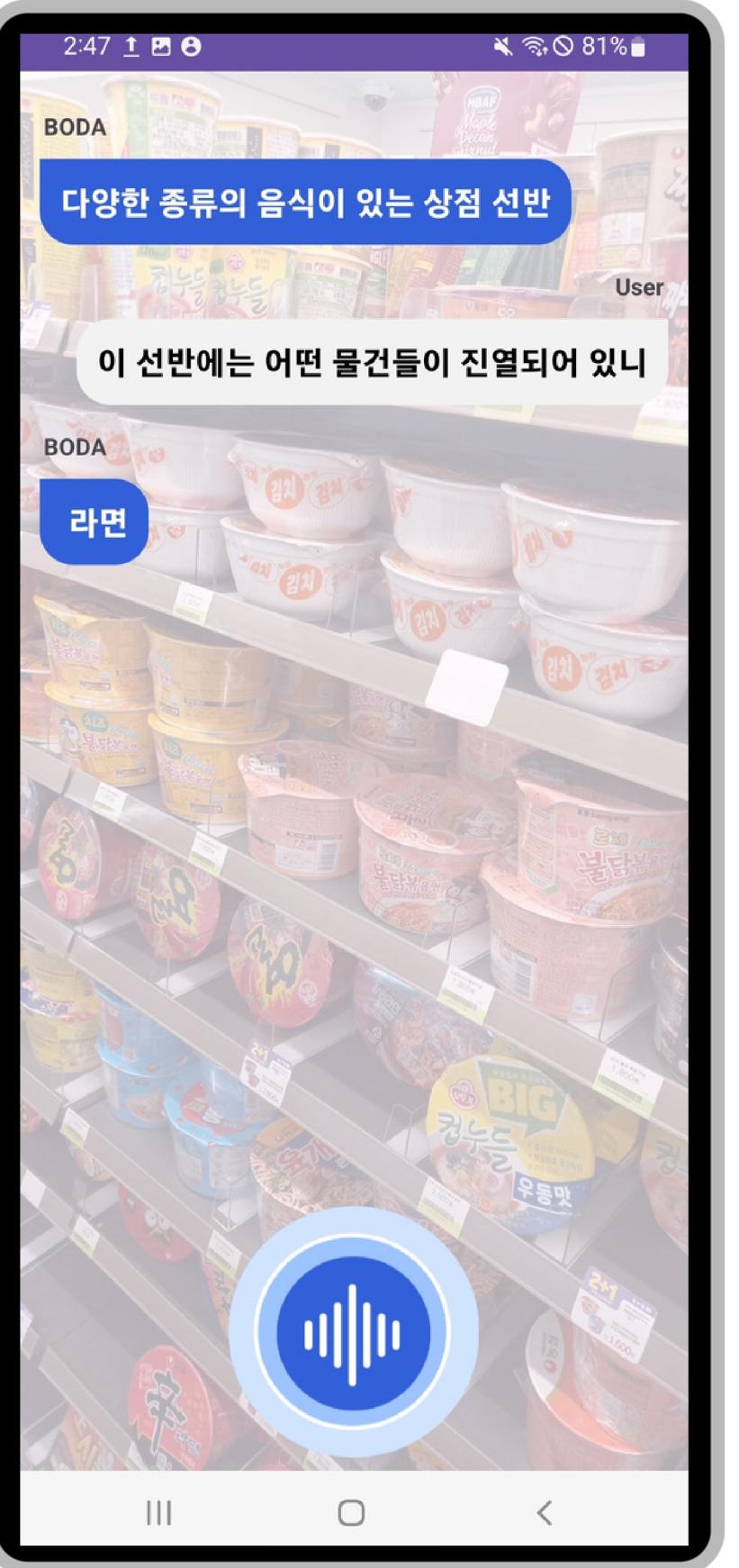




보다 (BODA)

저시력자 및 시각장애인들을 위한
시각적 질의응답 서비스



시각장애인의 실태와 요구사항 파악

KU 건국대학교
KONKUK UNIV.

장애학생지원센터

센터소개

지원업무

장애인식개선교육

게시판

☰

[스마트ICT융합공학과 팀 보다] 시각장애인을 위한 서비스 설문조사

작성자 장애학생지원센터 | 조회수 877 | 등록일 2024.03.15

안녕하세요!
저희는 건국대학교 스마트ICT융합공학과에 재학 중인 팀 보다(BODA, 고현민, 전수현, 정도영)입니다.

저희는 시각장애인의 일상 생활을 보다 편리하게 만들 수 있는 시각적 질의응답 서비스를 개발하고자 합니다.
이를 위해 시각장애인분들의 일상에서 겪는 어려움과 필요한 시각적 정보, 현재 사용하고 계신 보조 기술이나 앱에 대한 의견을 듣고 싶습니다.

추가적으로, 시각장애인 시각적 질의응답 서비스 개발을 위해 인터뷰가 가능하시다면 연락처 또는 마지막 문항에 응답해주세요.
인터뷰 진행자에 한해 소정의 상품을 지급할 예정입니다.

