

자연어 처리(NLP)를 이용한 인천대 챗봇

인챗

202002961 김현주 (팀장)

202001529 유다현 (팀원)

목차

① 제작 목적

② 작품 소개

③ 구현 방법 & 성능 평가

④ 작품 시연

⑤ 기대 효과

① 제작 목적

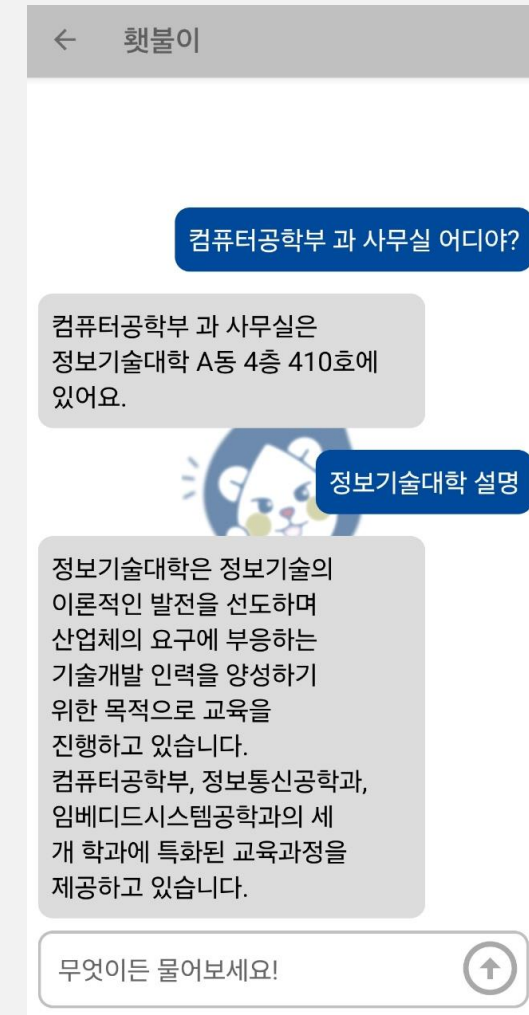
학교 생활 중 불편했던 문제점

- 인천대에 관한 정보를 찾는 동안
불필요한 웹서핑 시간 소요
- 유학생들 또는 인천대에 관해 모르는 사람들은
홈페이지를 통해 정보 찾는데 어려움
- 특정시간 이후 문의 전화 불가능
- 최근 전화 기피하는 사람 증가

