



연세대학교 빅데이터 학회

YBIGTA

포트폴리오 및 산학협력 제안 자료



YBIGTA는 **국내 최초의 대학생 빅데이터 학술 동아리**로, 연세대학교 학부 및 대학원생으로 구성되어 있습니다. 본 동아리는 빅데이터의 기초 지식을 습득하고, 최신 트렌드에 대한 이해를 넓히며, 구현 기술과 툴 사용법을 학습하여 이를 기반으로 다양한 학술 활동을 진행하는 것을 목표로 하고 있습니다. YBIGTA는 Data analytics, Data engineering, Data science 팀으로 구성되어 있으며, 빅데이터 전 영역에서 활발히 활동 중입니다.

YBIGTA는 꾸준히 여러 기업과 산학협력 및 연구용역 프로젝트를 수행해왔습니다.

- 2025년 상반기: (주)디투엘과 연구용역 계약 체결 / Upstage AI use case demo 제작
 - 제품 데이터 DB 기반 RAG(Text2SQL) 프로젝트 (디투엘)
 - 클레임 DB 기반 AI Agent 개선 대책서 자동 작성 프로젝트 (디투엘)
 - 아동 상담 기록 유형 분석 및 메일 자동화 Use case 데모 (Upstage AI)
- 이외에도 CHLNGERS, Carat, 렛서, 나노스정보, 글루코핏, 크라우드웍스 등과 다년간 협력 프로젝트 수행

YBIGTA는 단순한 대학 내 프로젝트를 넘어, 최신 빅데이터 기술을 활용한 산업 현장 맞춤형 데모 및 프로그램 제작, 연구용역 등 다양한 산학협력에 적극적으로 임하고 있습니다. 특히 업무 자동화, 모델링 등 여러 분야에서의 협업을 희망하며, 귀사에서 관련 프로젝트에 관심이 있으시다면 yonseibigdata@gmail.com 으로 연락 주시면 감사하겠습니다.

지도교수 소개

연세대학교 컴퓨터과학과 이원석 교수님

연구 분야

빅데이터 분석, 스트림 자율 학습, 재현데이터 (익명데이터) 처리, 영상 인식, 클라우드-엣지

주요 경력

1993-현재	연세대학교 컴퓨터과학과 교수
2012-현재	한국 빅데이터 포럼 운영위원장
2018	연세대학교 기획처 빅데이터 기반 교육혁신 TF 위원장
2013-2018	한국 빅데이터 포럼 운영위원장
2011-2013	(전) 대통령소속 국가정보화 전략위원회 실무위원



YBIGTA 연혁 및 팀 구성

2012

- YBIGTA 창립

. . .

Data Analytics

- EDA부터 인사이트 도출, 통계적 분석, 인과 추론, 그리고 모델 기반의 의사 결정까지 실무 중심의 분석 프로세스 학습
- A/B Testing, 추천 시스템 등의 통계 기반 분석 기법 등을 실무에 적용

Data Engineering

- Docker, Airflow, AWS, Kafka 등의 실무 도구를 익히고 목적에 맞는 파이프라인을 설계·구현
- MLOps 및 CI/CD의 개념과 필요성을 이해하고 이를 팀 프로젝트에 적용
- AI Agent 파이프라인을 설계하여 실제 DB를 바탕으로한 파이프라인 구축

Data Science

- CV, NLP, RL 등의 딥러닝 패러다임을 연구하며, TensorFlow, PyTorch, Keras 등의 딥러닝 프레임워크를 다룸
- 데이터의 종류에 따라 적합한 인공지능 모델을 구현할 수 있는 데이터 사이언티스트를 양성

YBIGTA 수상내역

2016

- Naver D2 CAMPUS 우수동아리 지원대상 선정

• • •

2020

- 삼성카드 데이터분석 온라인 가맹점 방문 고객 예측 알고리즘 부문 **1위**
- 문화관광 빅데이터 분석대회

• • •

2021

- CJ 대한통운 미래기술 챌린지 **최우수상**
- 공모전 시각화 스토리텔링부문 **대상**
- UNI-DTHON 2021 데이터톤 **대상**
- 한국 부동산 데이터 시각화 경진대회

• • •

2022

- 제2회 소방안전 빅데이터 활용 및 아이디어 경진대회 **최우수상**
- 제10회 문화데이터 활용 경진대회 **최우수상**
- 빅콘테스트 데이터분석리그 이노베이션 분야 **최우수상**

• • •

2023

- KPMG IDEATHON - NLP 활용 기업 서비스 제안 **대상**
- 2023 빅콘테스트 데이터신기술 분야 장려상
- 2023 빅콘테스트 비정형데이터 분야 **우수상**

2024

- 한국정책학회, 연세대 주최 사회문제 해결을 위한 해커톤 장려상
- 강남구 공공데이터 활용/분석 아이디어 공모전 **최우수상**
- 2024 [삼정KPMG] 2024 KPMG 아이디어톤 2등
- 임업진흥원, 산림빅데이터 진흥원 주최 산림빅데이터 활용 해커톤 **대상**
- DACON 2024 Samsung AI Challenge : Black-box Optimization **최우수상**
- 2024 문화체육관광 데이터 활용대회 **대상**
- 2024 빅콘테스트 생성형 AI 분야 장려상
- 2024 빅콘테스트 데이터 분야 장려상

• • •

2025

- (LH) 지식산업센터 공실률 최소화를 위한 해결방안 및 분석결과 도출 장려상
- 2025 AGI Agent 해커톤 **우수상**
- 난독화된 한글 리뷰 복원 AI 경진대회 3등
- 성남시 공공데이터 분석 및 활용 아이디어 공모전 장려상
- 업스테이지 AI agent 해커톤 **최우수상**