이는 현재 출입 프로젝트의 일환으로, 시각장애인 분들의 일상과 서비스 이용에 대해 깊이 이해하고자 합니다. 이를 통해 시각장애인 분들이 겪는 어려움을 더 잘 파악하고, 개선 방안을 모색하는 것이 목표입니다.

귀하의 소중한 경험과 아이디어는 이 분야의 이해를 깊게 하고 서비스 개선에 매우 중요한 역할을 할 것입니다. 이에 저희는 귀하와의 대면 인터뷰 또는 화상 인터뷰를 통해 귀하의 경험을 직접 듣고자 합니다. 인터뷰는 귀하의 편의에 따라 시간과 장소를 조정할 수 있으며, 대략 30분 정도 소요될 예정입니다.

저희는 귀하의 경험을 공유해주시다면 큰 도움이 될 것이며, 귀하의 의견은 서비스 개선을 위해 매우 소중하게 활용될 것입니다. 만약 귀하께서 이 인터뷰에 참여해 주실 의향이 있으시다면, 제기 제공한 연락처 정보로 연락 주시면 감사하겠습니다. 귀하의 편의에 맞는 시간에 인터뷰를 진행할 수 있도록 하겠습니다.

이 프로젝트에 대해 궁금하신 점이 있으시다면 언제든지 질문해 주시기 바랍니다. 귀하의 귀중한 시간을 활용해 주셔서 다시 한번 감사드리며, 귀하의 이야기를 듣고 개선을 위해 함께 노력하기를 간절히 기대합니다.

감사합니다.
담당자 연락처: 010-2393-3573 (고현민)

설문조사 링크
<https://forms.gle/7vCa1An8NBCCClkma8>

↑
TOP

시각장애인을 위한 서비스 설문조사

질문 응답 2 설정

시각장애인을 위한 서비스 설문조사

B I U GO X

안녕하세요!
저희는 건국대학교 스마트ICT융합공학과에 재학 중인 팀 보다(BODA, 고현민, 전수현, 정도영)입니다.

저희는 시각장애인의 일상 생활을 보다 편리하게 만들어줄 수 있는 시각적 질의응답 서비스를 개발하고자 합니다. 이를 위해 시각장애인분들의 일상에서 겪는 어려움과 필요한 시각적 정보, 현재 사용하고 계신 보조 기술이나 앱에 대한 의견을 듣고 싶습니다.

추가적으로, 시각장애인 시각적 질의응답 서비스 개발을 위해 인터뷰가 가능하시다면 담당자 연락처 또는 마지막 문항에 응답해 주세요. 인터뷰 진행자에 한해 소정의 상품을 제공할 예정입니다.

이는 현재 풀업 프로젝트의 일환으로, 시각장애인 분들의 일상과 서비스 이용에 대해 깊이 이해하고자 합니다. 이를 통해 시각장애인 분들이 겪는 어려움을 더 잘 파악하고, 개선 방안을 모색하는 것이 목표입니다.

귀하의 소중한 경험과 아이디어는 이 분야의 이해를 깊게 하고 서비스 개선에 매우 중요한 역할을 할 것입니다. 이에 저희는 귀하와의 대면 인터뷰 또는 화상 인터뷰를 통해 귀하의 경험을 직접 듣고자 합니다. 인터뷰는 귀하의 편의에 따라 시간과 장소를 조정 할 수 있으며, 대략 30분 정도 소요될 예정입니다.

저희는 귀하의 경험을 공유해주신다면 큰 도움이 될 것이며, 귀하의 의견은 서비스 개선을 위해 매우 소중하게 활용될 것입니다. 만약 귀하께서 이 인터뷰에 참여해 주실 의향이 있으시다면, 제가 제공한 연락처 정보로 연락 주시면 감사하겠습니다. 귀하의 편의에 맞는 시간에 인터뷰를 진행할 수 있도록 하겠습니다.

이 프로젝트에 대해 궁금하신 점이 있으시다면 언제든지 질문해 주시기 바랍니다. 귀하의 귀중한 시간을 할애해 주셔서 다시 한번 감사드리며, 귀하의 이야기를 듣고 개선을 위해 함께 노력하기를 간절히 기대합니다.

감사합니다.

담당자 연락처: 010-2393-3573 (고현민)

건국대학교 장애학생지원센터

 **한국시각장애인연합회**
KOREA BLIND UNION

한국시각장애인연합회
한국시각장애인연합회

English | 로그인 | 회원가입 | 사이트맵

연합회소개 사업소개 시설소개 알림광장 **열린광장** 자료실 시각장애인의 이해

| 열린광장

자유게시판 ...	문의하기	제활상담	자주하는질문	정보나눔
장애인뉴스클리핑				

○ 자유게시판 - 시각장애인을 위한 어플리케이션 설문조사

작성자	고령민	작성일시	2024-03-19 오후 4:31:23
첨부파일	없음		

안녕하세요!

