



AIM Lab.

LLM 기반 한국어 문법 오류 교정 서비스

한국판 Grammarly

이용빈식 캡스톤 디자인

팀장: 이용빈 20191120

팀원: 육정훈 20191780



국립
한밭대학교
HANBAT NATIONAL UNIVERSITY

CONTENTS

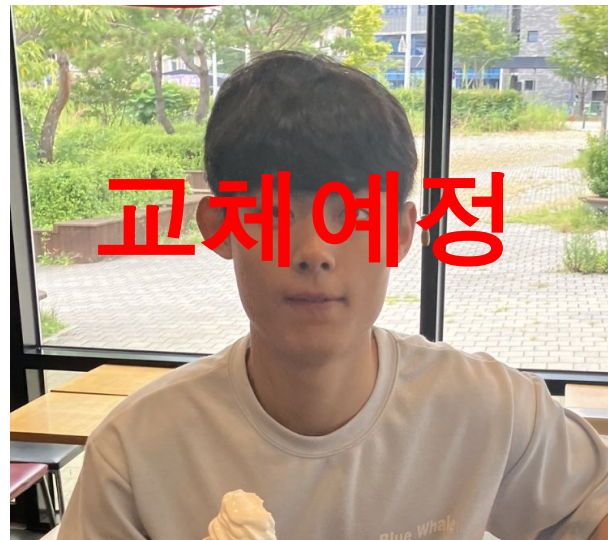
- 01 | 팀원 소개
- 02 | 프로젝트 배경
- 03 | 프로젝트 목표
- 04 | 데이터셋 예시
- 05 | 프로젝트 계획

1. 팀원 소개



이용빈

MLOps 구축
웹 어플리케이션 개발



육정훈

모델 연구 및 개발

2. 프로젝트 배경

글쓰기의 중요성↑, 글쓰기를 보조하는 AI 기반 서비스 일반화

ex) Grammarly, Trink, HyperWrite, Rytr 등

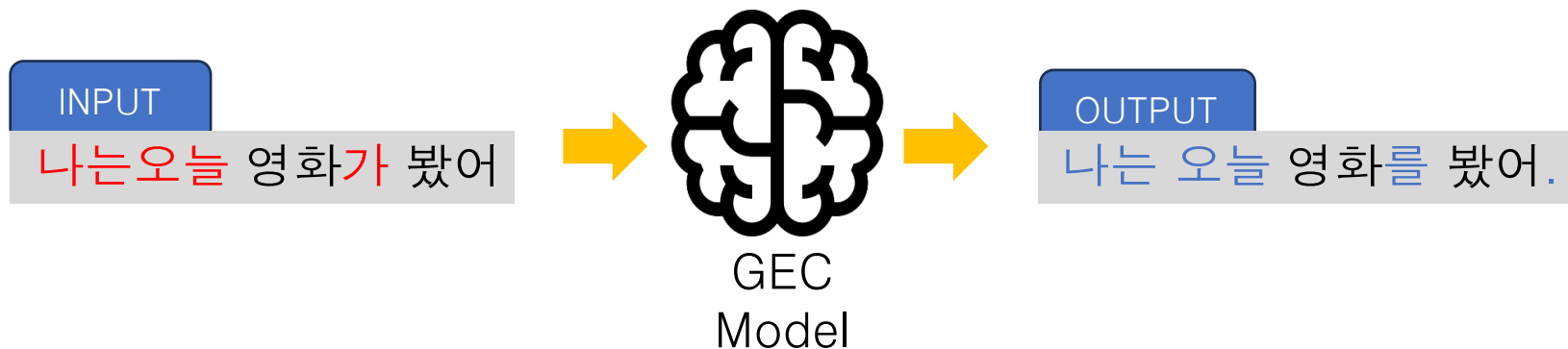


$$\begin{aligned} & \text{글쓰기 보조 서비스} \\ &= \text{GEC} + \text{AES} \end{aligned}$$

GEC = Grammatical Error Correction, 문법 오류 수정

AES = Automated Essay Scoring, 글의 수준 평가

- Grammatical Error Correction(GEC)
 - 인공지능 모델에 기반한 문법 오류 수정 Task



- Grammatical Error Correction(GEC)
 - 실제 서비스중인 Grammarly의 GEC

Untitled document

This **project** aims to develop a Korean version of Grammarly, a grammar and writing **assistent** program for the Korean language. Currently, most grammar correction and writing assessment tools are specialized for **english** and do not support Korean.