YBIGTA 활동 소개 - Data Engineering 팀

커리큘럼

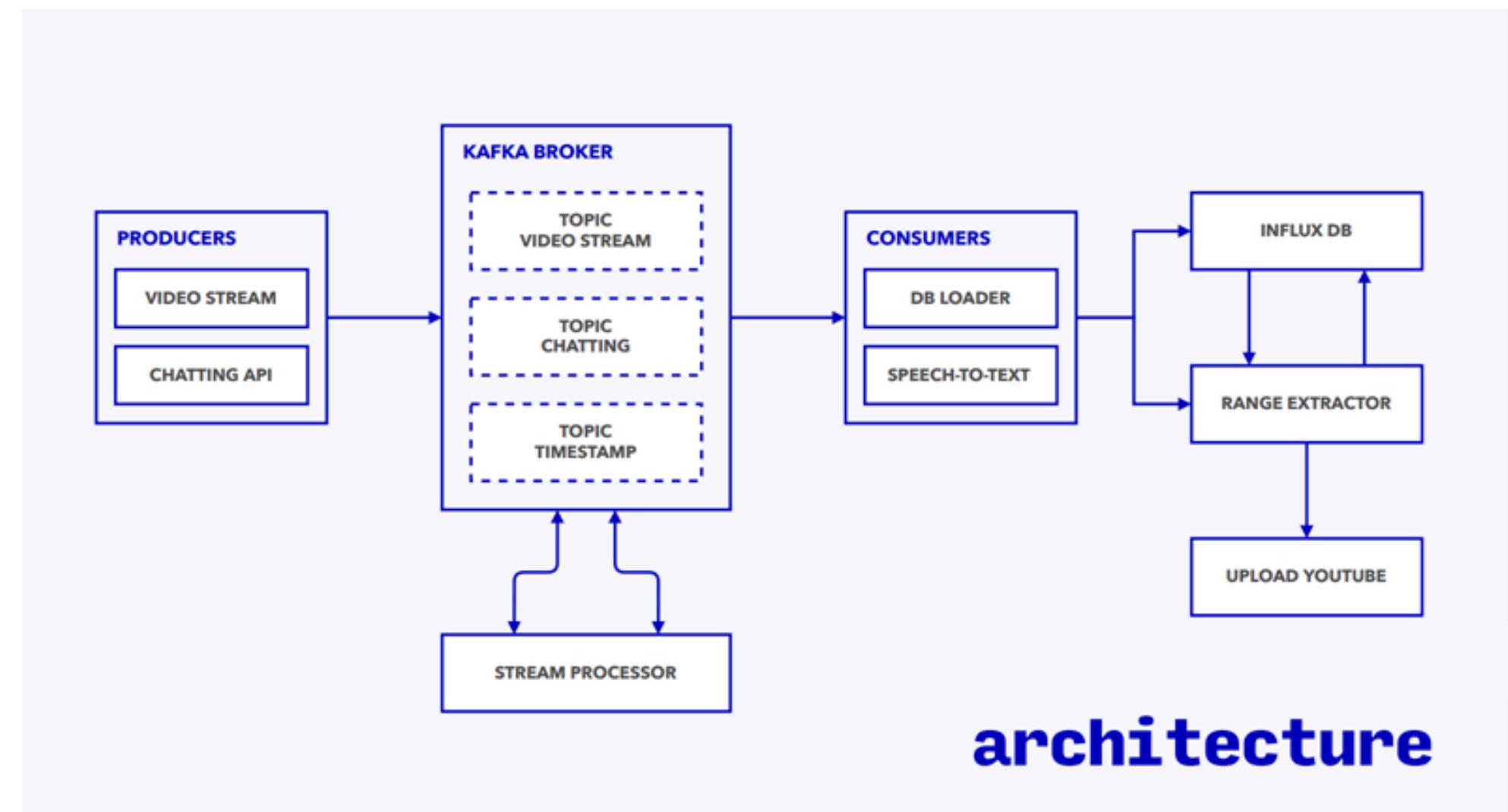
학기	01 Introduction (ETL, ELT)
	02 Docker, Docker Compose
	03 NoSQL, Airflow
	04 Cloud Computing, AWS
	05 Kafka, Kafka Streams
	06 Kafka Cluster, Kafka Connect
	07 MLOps

방학	01 Hadoop Cluster, MapReduce
	02 HDFS, Apache Parquet
	03 Spark 기초, RDD
	04 Spark SQL / DataFrame
	05 Spark Cluster, Spark Streaming
	06 Yarn