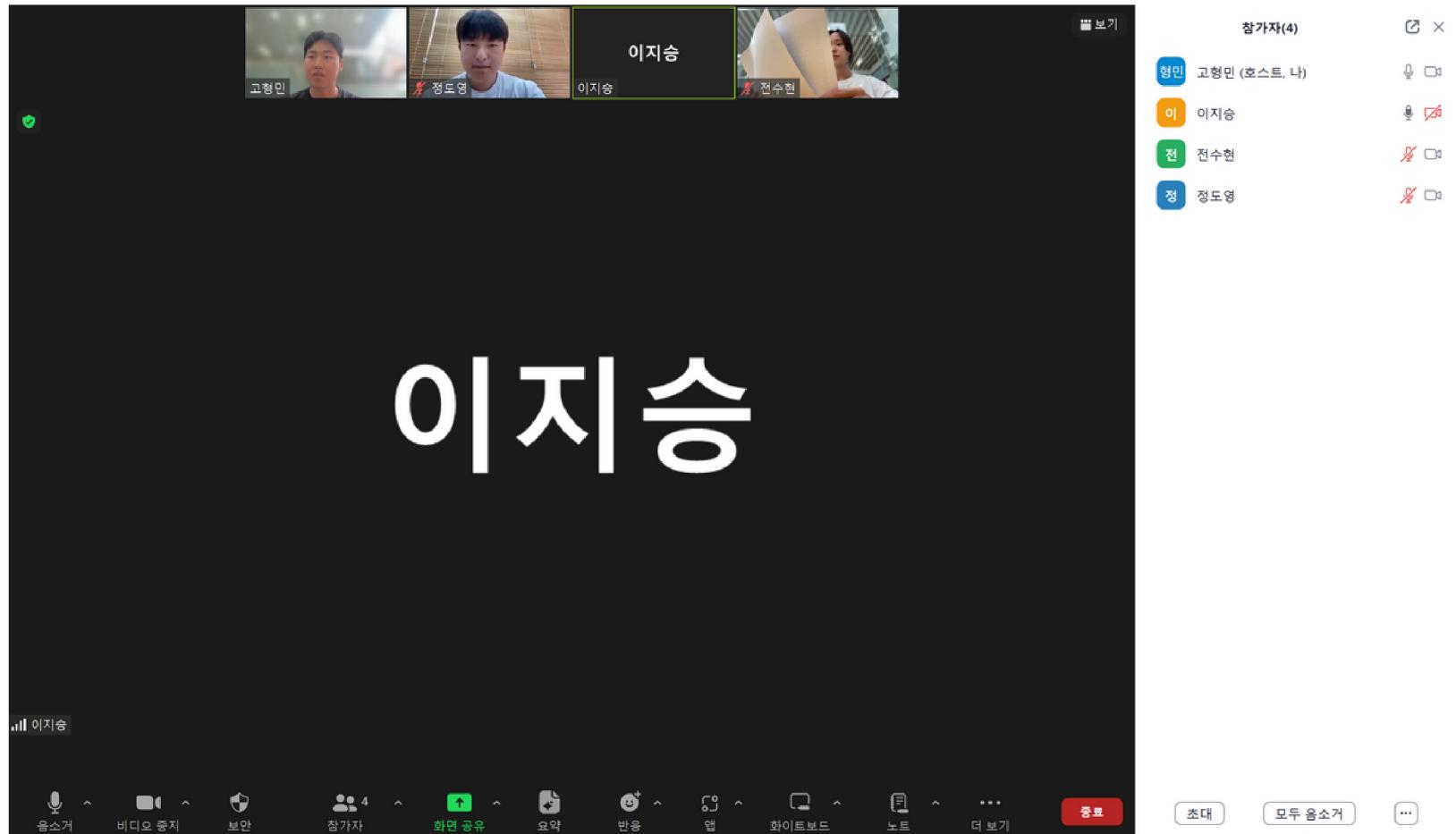
저희는 건국대학교 스마트ICT융합공학과에 재학 중인 팀 보다(BODA, 고령민, 전수현, 정도영)입니다.

저희는 시각장애인의 일상 생활을 보다 편리하게 만들어줄 수 있는 시각적 질의응답 서비스를 개발하고자 합니다.

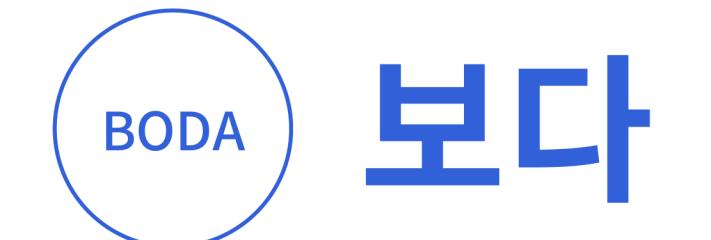
이를 위해 시각장애인분들의 일상에서 겪는 어려움과 필요한 시각적 정보, 현재 사용하고 계신 보조 기술이나 앱에 대한 의견을 듣고 싶습니다.