3 All suggestions



Correctness · Correct your spelling ⓘ

This **projaect** **project** aims to develop...

Accept

Dismiss

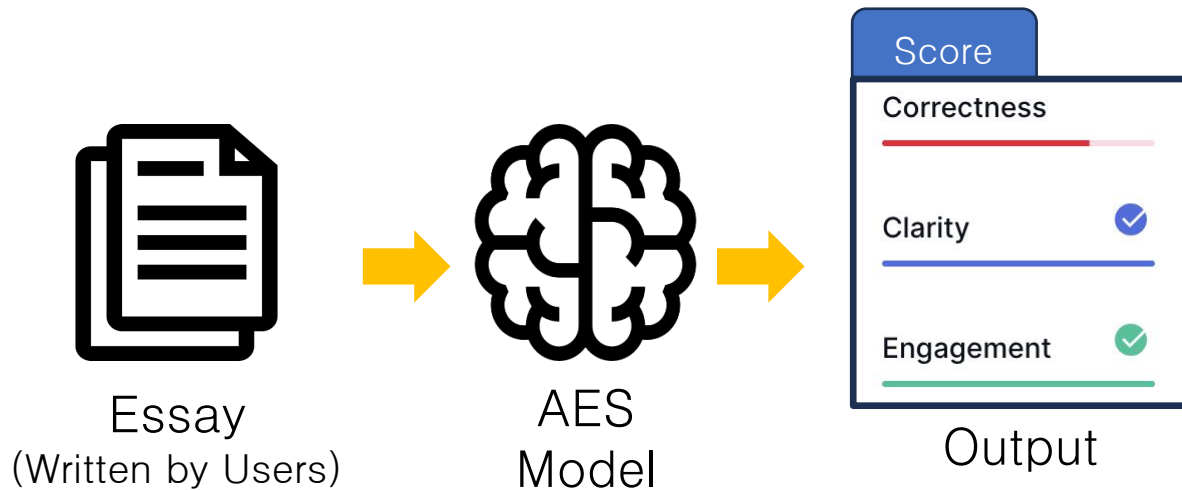


Correct your spelling
assistent



Change the capitalization
english

- Automated Essay Scoring(AES)
 - 인공지능 모델이 글의 수준을 평가하는 Task



- Automated Essay Scoring(AES)
 - 실제 서비스중인 Grammarly의 AES

Correctness



Clarity



Engagement



Delivery



Performance

Text score: 80 out of 100. This score represents the quality of writing in this document. You can increase it by addressing Grammarly's suggestions.

80

Word count

Characters	245	Reading time	8 sec
Words	37	Speaking time	17 sec
Sentences	2		

Readability

Metrics compared to other Grammarly users

Word length	5.5		Above average
Sentence length	18.5		Above average
Readability score	42		

3. 프로젝트 목표

- 한국어 특화 서비스의 부재
 - Grammarly 포함 대부분의 GEC+AES 서비스는 영문 대상
 - 일부 한국어 대상 서비스 또한 단순 맞춤법 교정에 불과함
 - Ex) 부산대 맞춤법 검사기

INPUT

안녕의 계세요. 저는 배가 고파서
단식을 해요. 배불러요!