시니어 프로젝트: Highlighter



<https://github.com/YBIGTA/24th-de-highlighter>



YBIGTA 활동 소개 - Data Science 팀

CV 프로젝트 – Sports Police



<https://github.com/qhdamm/Sports-Police>

다이빙 종목에서 AI가 선수의 동작을 평가하여 채점 점수를 보여주는 서비스

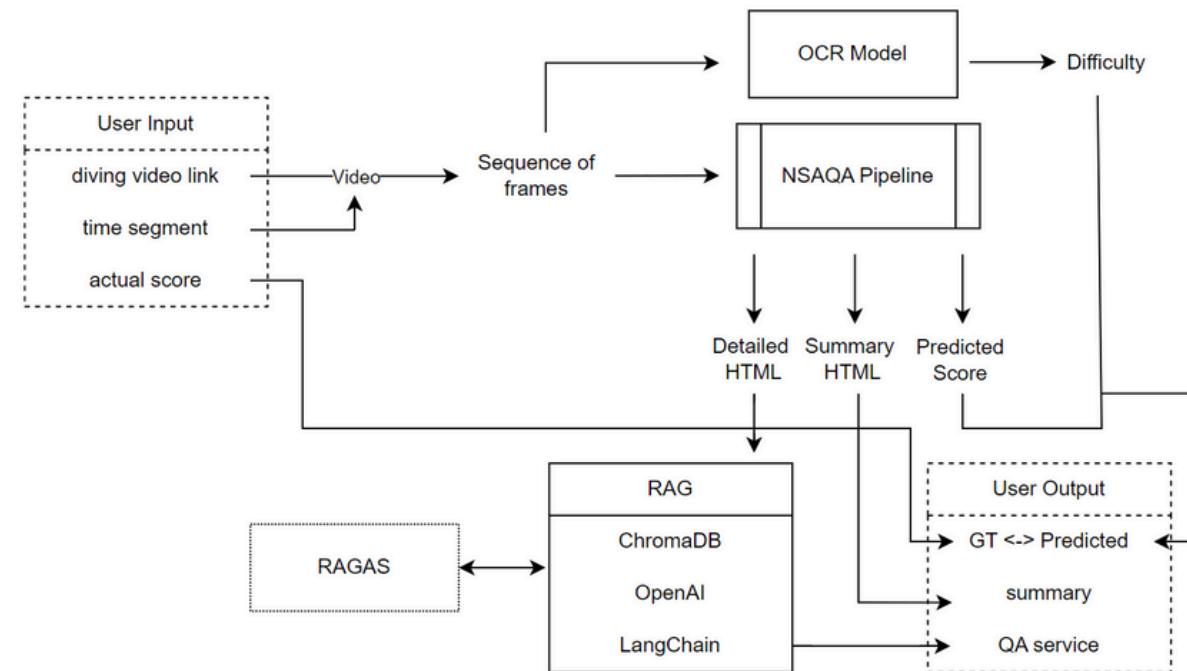
기능1. 다이빙 영상 분석 및 채점

- 다이빙 영상 분석을 통한 예측된 점수, 항목별 예측 점수 제공

기능2. QA 서비스

- 분석 결과에 대한 질의, 다이빙 규정에 대한 질의 가능

다이빙 종목에서 AI가 선수의 동작을 평가하고 그 결과를 보여주는 서비스

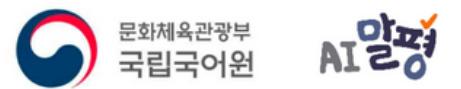


NLP 프로젝트 – 모두의 말썽치



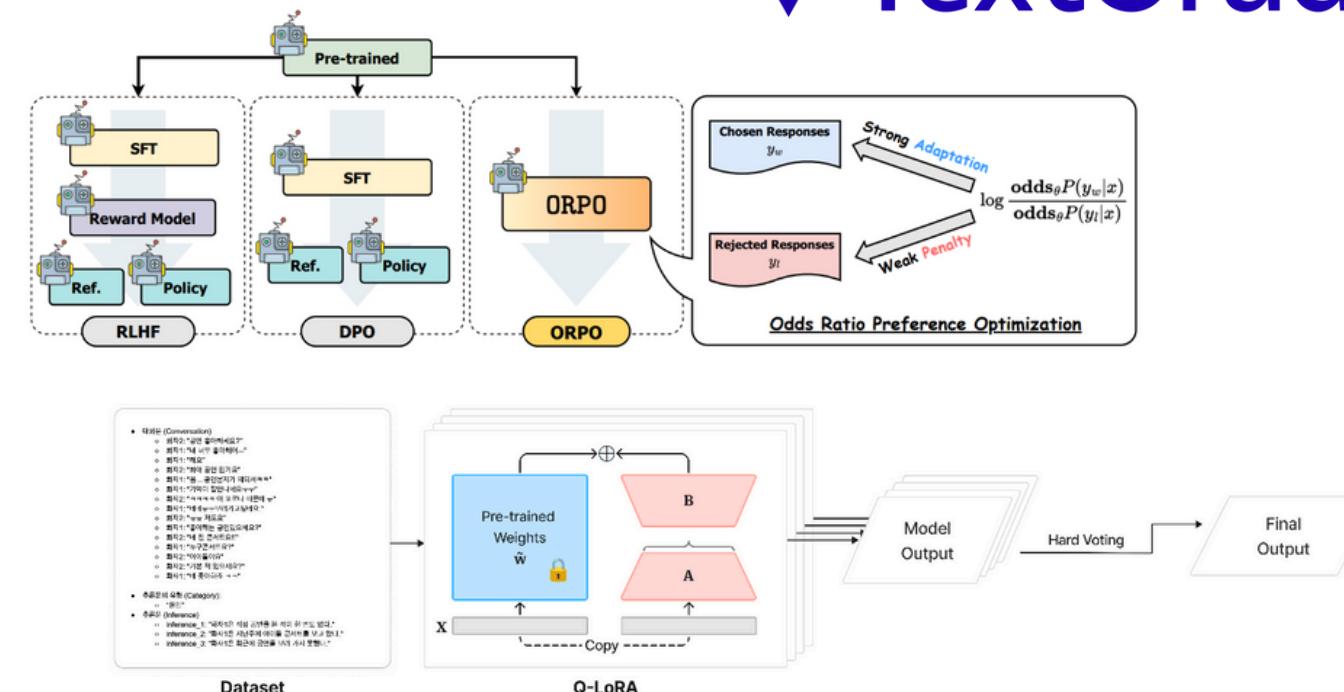
https://github.com/seodaegal/Ko_Conversational_Context_Inference

2024 국립국어원 인공지능의 한국어 능력 평가 목표



주어진 대화 내용을 바탕으로, 특정된 대상 발화로부터 다섯 가지의 추론문 유형인 ‘원인’, ‘후행 사건’, ‘전제 조건’, ‘내적 동기’, ‘감정 반응’ 중 하나에 상응하는 문장을 정확하게 추론

TextGrad



YBIGTA 활동 소개 - Data Analytics 팀

커리큘럼

학기	01	통계 기반 이상치 탐지
	02	ML 기반 이상치 탐지 기법
	03	DL 기반 이상치 탐지 기법
	04	시계열/시퀀셜 데이터 이상치 탐지
	05	컨텐츠 기반, 협업 기반, 하이브리드 필터링
	06	MLP 기반 추천 모델
	07	Sequential 추천 : RNN, Transformer 기반 모델
	08	GNN 기반 추천 모델
	09	Autoencoder 기반 모델, PoI 추천 알고리즘
방학	01	시니어 프로젝트 01
	02	시니어 프로젝트 02

시니어 프로젝트

1. All-In-One 특허 출원 보조 서비스
(KPMG 아이디어톤 2위)



2024 DA 겨울 시니어 프로젝트 1차 & 2차 최종 발표 ■

DA 팀 김지수 김체연 양진성 정지원
DE 팀 성현준 이우홍



Patent Assistant Productivity Accelerator

생성형 AI를 활용한 전문적인 특허 명세서 작성, 국내외 특허 유사도 분석, 연구동향 리서치 기능을 제공하는 All-In-One 특허 출원 보조 서비스 구현

- (2) 머신러닝 및 시공간 데이터 분석을 통한 강남구 음주운전 단속 솔루션
(강남구 공공데이터 활용 분석 아이디어 공모전 최우수상)

회귀 모델을 활용해 음주운전 사고다발지역, 유동 인구, 실시간 교통량 정보를 고려한 실시간 음주운전 단속
필요 지역 추천 서비스 구현

YBIGTA CONFERENCE

YBIGTA CONFERENCE는 매 학기 말 열리는 YBIGTA의 대표 행사이자 전 학회원이 모이는 최대 규모의 학술 행사이입니다. 한 학기 동안 연구·학습한 내용을 바탕으로 다양한 주제를 발제하고, 기존 팀과 다른 학회원들이 모여 새로운 프로젝트 팀을 구성합니다.