구글 설문조사 품

Problem



Solution |



시각장애인의 **눈**이 되어주는
손안의 마법 같은 지팡이

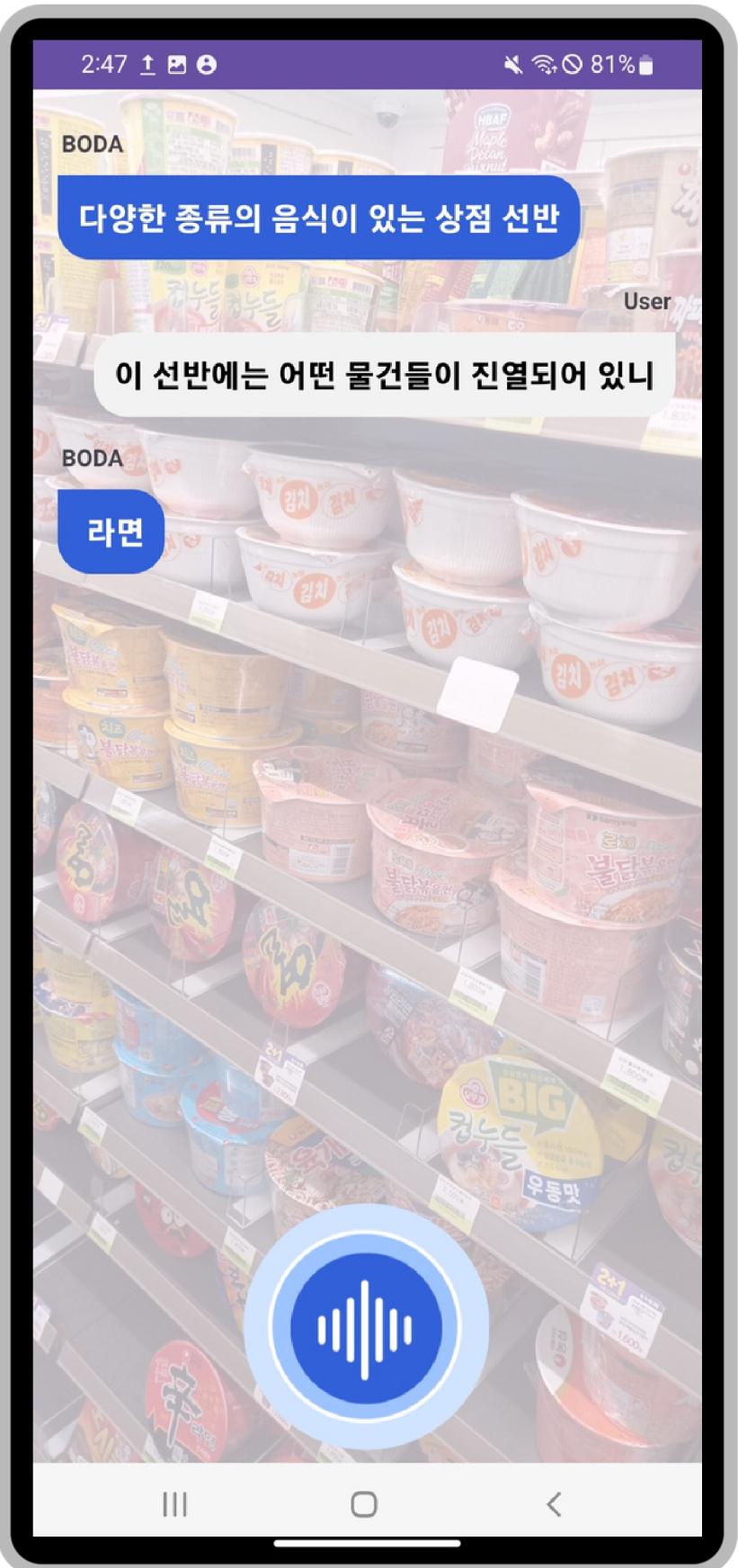
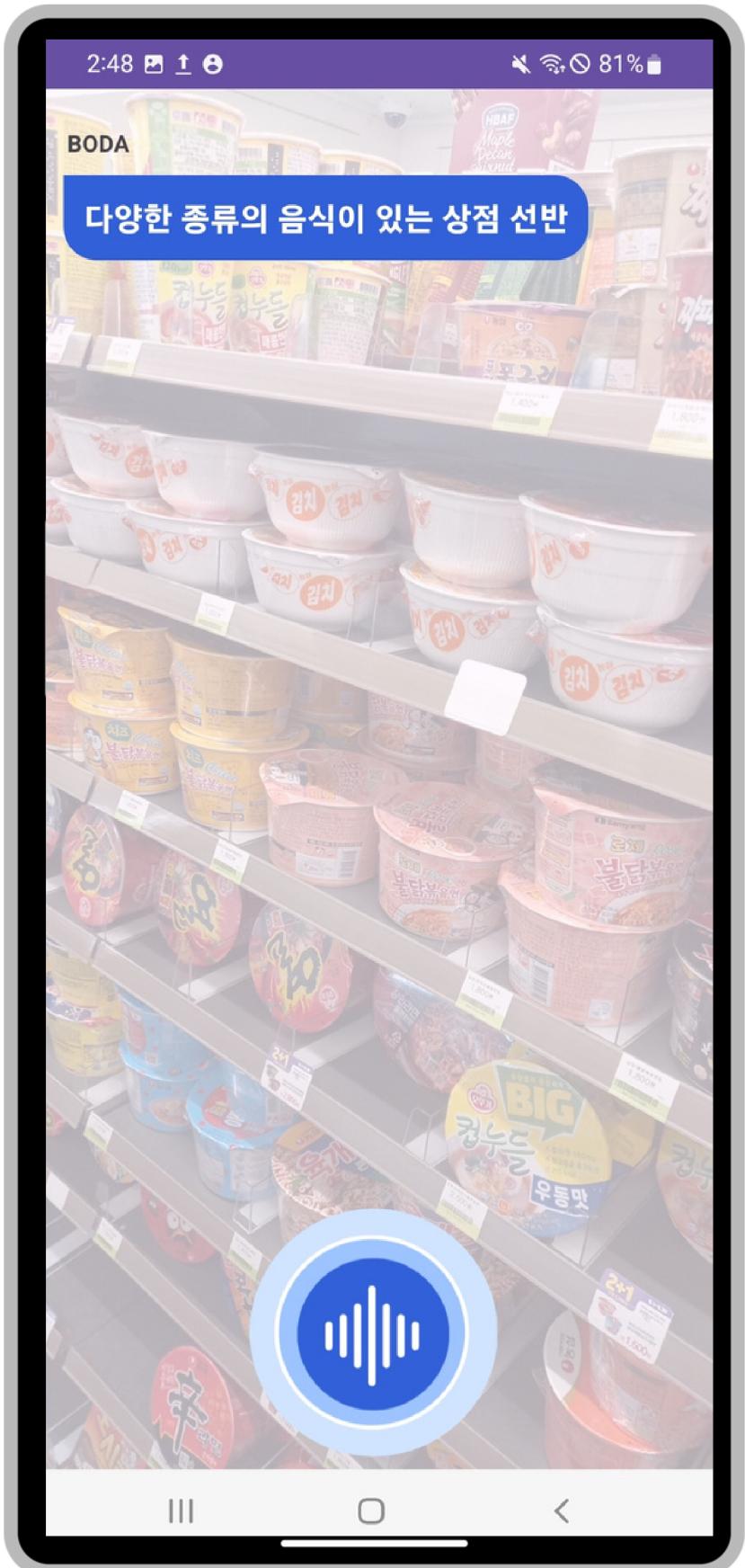


