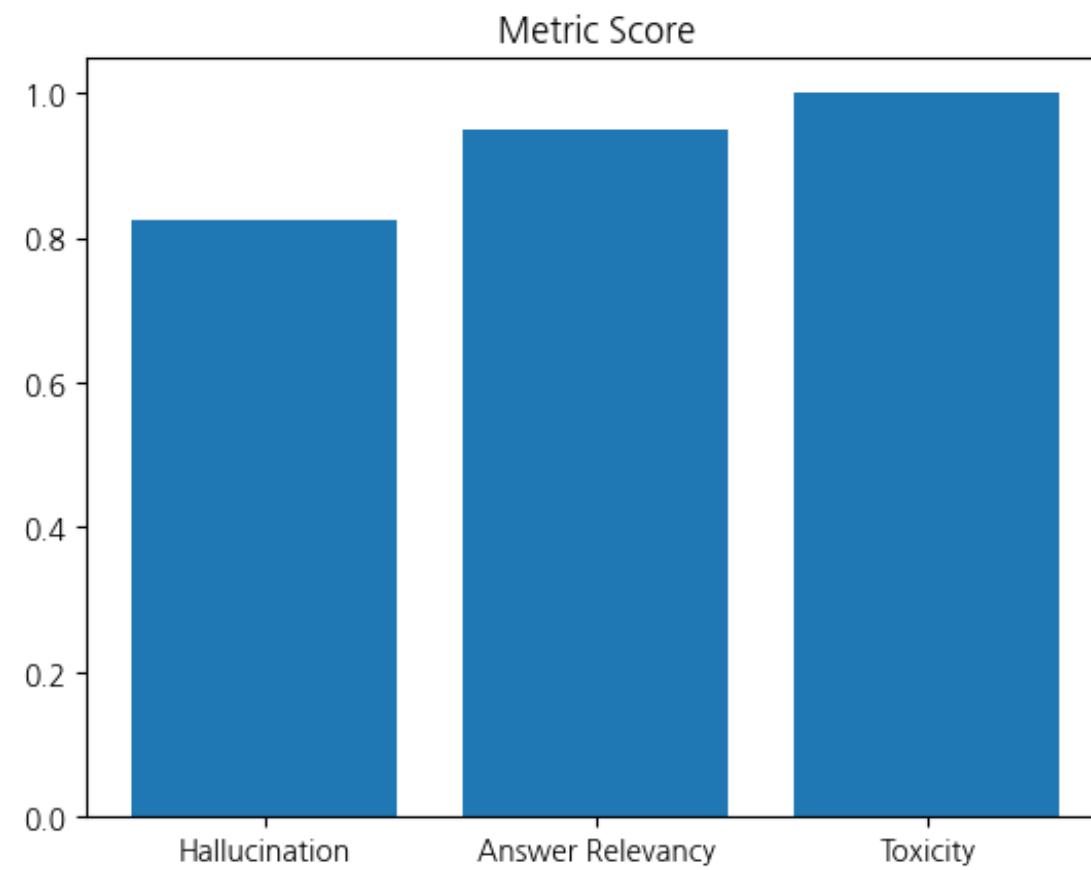


Chart 1. Metric Score