프로젝트 주제는 자연어처리(NLP), 컴퓨터비전(CV), 추천시스템, 시계열 분석, 멀티모달 AI 등 폭넓은 AI/데이터 분야를 다루며, 데이터 전처리부터 모델 설계·학습, 서비스 배포까지 End-to-End 완성형 결과물 제작을 목표로 합니다. 이를 통해 실무에 가까운 프로젝트 사이클을 경험할 수 있습니다.

또한, 기업·연구기관과의 산학협력을 통해 실제 문제 해결형 데이터 분석·엔지니어링·AI/ML 솔루션을 제안·구현하며, 실무에서 바로 적용 가능한 인사이트를 창출합니다. 행사 당일에는 각 팀이 프로젝트 결과를 발표하고 서비스·모델을 시연하며, 학회원과 외부 심사위원으로부터 피드백을 받아 프레젠테이션·협업·문제 해결 능력을 함께 향상시킵니다.

컨퍼런스 자료

<https://yonseibigdata.notion.site/YBIGTA-eb317c1ba96640a4bf591274c61c1181>

YBIGTA Conference 우수 사례 – RAG & AI Agent (DTOL - YBIGTA 25 상반기 산학 프로젝트)

업무 보조 AI 서비스 제작

RAG & AI Agent

프로젝트 목표

Problem

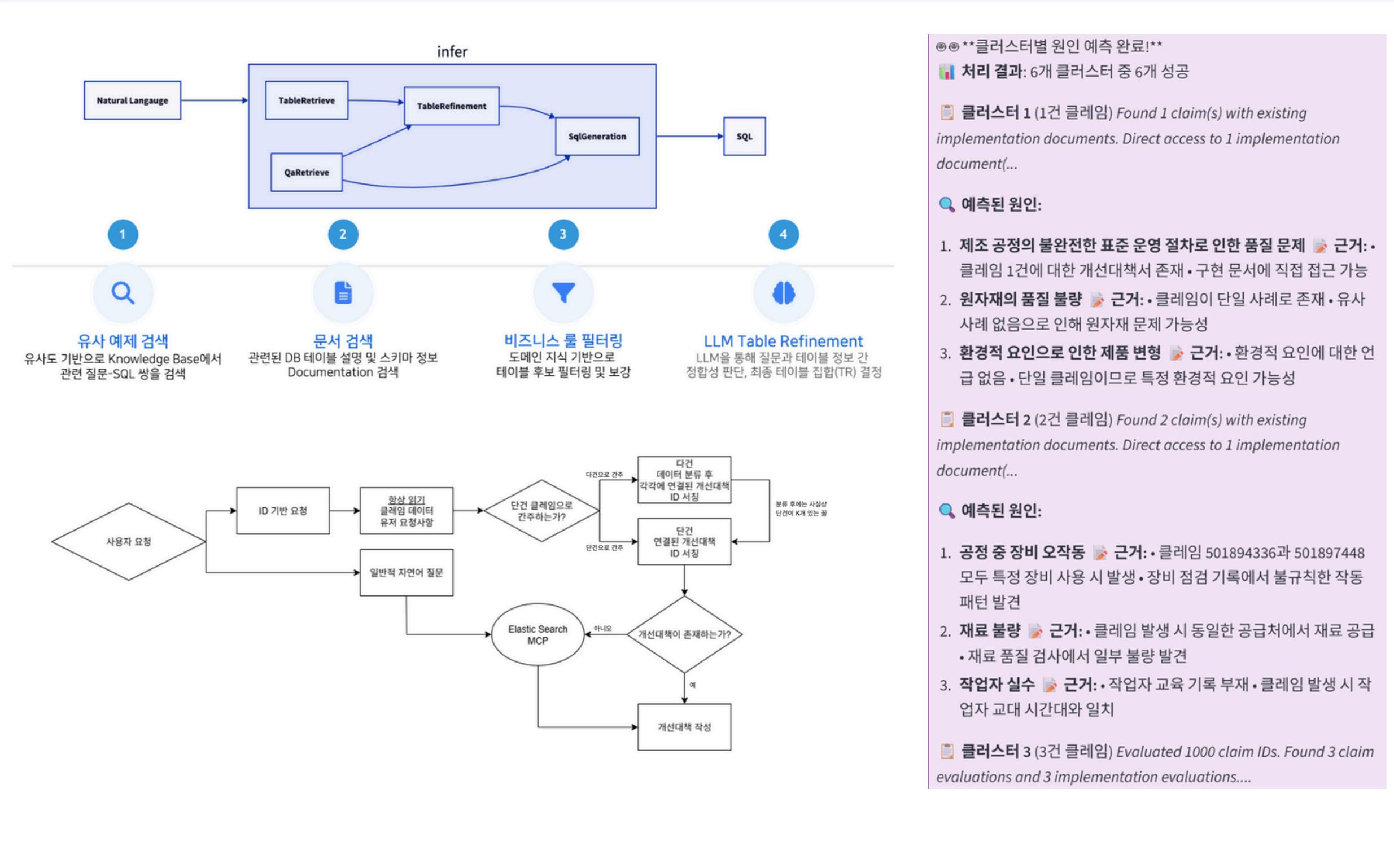
1. 방대한 데이터 탐색을 위한 DB 서칭용 RAG 필요
(Text2SQL)
 2. 반복적인 개선 대책서 및 클레임 원인 예측 업무 보조 Agent 필요