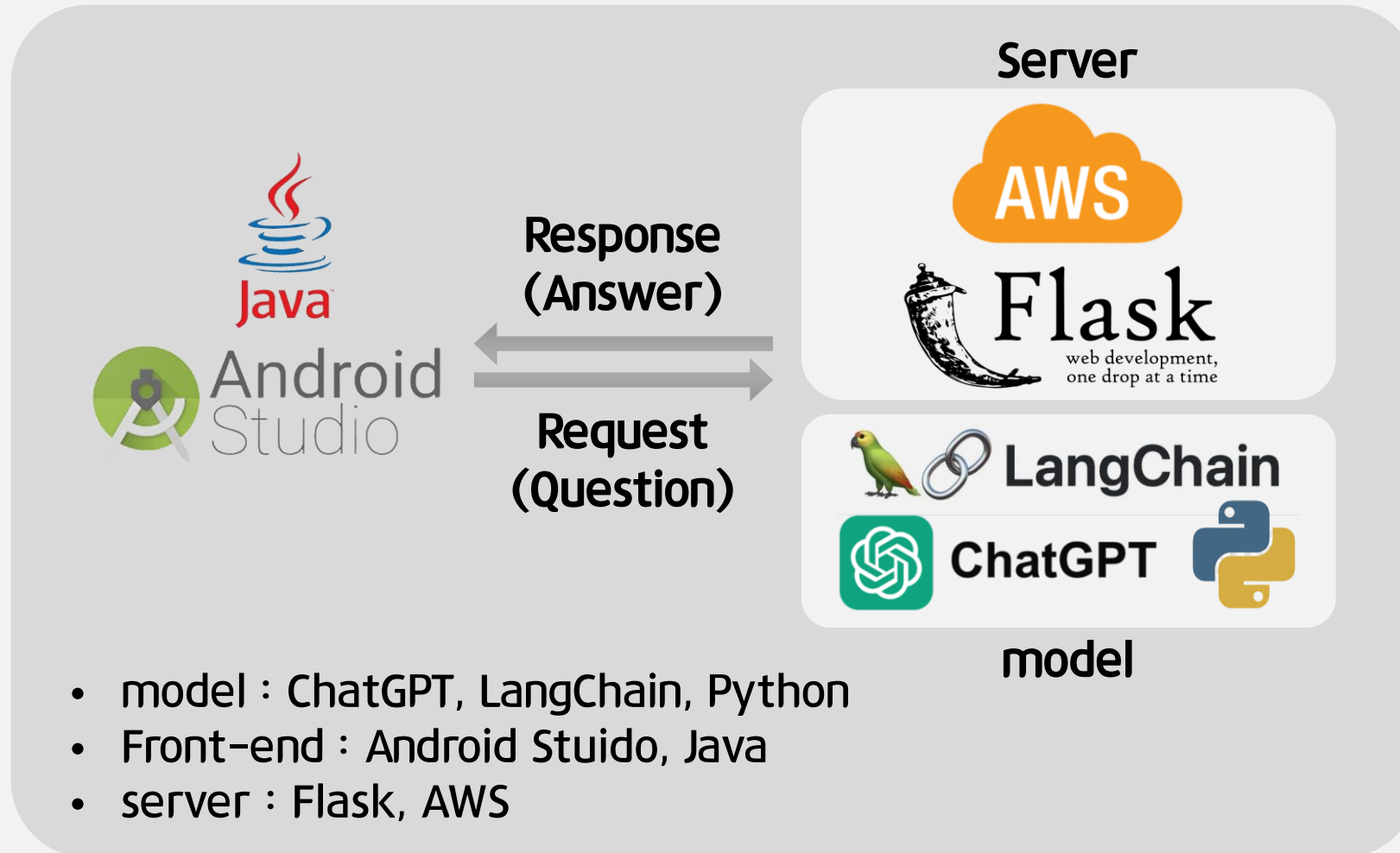
② 작품 소개

GPT 모델 기반의 챗봇 어플

- 한국어와 영어를 사용하는
인천대 재학생과 유학생을 위한 챗봇
- 정해진 형식의 질문에 답변만 주는 것이 아닌
학교 홈페이지 pdf를 검색해 기계를 통해
답변을 주는 챗봇