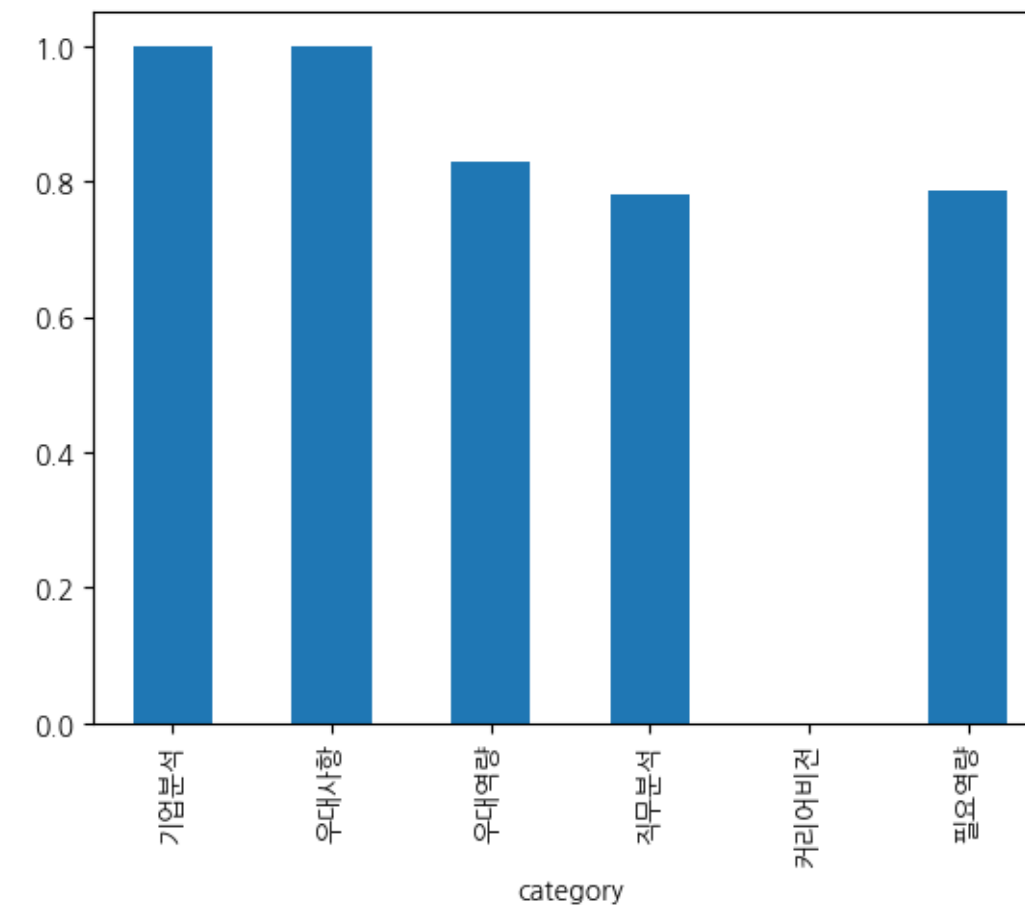


Chart 2. 질문 분류 카테고리 별 정확도

02. QA 데이터셋 구축