Solution

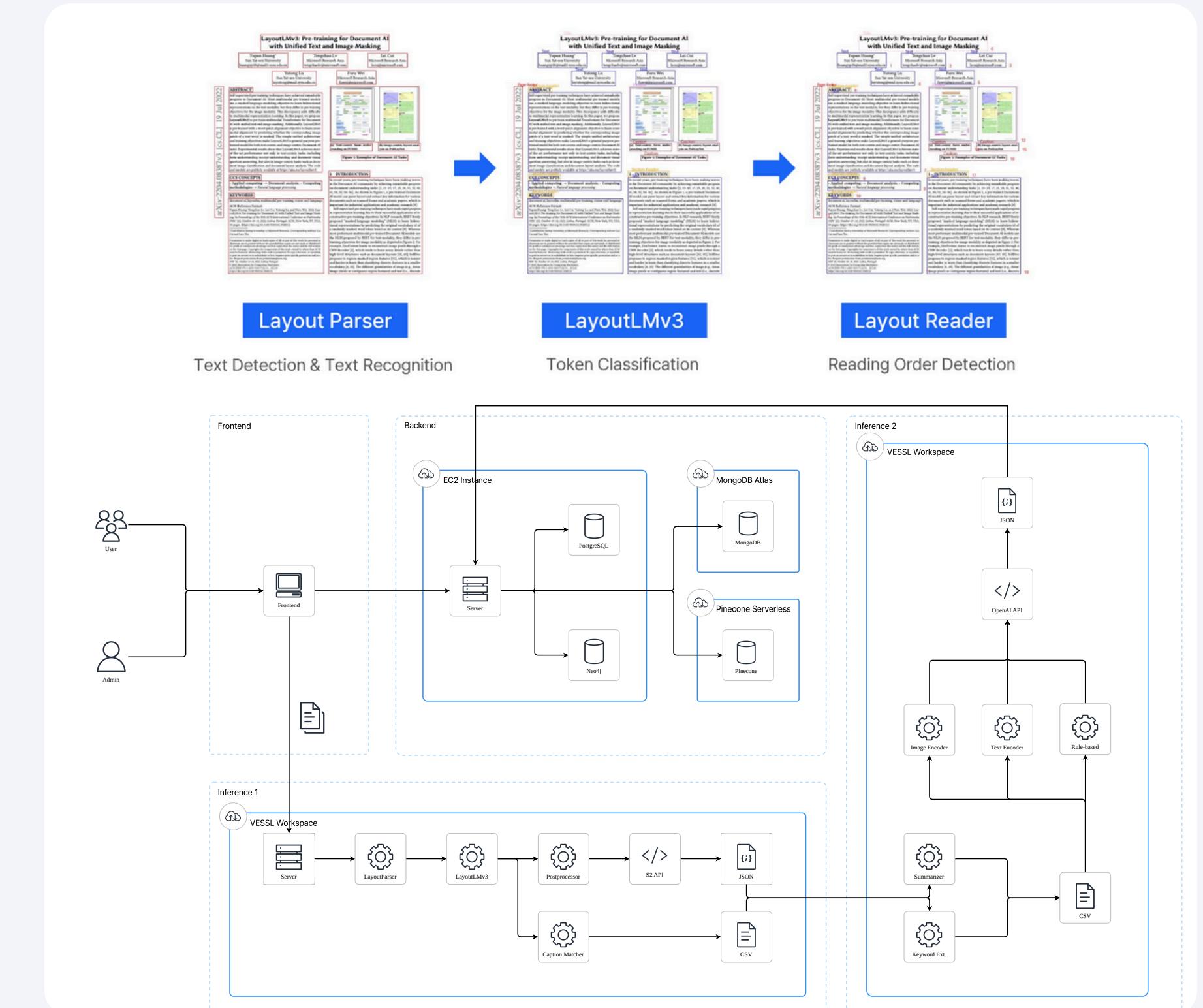
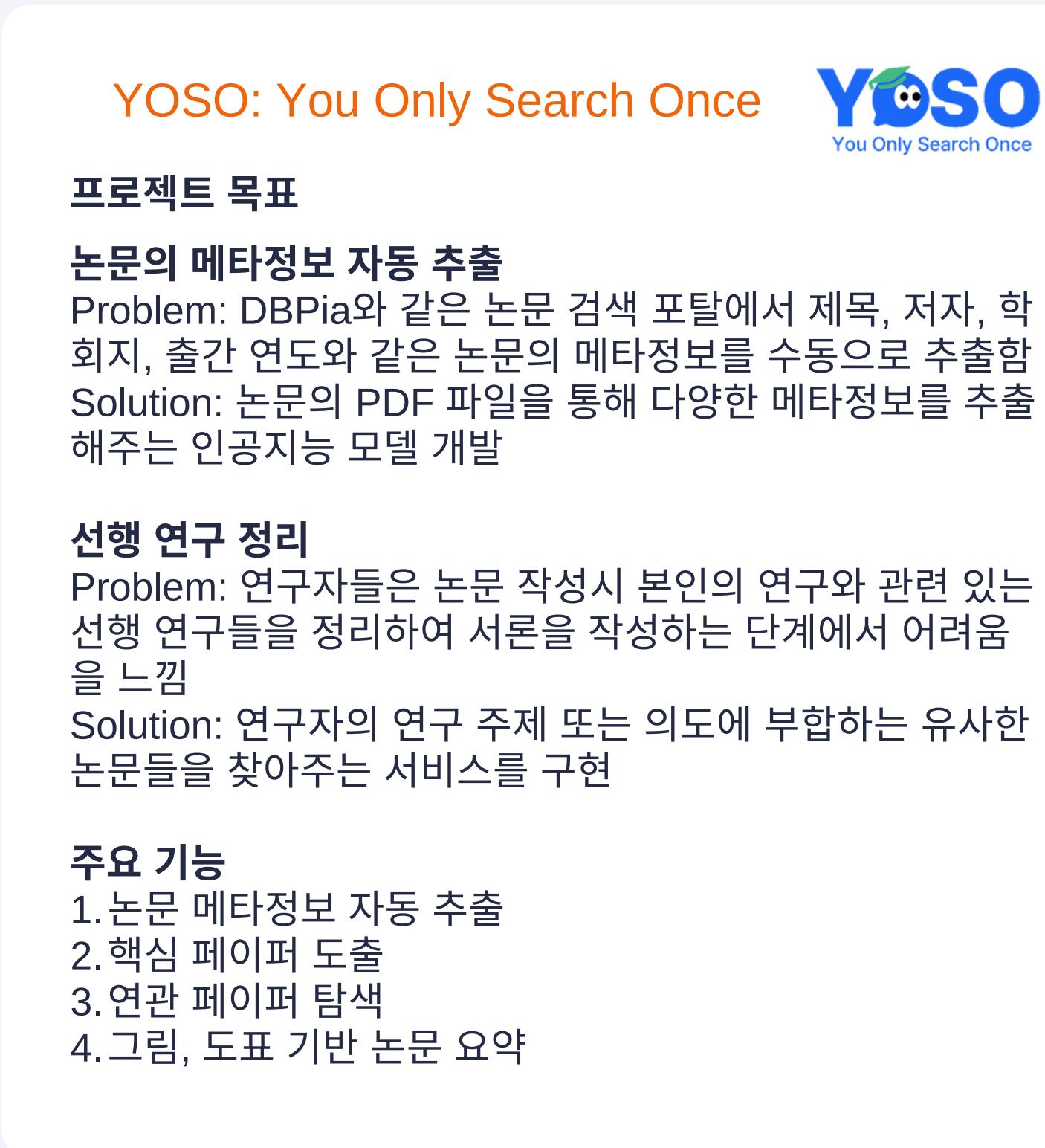
최신 LangChain LangGraph 같은 프레임워크를 활용해서 RAG, AI Agent 보조 서비스 제작