③ 구현 방법 & 성능평가

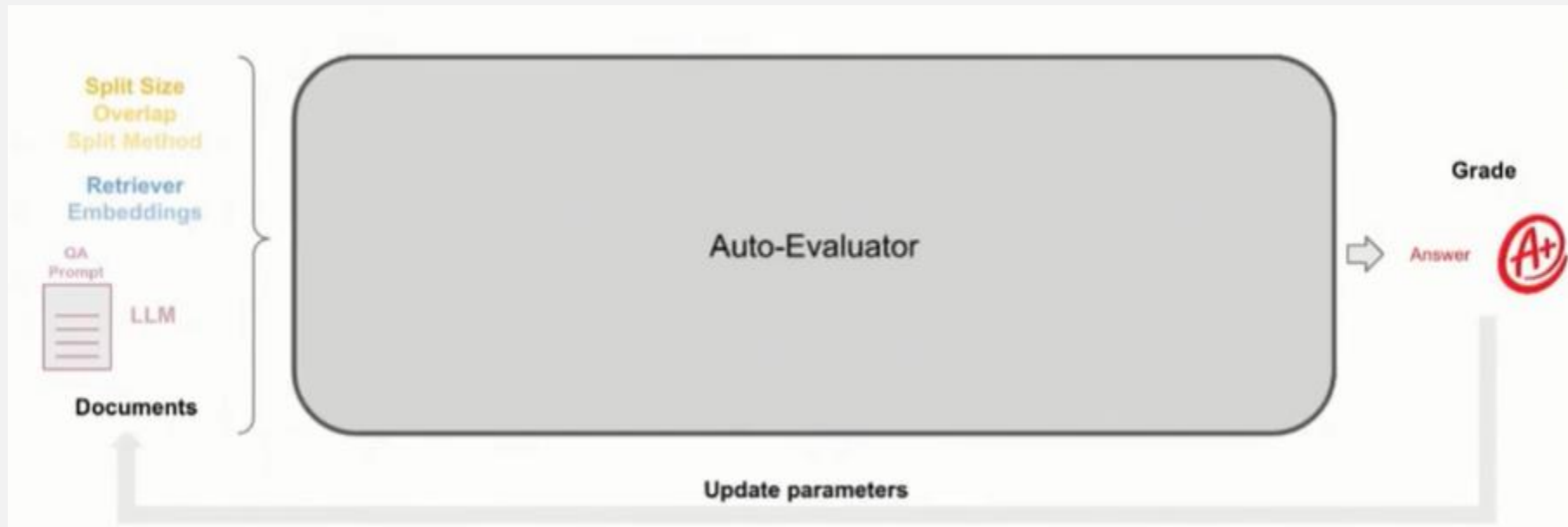


③ 구현 방법 & 성능평가



LangChain Auto-Evaluator

문서들로부터 LLM(GPT-3.5-turbo)을 사용하여
Question-Answer 세트를 생성하여 자동으로 평가

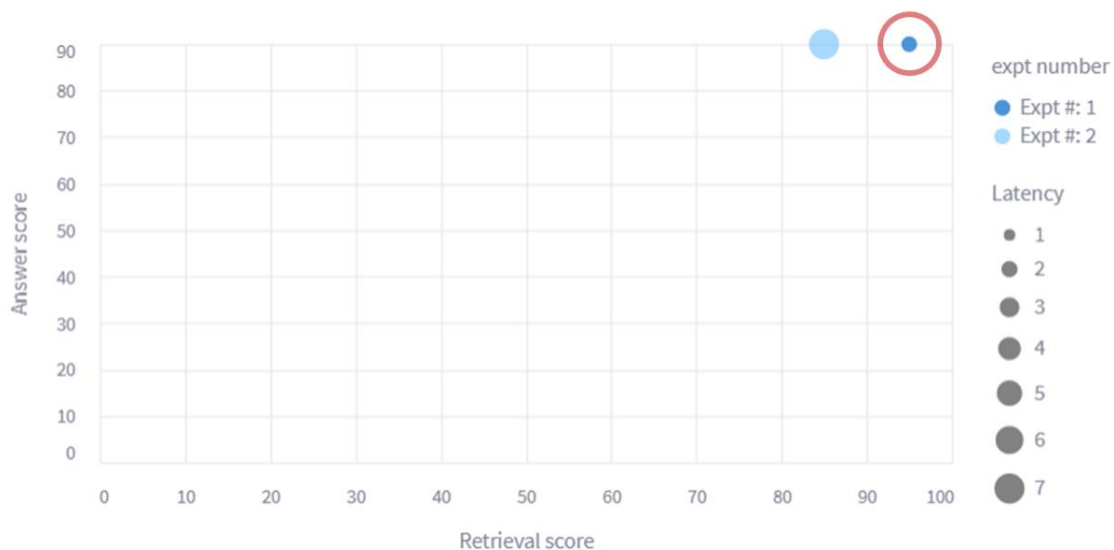


③ 구현 방법 & 성능평가

GPT 버전 비교

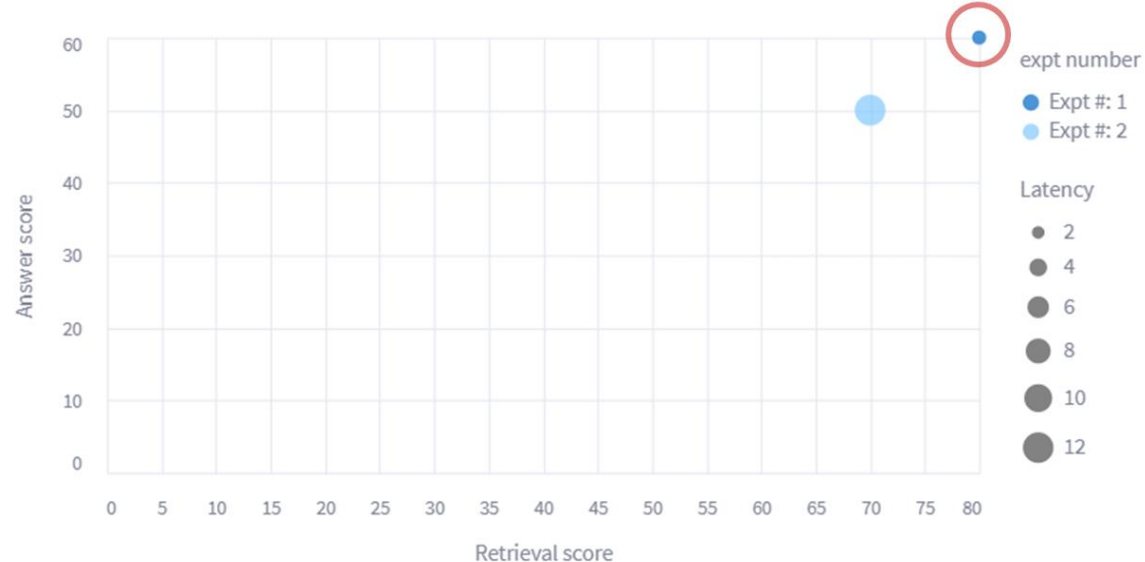
1) 영어

					응답속도	성능		
		model	retriever	embedding	num_neighbors	Latency	Retrieval score	Answer score
0	itter	gpt-3.5-turbo	TF-IDF	OpenAI	3	1.873	95	90
1	itter	gpt-4	TF-IDF	OpenAI	3	7.0056	85	90



2) 한국어

					응답속도	성능		
		model	retriever	embedding	num_neighbors	Latency	Retrieval score	Answer score
0	er	gpt-3.5-turbo	similarity-search	OpenAI	3	2.5727	80	
1	er	gpt-4	similarity-search	OpenAI	3	12.272	70	