교정 내용

입력 내용 고파서 단식을 해요

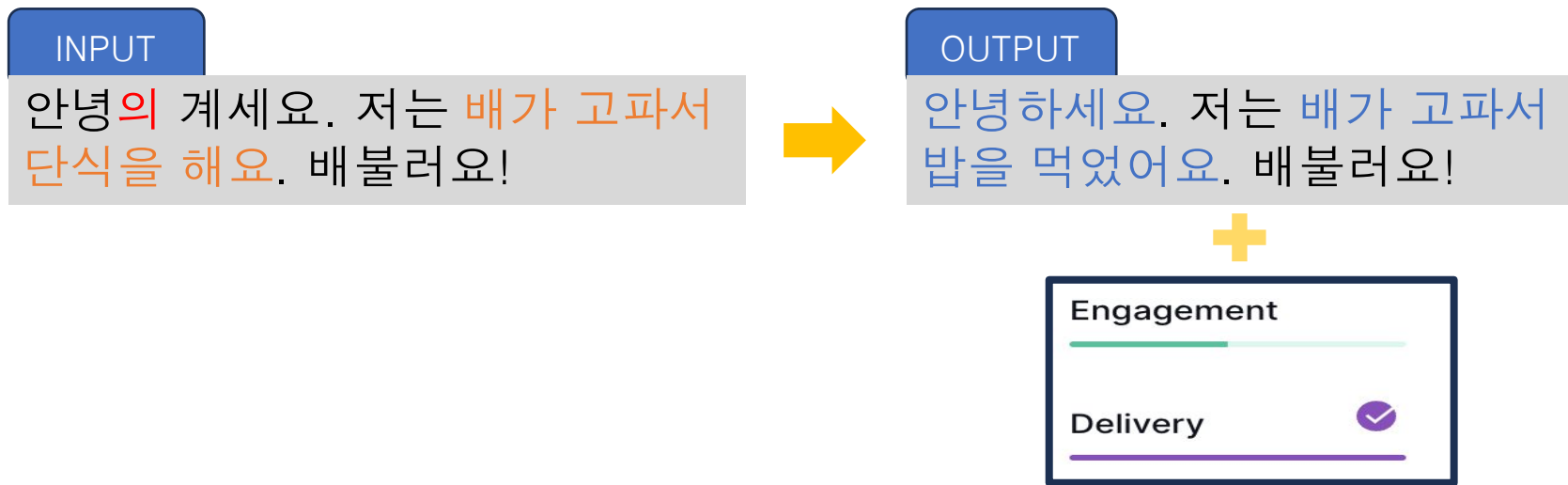
대치어 고파서 단식해요

직접 수정 원하는 대치어를 직접 입력하세요.

적용

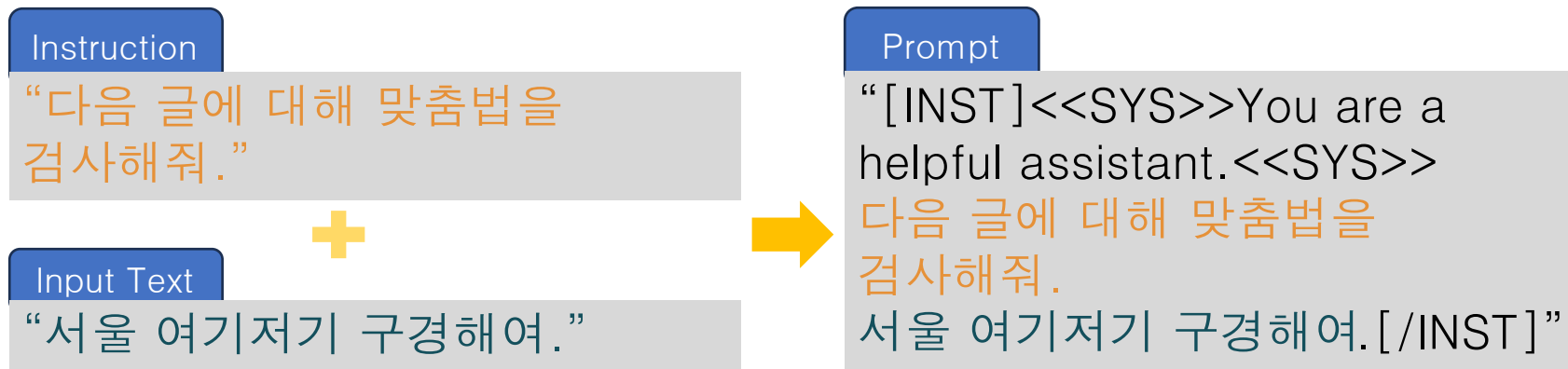
도움말 굳이 조사 '을/를/이/가'를 쓰지 않아도 된다면 쓰지 않습니다.

- 최종 목표: 한국어 특화 LLM 기반 글쓰기 보조 서비스 개발
 - 모델이 한국어 문법을 이해하고 그에 따른 교정 제공
 - 문맥 파악에 따른 전체 글의 다양한 평가 제공