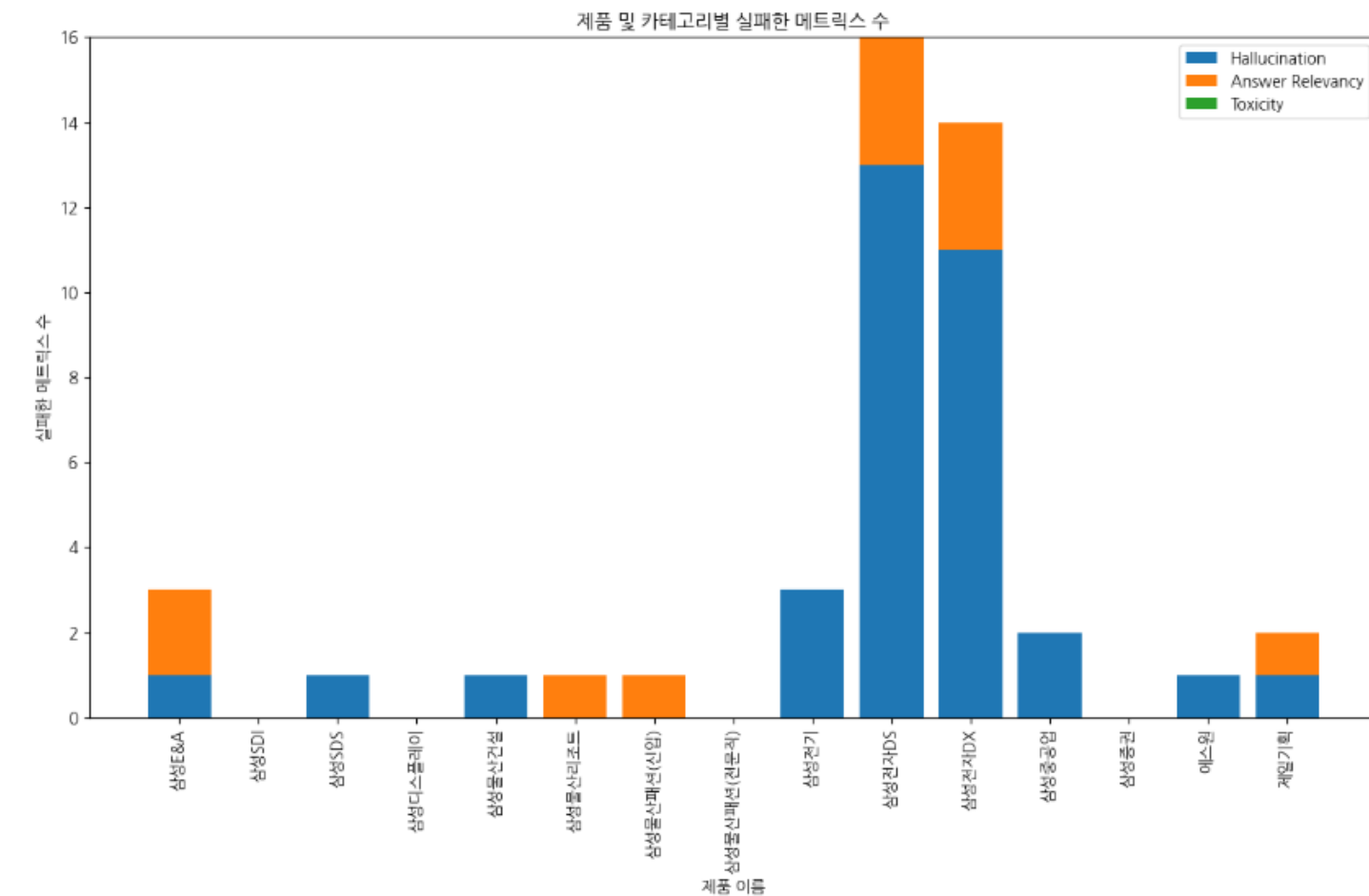
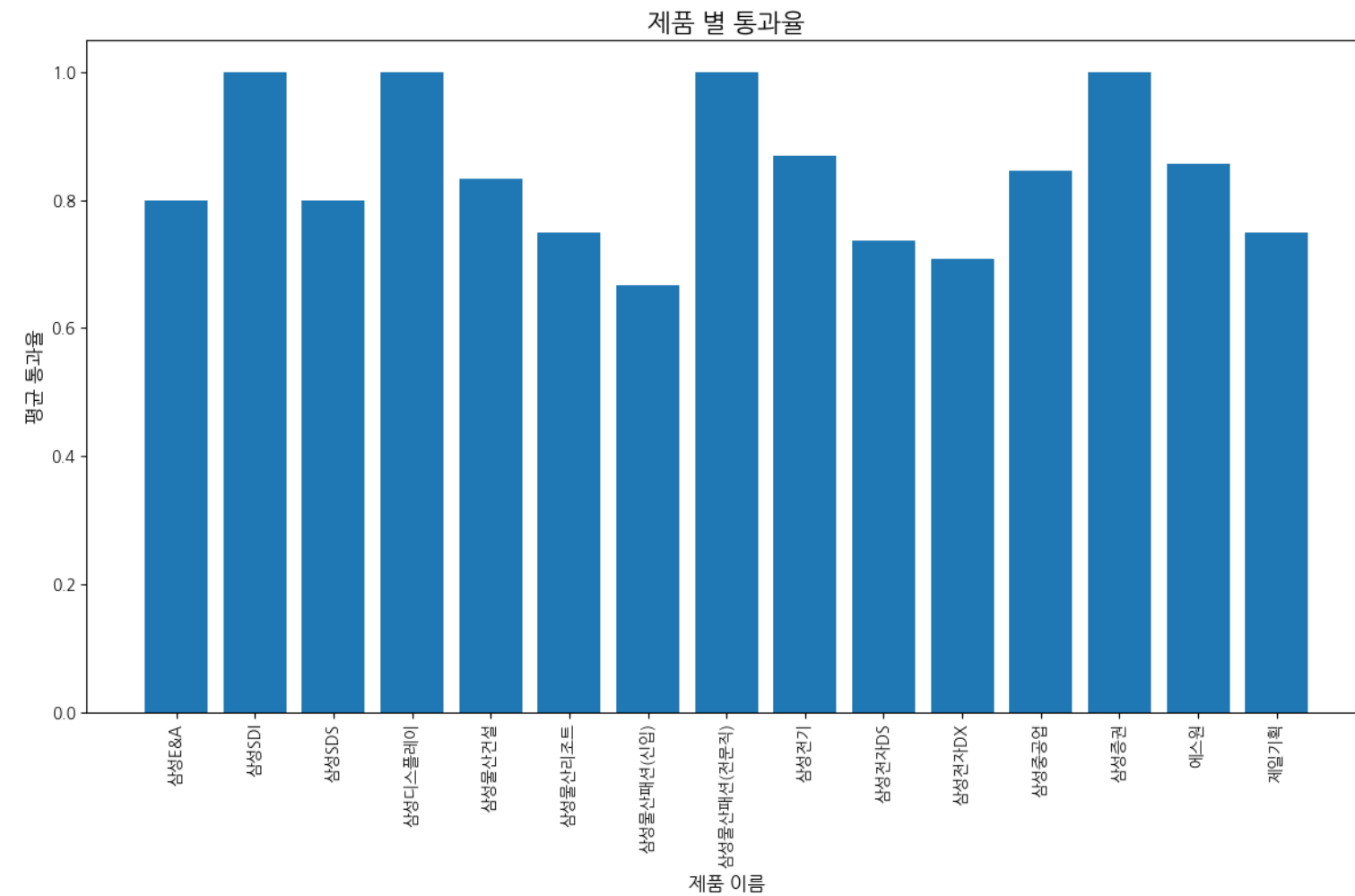