③ 구현 방법 & 성능평가

Retriever 비교

1) 영어

응답속도 성능

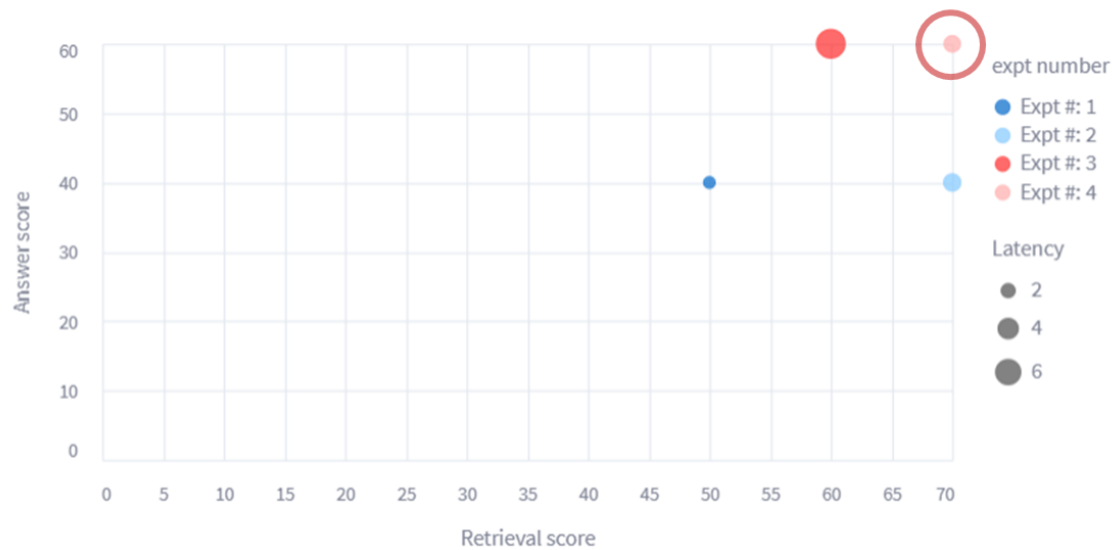
		model	retriever	embedding	num_neighbors	Latency	Retrieval score	Answer score
0	itter	gpt-3.5-turbo	TF-IDF	OpenAI	3	2.028	100	90
1	itter	gpt-3.5-turbo	SVM	OpenAI	3	2.4486	85	80
2	itter	gpt-3.5-turbo	Llama-Inc	OpenAI	3	5.0283	80	75
3	itter	gpt-3.5-turbo	similarity-	OpenAI	3	2.2808	80	85



2) 한국어

응답속도 성능

		model	retriever	embedding	num_neighbors	Latency	Retrieval score	Answer score
0	itter	gpt-3.5-turbo	TF-IDF	OpenAI	3	1.4581	50	40
1	itter	gpt-3.5-turbo	SVM	OpenAI	3	2.964	70	40
2	itter	gpt-3.5-turbo	Llama-Inc	OpenAI	3	7.7729	60	60
3	itter	gpt-3.5-turbo	similarity-	OpenAI	3	2.72	70	60