주요기능

1. Text2SQL
 2. DB 자동 서치 후 클레임 원인
예측 / 개선 대책서 작성



YBIGTA Conference 우수 사례 – YOSO: 선행 연구 기반의 논문 작성 보조기



YBIGTA Conference 우수 사례 – YouTube

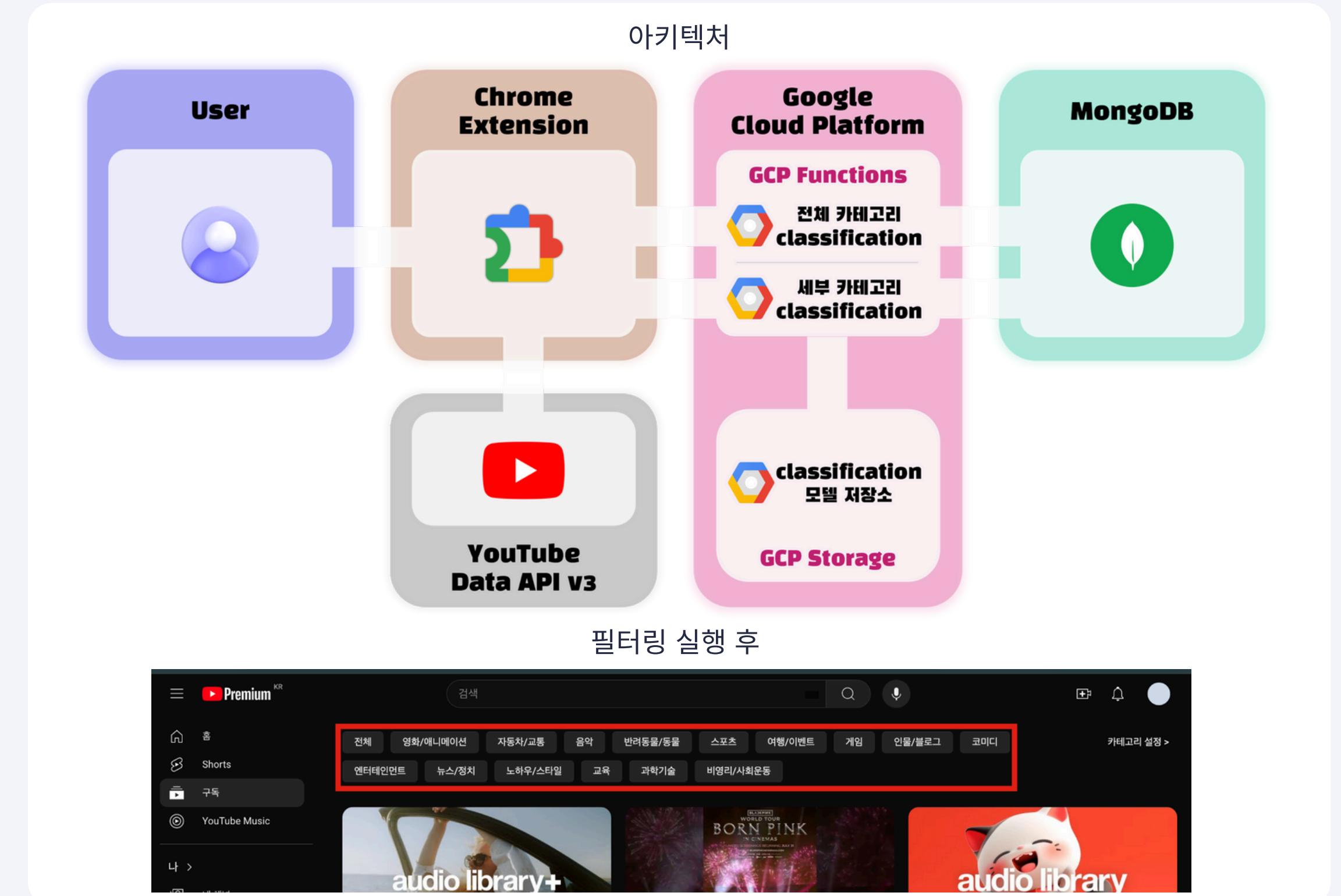


Personalized YouTube Algorithm Washer

프로젝트 목표

문제 인식: YouTube 사용자들은 추천 알고리즘으로 인해 자신의 원래 목적과 관련이 없는 컨텐츠를 소비하는 경우가 많다. 이를 해결하기 위해 사용자들은 자신이 관심 있는 채널을 구독하고 해당 채널의 영상을 시청하지만, 구독한 채널이 많고 다양한 경우에는 자신이 원하는 컨텐츠를 찾기 어렵다.

-> 해결책: 구독한 채널의 컨텐츠를 카테고리별로 분류하고, 사용자가 설정 가능한 세부 카테고리로 나누어 영상을 필터링할 수 있는 크롬 확장 프로그램