사진 촬영을 통해 간단히 설명!

일상생활에서 간단하게 사진을 찍어보세요.
BODA가 사진에 대한 전체적인 설명을 음성으로 알려드려요.

“다양한 종류의 음식이 있는 상점 선반”

“테이블 위에 올려진 립스틱”



Solution |

말 한마디로 한 번에 질문!

이제 질문도 말로 하세요. 텍스트로 입력하지 않아도 문제없어요.
당신이 원하는 대답을 가장 빠르고 쉽게 알려드려요.

“이 물건의 색은 어떤 색이야?”

“이 선반에는 어떤 물건들이 진열되어 있니?”

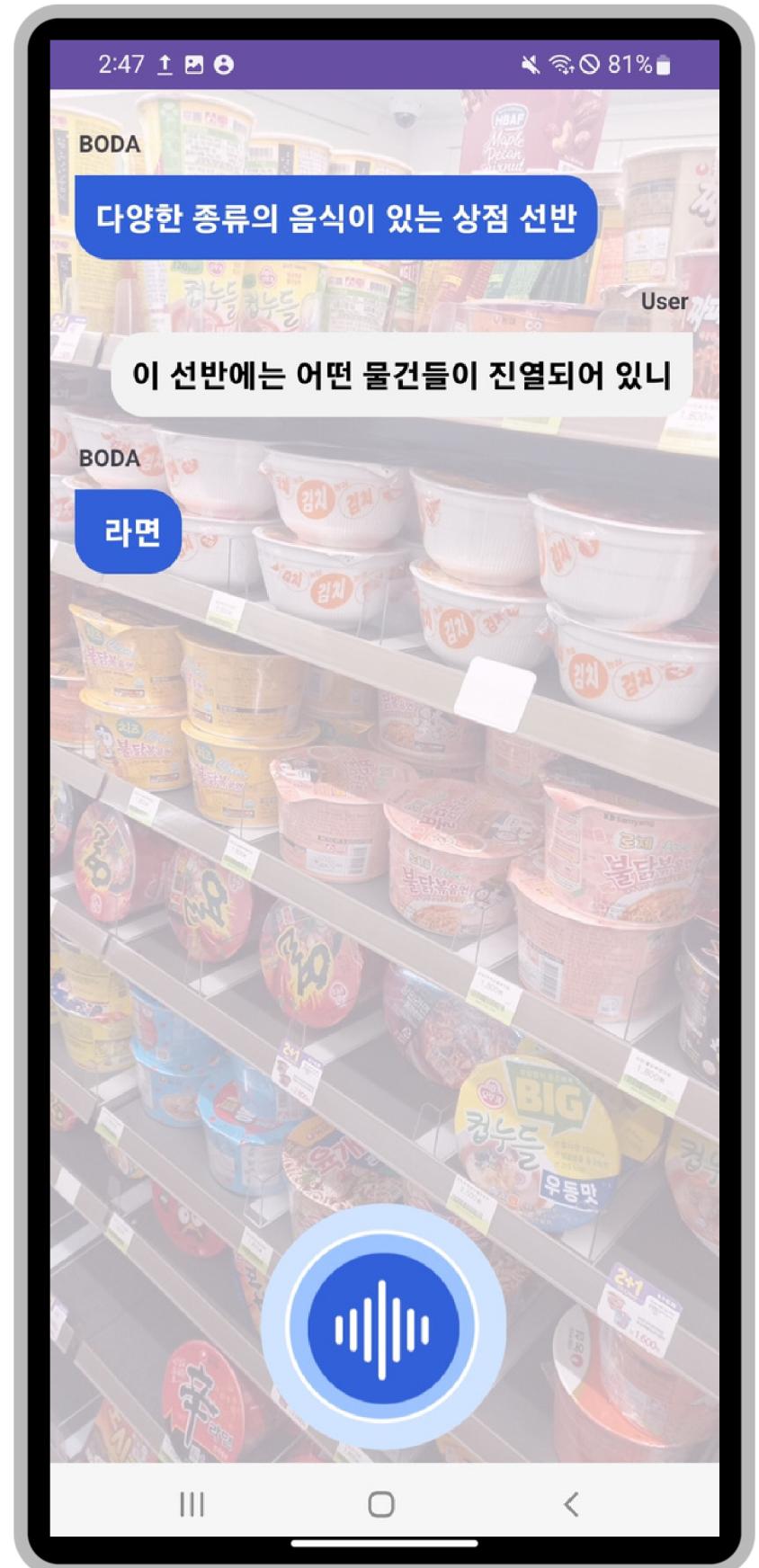


Image Captioning

사진에 대한 전체적인 설명을 제공

The Illustrated Image Captioning using transformers

- Image Encoder인 ViT로 이미지 특징을 추출한 후, Text Decoder인 GPT-2로 캡션을 출력하는 모델.
- COCO dataset으로 학습시켜 허깅페이스에 공개로 배포된 [nlpconnect/vit-gpt2-image-captioning](https://huggingface.co/nlpconnect/vit-gpt2-image-captioning) 모델을 다운받아 사용