4. 데이터셋 예시

- 데이터셋 구성 예시 – GEC
 - 특정한 Template을 따르는 Instruction 형태로 모델에 전달

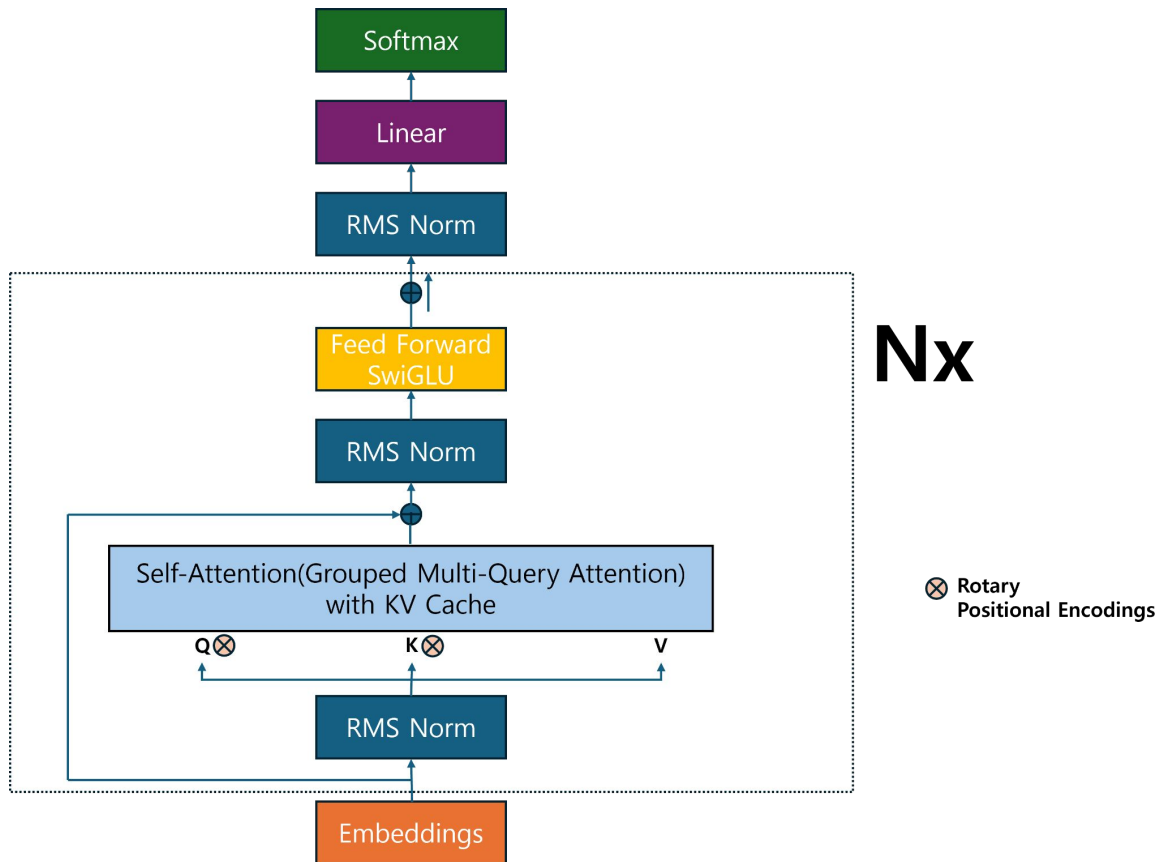


- 데이터셋 구성 예시 - AES
 - 입력 텍스트와 주제의 수준, 평가 점수로 구성



5. 프로젝트 계획

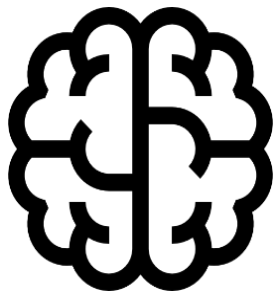
- 모델 개요
 - LLaMA



- 학습 개요
 - 순차적으로 학습
 1. GEC
 2. AES



Input
Sentence



Model



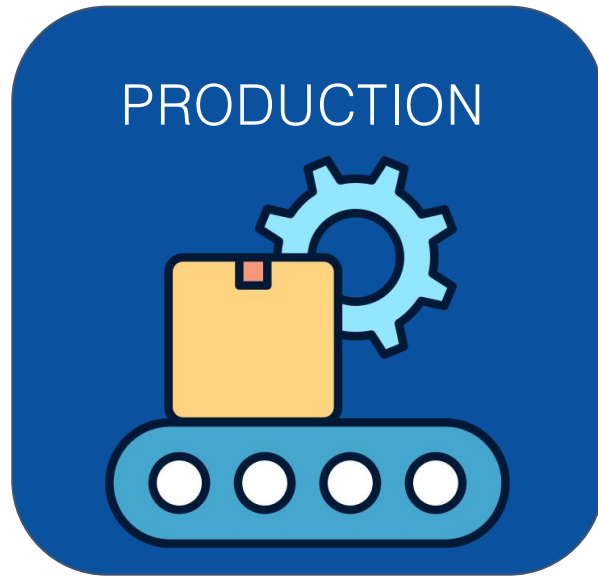
“80”