<https://github.com/YBIGTA/24th-conference-youtube>

YBIGTA Conference 우수 사례 - SIGMA

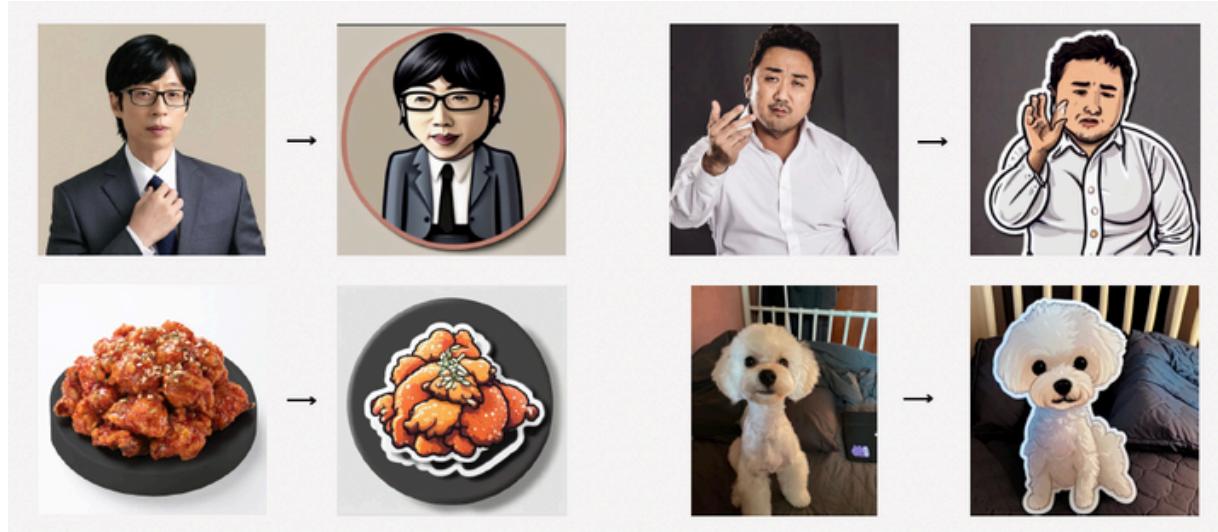
Sticker Image Generation by Multimodal Assistant

프로젝트 목표

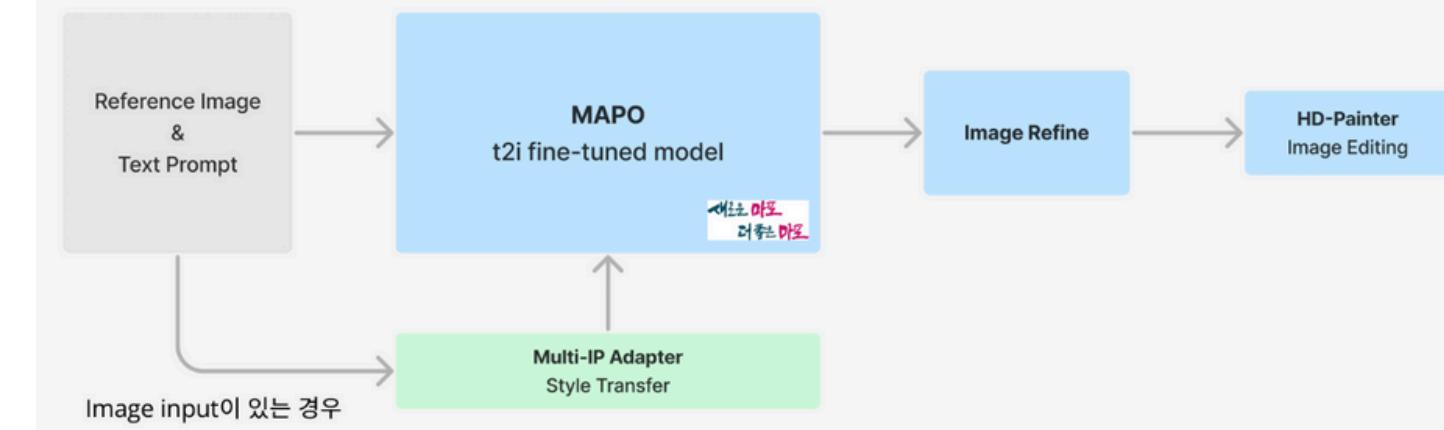
마케팅 에셋, 다이어리 꾸미기, 어린이 학습 교안 등을 위한 개인 맞춤형 스티커 생성 모델 개발

주요 서비스 및 파이프라인

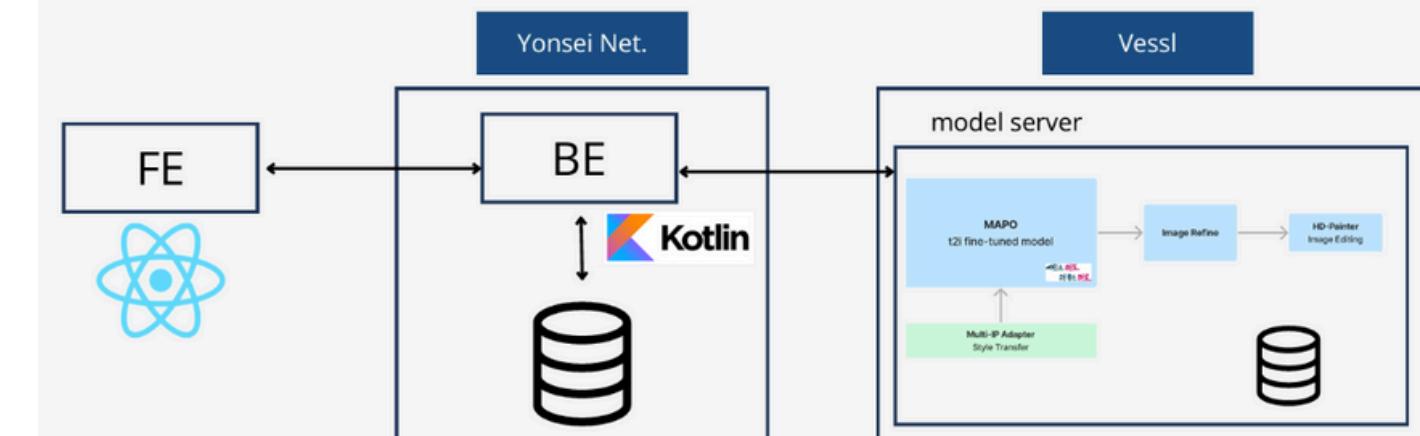
1. 유저가 원하는 텍스트를 입력했을 때 그에 맞는 스티커 생성
2. 스티커로 만들고 싶은 사진이 있을 때, 텍스트와 사진을 같이 입력했을 때 스티커 생성
3. 1차 생성이 끝난 후 수정하고 싶은 부분을 입력 받아 재생성