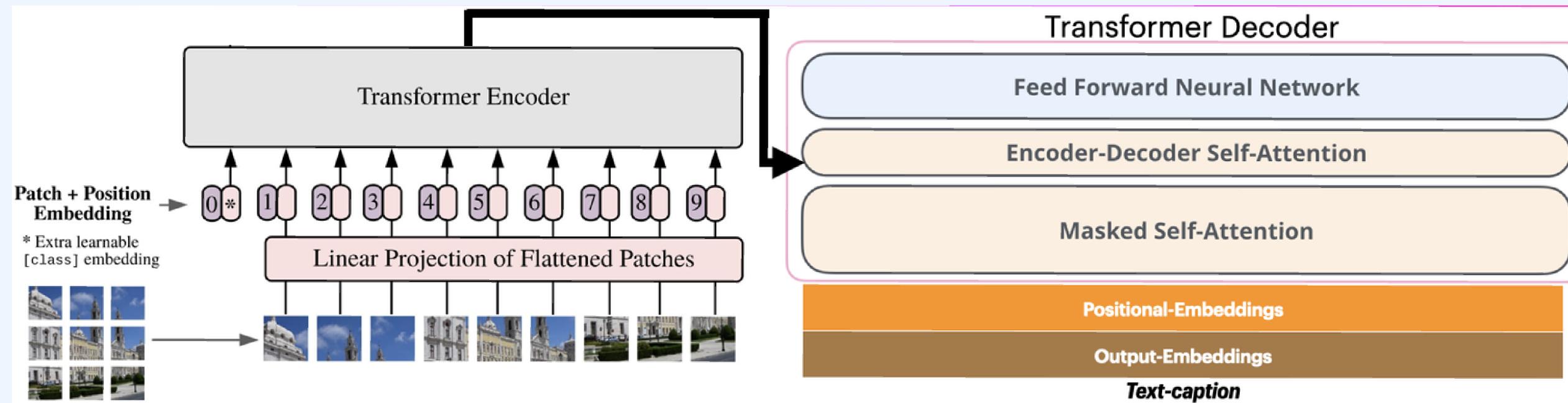


Image Captioning 모델

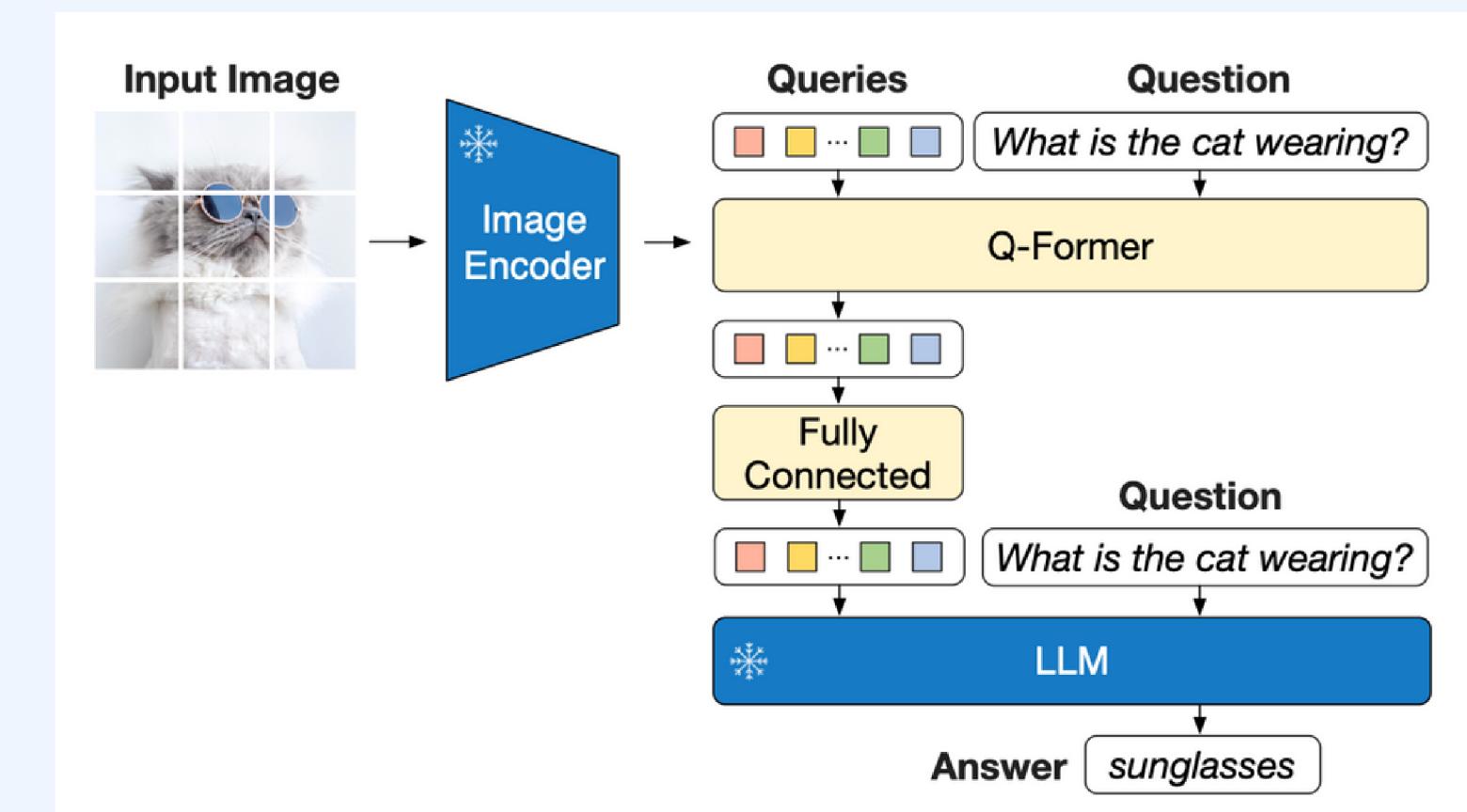
- <https://huggingface.co/nlpconnect/vit-gpt2-image-captioning> : 개인(Ankur Kumar)이 pre-train시켜 배포한 모델 사용

Visual Question and Answering

사용자의 질문에 대한 실시간 답변 가능

SOTA*성능의 BLIP-2 모델 사용

- frozen pretrained Image Encoder와 Large Language Model(LLM)을 사용하여 Image Captioning과 Visual Question Answering 태스크에서 SOTA 성능을 달성한 모델.
- Zero-shot Learning에서도 효과적일 뿐만 아니라, frozen된 LLM을 제외하고 나머지 부분만 훈련시켜 효율적인 파인튜닝이 가능.