1. Corrected Sentence
2. Score

Output

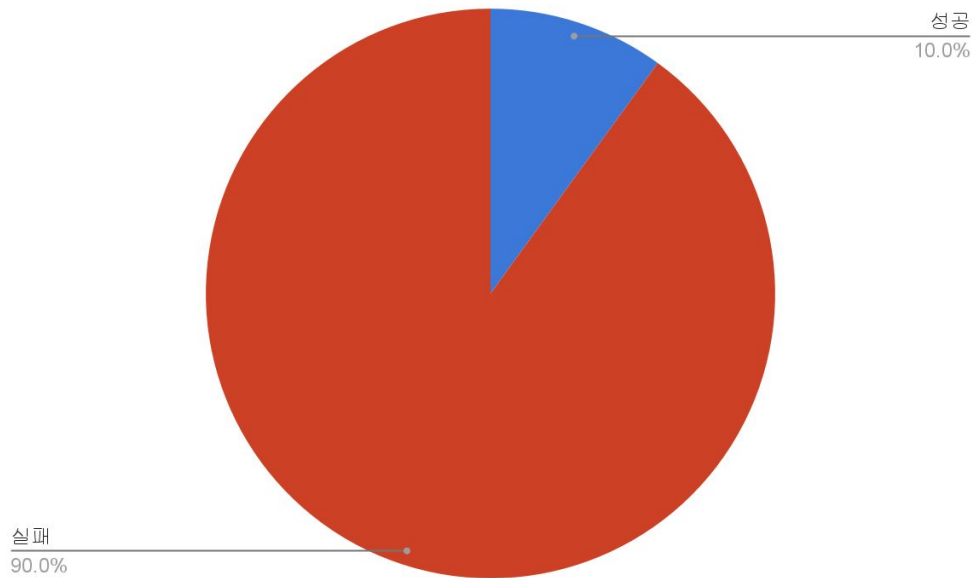


TO

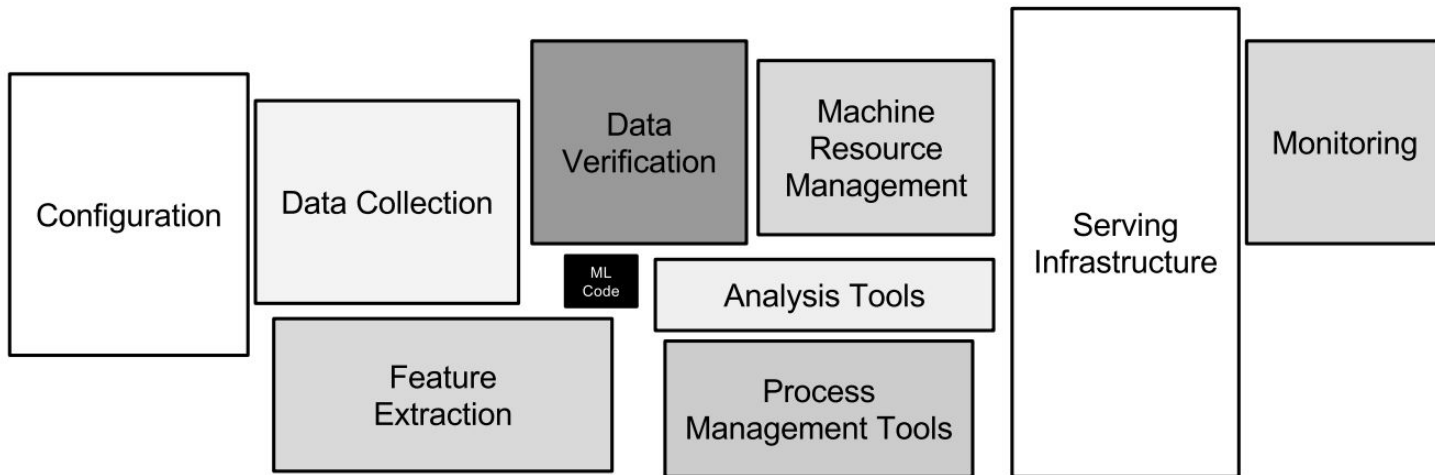


머신러닝 프로젝트

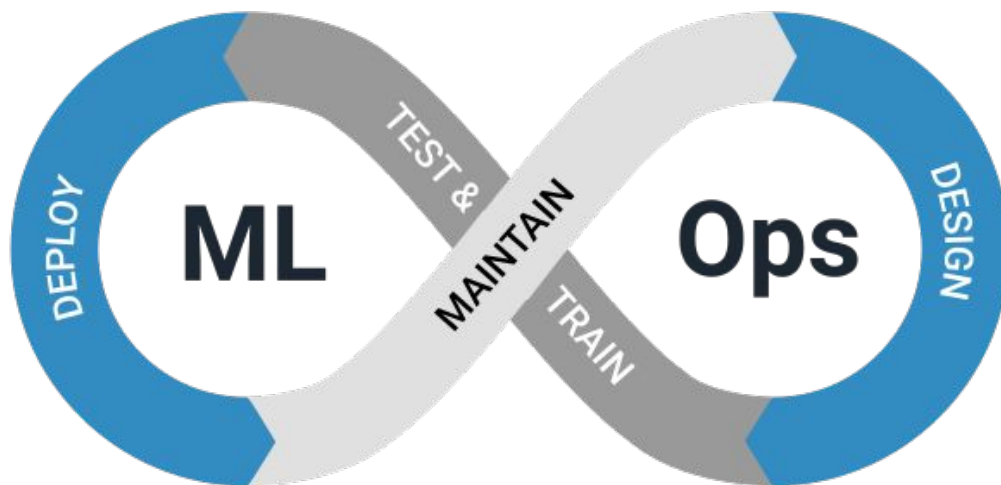
연구 성과 != 비즈니스 성과



모델 코드는 머신러닝 시스템의 일부

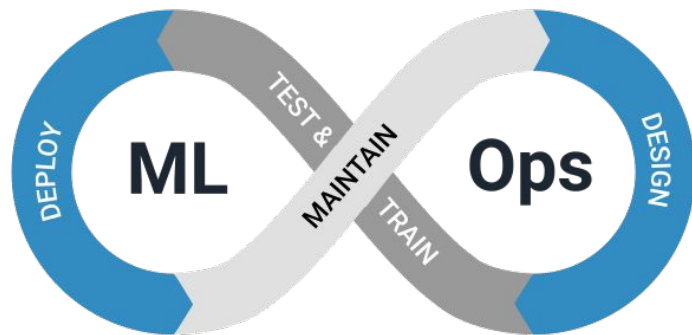