③ 구현 방법 & 성능평가

2) 한국어

한국어 데이터지만 **영어로 Question-Answer를 생성하여 평가**
→ 한국어 챗봇의 성능 저하

	question	answer
0	What are the main departments within the School of Business at Incheon University?	The main departments
1	What is the criteria for determining the scholarship amount based on the average GPA?	The scholarship amount
2	What is the purpose of the College of Business at Incheon University?	The purpose of the Col
3	What are the three majors offered in the Department of Northeast Asian International Business?	The three majors offered
4	What is the track name for students majoring in Energy Chemical Engineering and air conditioning?	Carbon Neutral Track
5	What percentage of the tuition fee is waived for students who achieve a TOPIK score of 70% or higher?	70%
6	What is the main purpose of the Department of Information and Communication Engineering?	The main purpose of th
7	When was the Graduate School of Culture established?	The Graduate School o
8	What is the process for applying for a split payment of tuition fees for the 2024 spring semester?	The application period
9	Where can students apply for an Internet certificate at Incheon University?	Students can apply for

③ 구현 방법 & 성능평가

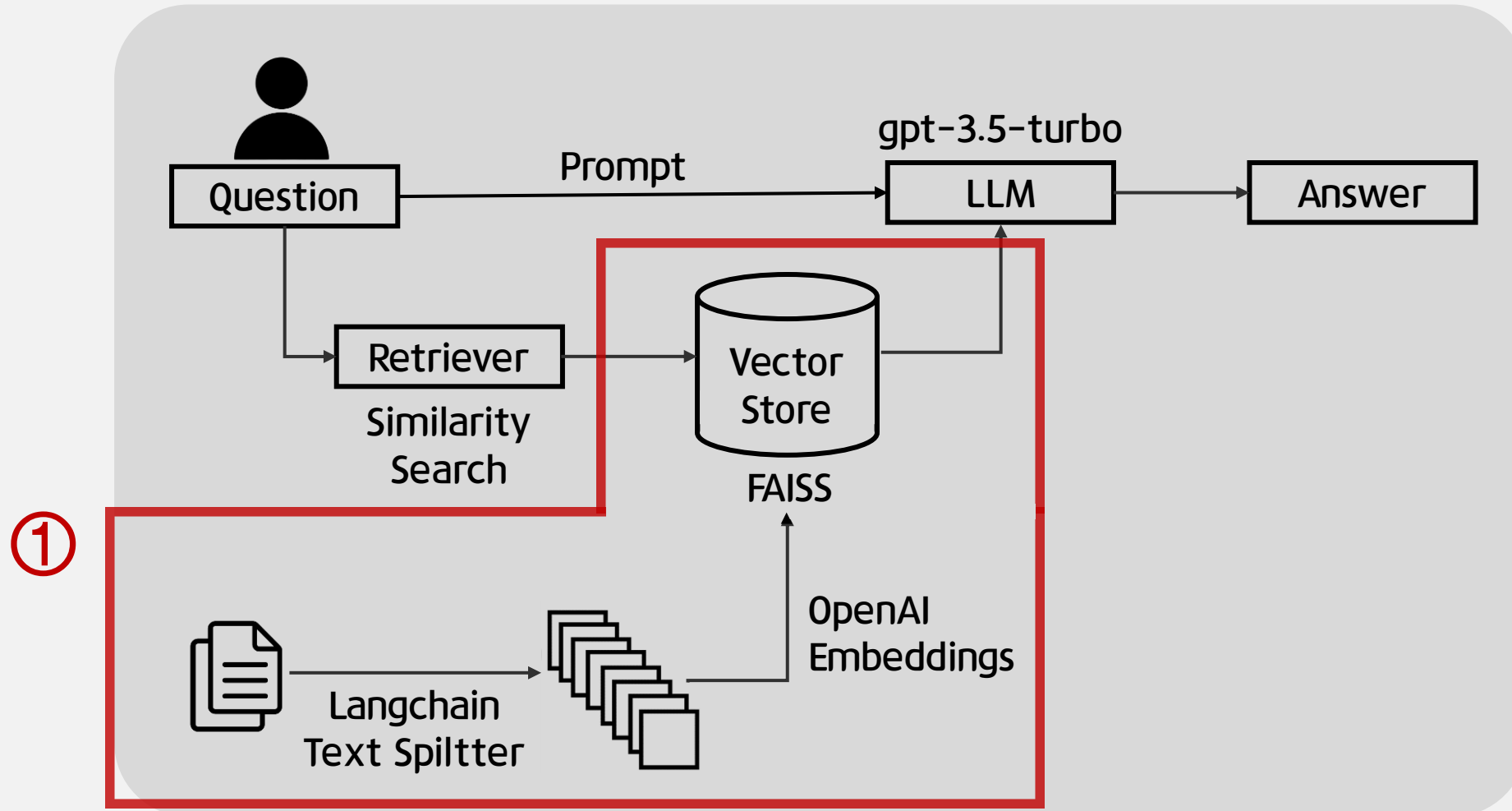
2) 한국어

직접 만든 Question-Answer

1	질문	Answer	Retrieval
2	2학기 기말고사가 언제야?	O	O
3	1학기 기말고사가 언제야?	O	O
4	1학기 중간고사가 언제야?	O	O
5	등록금에 대해 알려줘	O	O
6	기계공학과에 대해 알려줘	O	O
7	공결에 대해 알려줘	O	O
8	경조사로 인한 결석은 몇 일까지 인정돼?	O	O
9	부전공에 대해 설명해줘	X	X
10	복수전공 이수 중에 포기할 수 있어?	O	O
11	전과신청 자격이 있을까?	O	O
12	컴공 전화번호를 알려줘	X	X
13	컴퓨터공학부 전화번호를 알려줘	O	O
14	기계공학과 위치가 어디야?	O	O

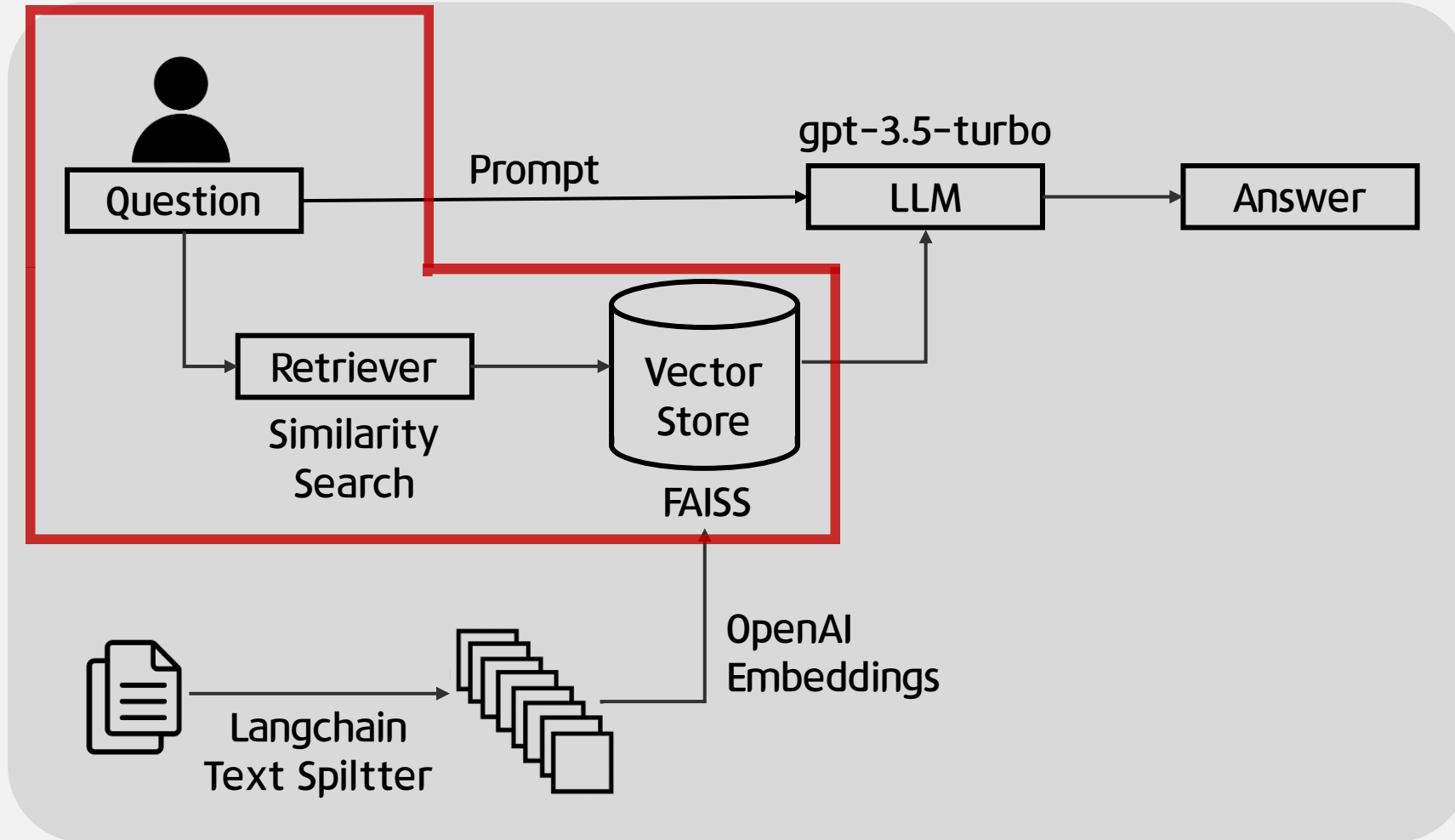
- **Retrieval Score : 85**
질문에 대해 검색된 **문서의 관련성** 평가
- **Answer Score : 83**
모델이 생성한 **답변의 품질** 평가