Chart 3. 기업별 실패 발생 빈도, 기업 별 통과율 그래프

02. QA 데이터셋 구축

평가 결과 요약

질문 카테고리

프롬프팅 과정에서 발생한 edge 카테고리인
커리어 비전을 제외하고, 모든 카테고리에서 80%
~ 100%의 정확도를 보여줌

기업별 분류

직무기술서, 채용공고 페이지에 따라 품질이
달라지는 경향.
60%~100%의 큰 정확도 편차가 일어남.

데이터 오류 별 분류

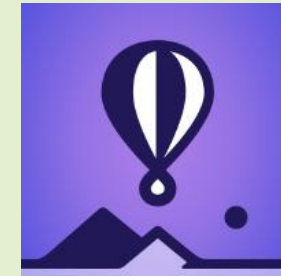
Toxicity : 오류 발생 건수 없음.
Answer Relevancy : 95% 이상 정확도 기록
Hallucination : 가장 많은 오류 발생

03. 서비스 구축 및 배포



FRONTEND

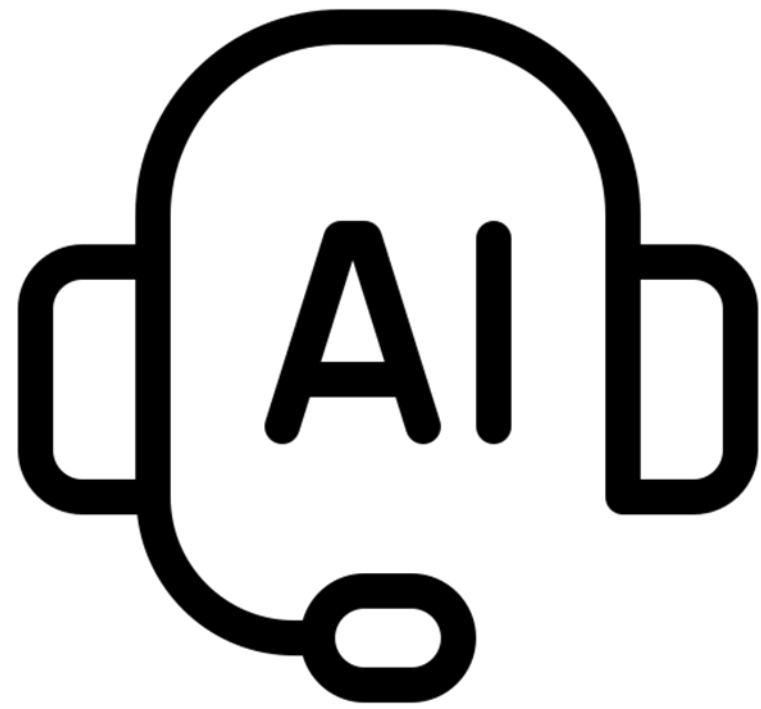
1. 기본 제공 CHATBOT UI 템플릿을 활용
2. VERCEL을 통한 배포



BACKEND

1. 수집 데이터 + QA 데이터셋을 벡터 형식으로 임베딩 (Upstage beddings)
2. Pinecone vector store를 사용하여
벡터 저장 및 RAG 구현
3. FLY.IO를 통한 백엔드 배포

04. 프로젝트 개선 사항



개선 과제

1. 기업 별 정확도의 편차 발생 문제 (QA 데이터셋 구축 과정)
2. 기본 챗봇 인터페이스의 확장성 문제 (서비스 구축 과정)

향후 확장 과제

1. 데이터 추가 수집 및 정제를 통한 기업 정보 확장
2. 사용자의 이용 상황에 알맞은 메인 페이지 구성

영상 시연

감사합니다