-> MLOps를 도입하여 서비스를 구축



MLOps의 요소

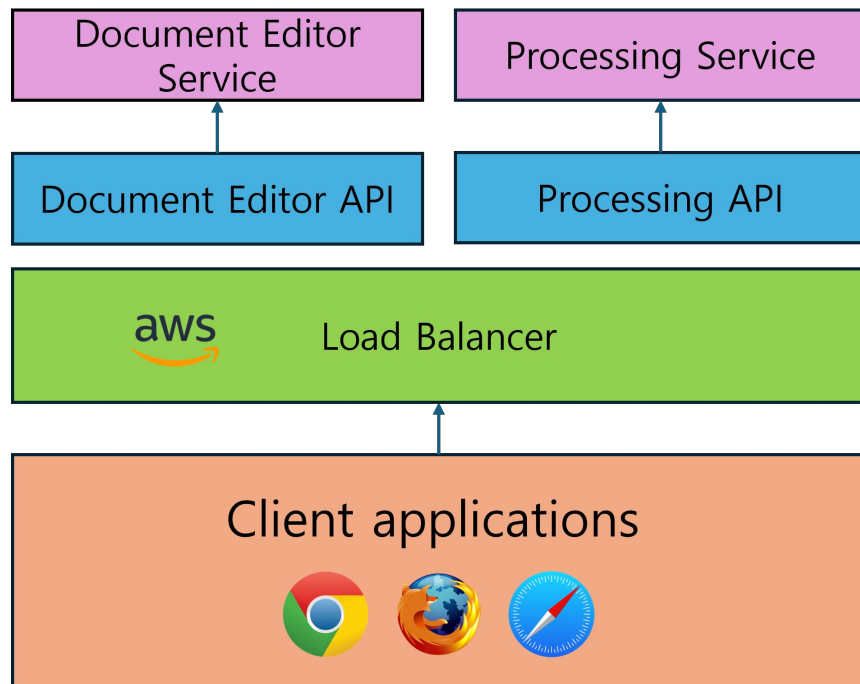
- 모델 파이프라인 자동화
- 모니터링
- 지속적 학습
- 배포 자동화



MLOps 원칙을 준수하여 웹 어플리케이션 구축

-> 확장성, 신뢰성, 유지보수성 등을 갖춘 머신러닝 시스템

예상 서비스 기본 아키텍처



Thank You!

Question & Answer