③ 구현 방법 & 성능평가



③ 구현 방법 & 성능평가

②



③ 구현 방법 & 성능평가



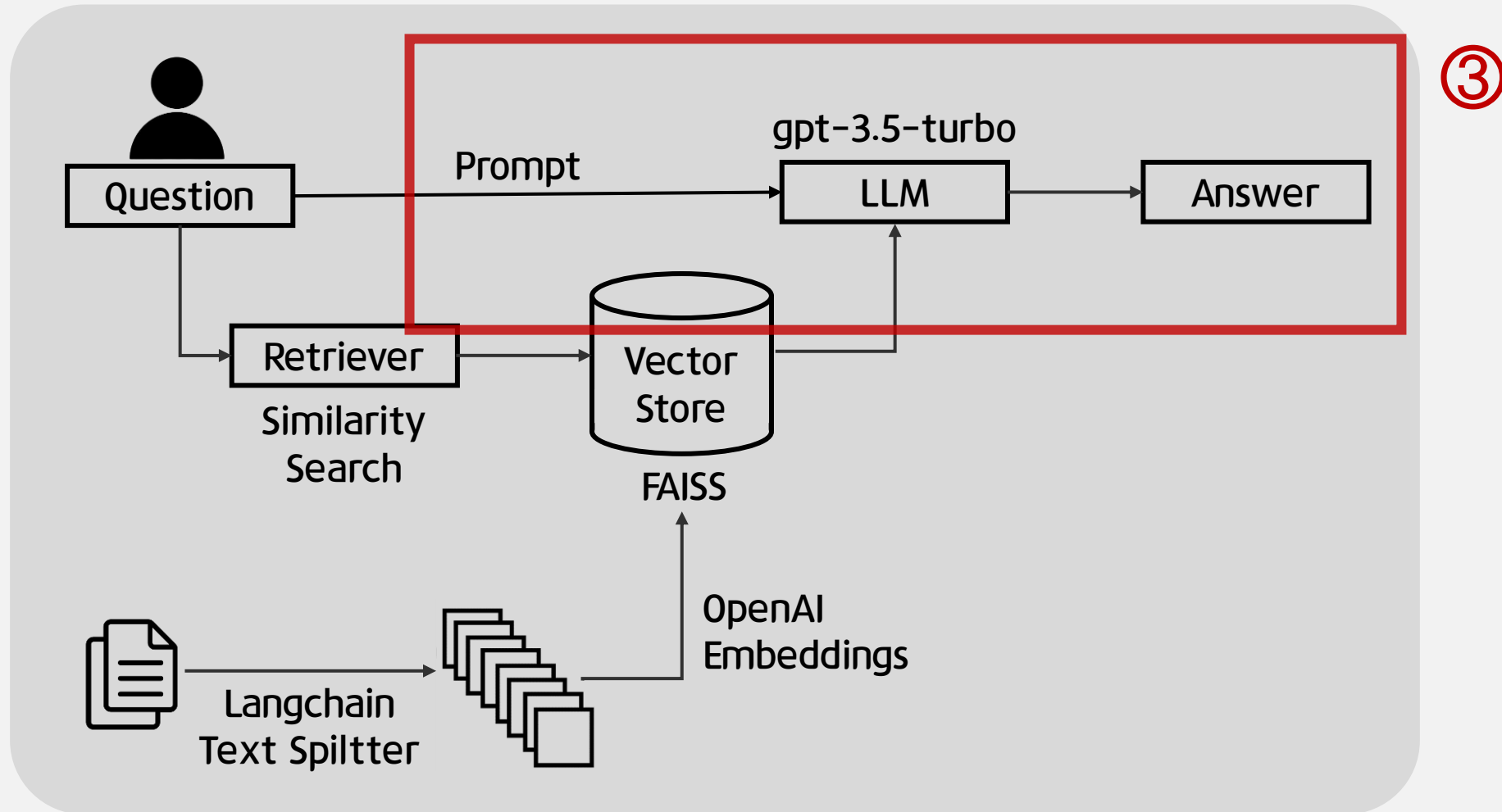
- **Prompt** : 모델의 응답을 안내하고 문맥을 이해하며 답변 출력을 생성하는 데 도움을 주는 역할

```
system_template = """
You are a knowledgeable and helpful assistant specializing in answering questions about Incheon National University (INU).
You have access to various documents about INU to provide accurate and detailed responses.
Use only the information from the provided documents to answer the questions.

Guidelines:
1. Accuracy: Ensure all responses are factually correct and directly supported by the provided documents.
2. Brevity and Clarity: Provide concise answers, but include enough detail to be helpful.
3. Language: Respond in the language of the question. Use polite and formal language when responding in Korean, and professional tone in English.
4. Unavailable Information: If the document does not contain enough information to answer a question, respond with "질문에 대한 답을 찾지 못했습니다".
5. Terminology: Understand and use the following equivalencies:
    - '기말고사' = '기말시험' (final exam)
    - '중간고사' = '중간시험' (midterm exam)
    - '과사' = '과 사무실' (department office)
    - '전번' = '전화번호' (phone number)
    - '정보대' = '정보기술대학' (College of Information Technology)
    - '컴공' = '컴퓨터공학부' (Department of Computer Engineering)
6. Contextual Adaptation: Adapt your response based on the context and details within the documents.

Documents:
{docs}
"""
```