05 Overall Pipeline T2I & I2I Image Generation + Image Editing



05 Overall Pipeline service



YBIGTA Conference 우수 사례 – SONIG : Subject Consistent Image Generation

Subject Consistent Image Generation

프로젝트 목표

Diffusion Model의 fine-tuning 과정에서 모델에 발생하는 손실을 극복하기 위한 방법론 개발

프로젝트 설명

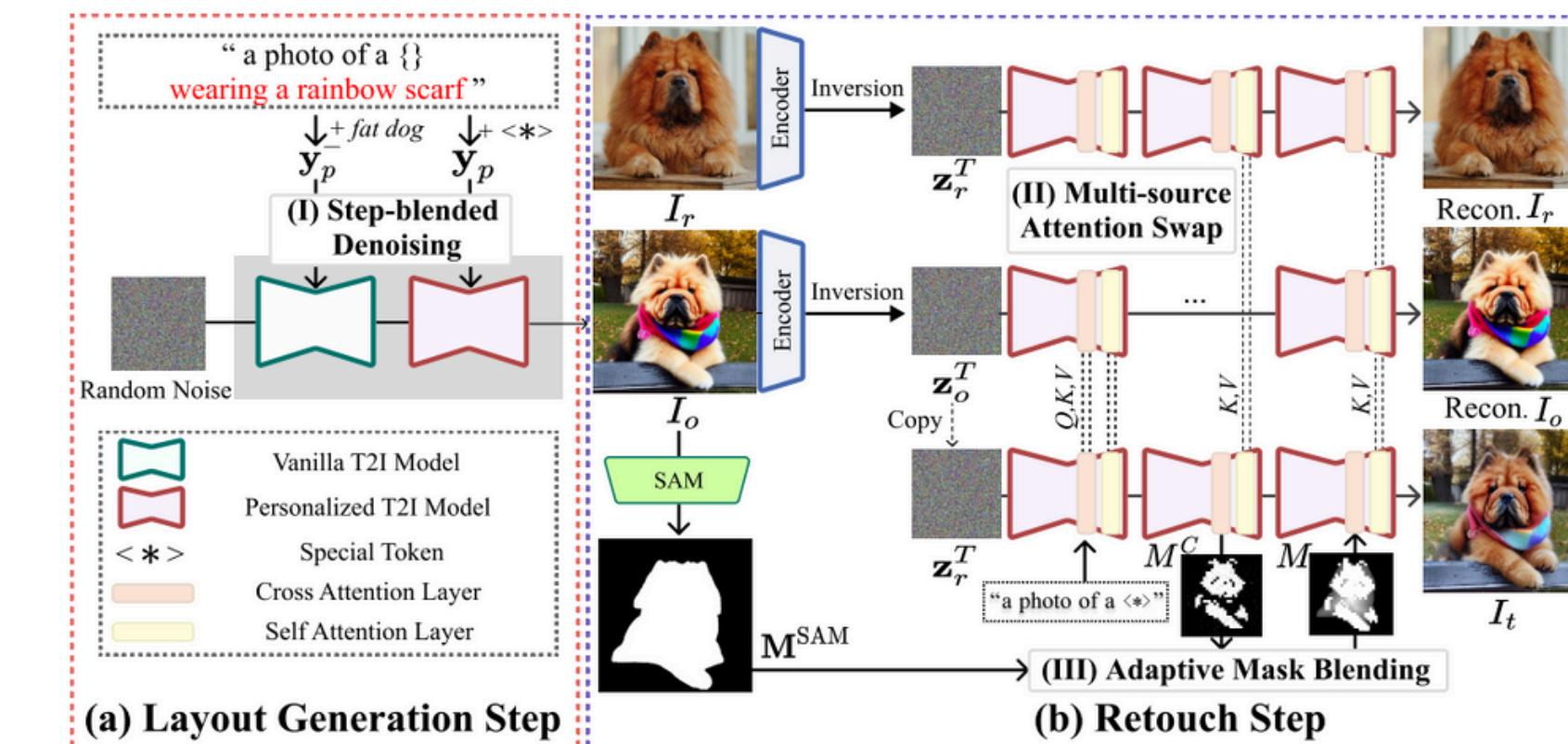
- Fine-tuning 이후에도 높은 프롬프트 수행 능력과 생성 이미지의 다양성을 보존할 수 있는 orthogonal한 방법론 제안
- 이후 다양한 fine-tuning 방법론에 적용하는 experiment를 진행함

Layout-and-Retouch: A Dual-stage Framework for Improving Diversity in Personalized Image Generation

Kangyeol Kim^{1,3*}, Wooseok Seo^{2*}, Sehyun Nam², Bodam Kim²,
Suhyeon Jeong², Wonwoo Cho^{1,3}, Jaegul Choo^{1,3†}, Youngjae Yu^{2†}

¹ KAIST AI ² Yonsei University ³ Letsur Inc.

{kangyeolk, wcho, jchoo}@kaist.ac.kr
{justin_seo, daniel5253, qhdamm23,
pikachuisabird, yjy}@yonsei.ac.kr



YBIGTA Conference 우수 사례 – 마인드벗

인지행동치료(CBT) 기반의 심리 상담 보조 서비스

프로젝트 목표

Problem

대한민국 정신 장애 평생 유병률은 높지만, 진단 후 전문가 상담 비율은 10%밖에 되지 않는다. 주요 원인은 비싼 상담 가격, 사회적 편견, 시공간의 제약이다.

Solution

인지 행동 치료 방법을 적용한 심리 상담 보조 챗봇을 제작해서 기존의 문제였던 비용, 프라이버시, 접근성의 허들을 극복하자.

주요기능

1. 사용자 설문
2. 무드 트래킹 & 간편 상담



<https://github.com/MindBut/mindbut-project>

YBIGTA Conference 우수 사례 – Legent

Graph와 RAG를 이용한 법률 AI Agent

프로젝트 목표

교통사고 발생 시 Hypergraph RAG 기술로 책임 소재(과실 비율)를 판단하고 법률 정보를 제공하는 AI 에이전트 개발

Problem

기존 판례 검색 서비스는 사건들 간의 유기적 연결과 책임 판단 기준의 흐름을 체계적으로 반영하지 못하는 한계 존재

Solution

교통사고 판례의 복잡한 관계를 HyperGraph로 구조화하고, 사용자의 사고 영상/텍스트를 기반으로 RAG 기술을 통해 가장 관련성 높은 판례를 찾아 정확한 분석 리포트를 생성

주요기능

Input: 블랙박스 영상 및 사용자의 상황 설명 텍스트

Output:

- 과실 비율 리포트: 사고 상황 요약 및 유사 판례 기반의 상세 과실 비율 분석.
- 자동 생성 영상: 분석 결과를 '한문철 TV' 스타일 영상 및 TTS 음성으로 자동 생성.