③ 구현 방법 & 성능평가



③ 구현 방법 & 성능평가



④ 작품 시연



⑤ 기대 효과

인챗의 장점


- 24시간 운영
- 단순 UI를 이용하여 누구나 이용 가능, 편리
- 인천대 홈페이지를 직접 방문하지 않고 필요한 정보에 접근 가능
- 불필요한 웹서핑 없이 정보 검색 시간 절약 가능
- 인천대 관련 궁금증을 해결하고자 방문하는 사용자들에게 인천대의 홍보 효과 기대 가능
- PDF를 통해 문서를 검색하여 챗봇에 사용되는 데이터를 언제든지 간단하게 업데이트 가능

⑤ 기대 효과

다른 인천대 챗봇들과의 차별점



인혁이

컴퓨터공학부 졸업요건 알려줘
오전 11:54

**인혁이**
인천대학교는 10개의 단과대학에서 외국인 학생들을 모집합니다

- ◆ 신입생은 생명공학부를 제외한 모든 학과에서 모집
- ◆ 편입생은 생명공학부, 한국화전공, 서양화전공, 체육학부, 운동건강학부를 제외한 모든 학과에서 모집
- ◆ 소비자학과 신입학, 2학년 편입은 글로벌경영대학 소비자학과, 3학년 편입은 자연과학대학 소비자동학과로 신청해야 함

잘가
오전 11:56

**인혁이**
제가 잘 모르는 내용이에요 😞
다시한번 질문에 줄래요?
 다른 언어로 변경해 대화 할 수도 있어요!
If you want to change the language, click the button in the top left.

INU챗봇

학식 알려줘
오전 11:22
안녕

INU봇
이해하기 어려워요
오전 11:22
2023년 10월 15일 일요일 >

과사번호알려줘
오후 12:49

INU봇
총 학생회 SNS 링크입니다.
제휴업체를 확인하세요
INSTAGRAM
FACEBOOK
플러스친구
오후 12:49
안녕

INU봇
무엇을 원하시는지 잘 모르겠어요
오후 12:50
컴퓨터공학부 전화번호 알려줘
오후 12:50

⑤ 기대 효과

다른 인천대 챗봇들과의 차별점

인챗	인혁이	INU챗봇
문서검색 방식	시나리오 방식	시나리오 방식
어플	웹	카카오톡 플러스 친구
재학생과 유학생 정보	외국인 (신)입학생 정보	학식, 학교의 날씨 등 한정된 소량의 정보

발표 들어주셔서 감사합니다