* SOTA : State-of-the-Art. 현재 최고 수준의 모델

** BLIP-2 : <https://huggingface.co/Salesforce/blip2-opt-2.7b> : salesforce가 배포한 blip2-opt-2.7b 모델 사용

Solution

VizWiz Question Answering Dataset

시각장애인들이 실제 일상에서 겪는 어려움을 해결하기 위한 데이터셋

시각장애인이 직접 촬영한 사진에 대한 질의응답 데이터셋 -> 흔들리거나 사진과 관련없는 질문 다수



Q: Does this foundation have any sunscreen?
A: yes



Q: What is this?
A: 10 euros



Q: What color is this?
A: green



Q: Please can you tell me what this item is?
A: butternut squash red pepper soup



Q: Is it sunny outside?
A: yes



Q: Is this air conditioner on fan, dehumidifier, or air conditioning?
A: air conditioning



Q: What type of pills are these?
A: unsuitable image



Q: What type of soup is this?
A: unsuitable image



Q: Who is this mail for?
A: unanswerable



Q: When is the expiration date?
A: unanswerable



Q: What is this?
A: unanswerable



Q: Can you please tell me what the oven temperature is set to?
A: unanswerable



1. VizWiz QA dataset 중 이미지와 질문의 연관성이 높은 데이터만을 추출
2. 파인튜닝을 위해 데이터 수 조절



Q : What's the name of this product?
A : basil leaves



Q : What's this?
A : boots



Q : What color... what color is this skirt?
A : white



Q : What is this game?
A : grand theft auto vice city

training image/question pairs : 20,523

validation image/question pairs : 4,319

test image/question pairs : 8,000

training image/question pairs : 10,000

validation image/question pairs : 1,000

test image/question pairs : 100

*vizwiz dataset : <https://vizwiz.org/>

** vizwiz question and answering dataset (파인튜닝에 실제로 사용한 데이터 url) : <https://www.kaggle.com/datasets/lhanhsin/vizwiz>

Solution

VizWiz Question Answering Dataset

시각장애인들이 실제 일상에서 겪는 어려움을 해결하기 위한 데이터셋

시각장애인이 직접 촬영한 사진에 대한 질의응답 데이터셋 -> 흔들리거나 사진과 관련없는 질문 다수



Q: Does this foundation have any sunscreen?
A: yes



Q: What is this?
A: 10 euros



Q: What color is this?
A: green



Q: Please can you tell me what this item is?
A: butternut squash red pepper soup



Q: Is it sunny outside?
A: yes



Q: Is this air conditioner on fan, dehumidifier, or air conditioning?
A: air conditioning



Q: What type of pills are these?
A: unsuitable image



Q: What type of soup is this?
A: unsuitable image



Q: Who is this mail for?
A: unanswerable



Q: When is the expiration date?
A: unanswerable



Q: What is this?
A: unanswerable



Q: Can you please tell me what the oven temperature is set to?
A: unanswerable



1. VizWiz QA dataset 중 이미지와 질문의 연관성이 높은 데이터만을 추출
2. 파인튜닝을 위해 데이터 수 조절



Q : What's the name of this product?
A : basil leaves



Q : What's this?
A : boots



Q : What color... what color is this skirt?
A : white



Q : What is this game?
A : grand theft auto vice city

training image/question pairs : 20,523

validation image/question pairs : 4,319

test image/question pairs : 8,000

training image/question pairs : 10,000
validation image/question pairs : 1,000
test image/question pairs : 100

*vizwiz dataset : <https://vizwiz.org/>

** vizwiz question and answering dataset (파인튜닝에 실제로 사용한 데이터 url) : <https://www.kaggle.com/datasets/lhanhsin/vizwiz>

Solution

시각장애인 시스템 개발을 위한 VQA 데이터셋

일상생활에서 시각장애인들이 겪는 불편함과 궁금증을 해결하기 위한 시각적 정보의 질문과 답변



Q : What is the material of the sink top?

A : marble



Q : How many doors does a white closet have?

A : 7



Q : What color is the chair in front of the piano?

A : pink



Q : Which side of the bed is the bookshelf on?

A : right

- 우리나라의 실내 및 실외 생활 거주 환경에서 촬영된 **한국어 질문과 한국어 답변 쌍**
- 객체와 상황에 대한 이해를 요구하는 질문들로 구성.
- 대부분의 데이터가 이미지의 객체에 대해 물체의 재질, 색상, 개수 그리고 위치를 묻는 질문과 답변으로 구성

training image/question pairs : 224,499

test image/question pairs : 15,128



1. google translation의 문서 번역 시스템으로 한->영 번역
2. train data/validation data/test data 무작위 추출
3. 파인튜닝을 위해 데이터 수 조절

이미지에 대한 영어 질문-답 쌍으로 구성

training image/question pairs : 20,000

validation image/question pairs : 2,000

test image/question pairs : 100

*AI Hub에서 배포한 “시각장애인 시스템 개발을 위한 VQA 모델”이름의 데이터셋 사용

<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=454>

시각장애인 시스템 개발을 위한 VQA 데이터셋

일상생활에서 시각장애인들이 겪는 불편함과 궁금증을 해결하기 위한 시각적 정보의 질문과 답변



Q : What is the material of the sink top?

A : marble



Q : How many doors does a white closet have?

A : 7



Q : What color is the chair in front of the piano?

A : pink



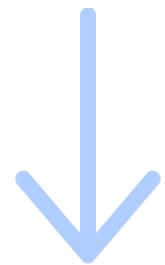
Q : Which side of the bed is the bookshelf on?

A : right

- 우리나라의 실내 및 실외 생활 거주 환경에서 촬영된 **한국어 질문과 한국어 답변 쌍**
- 객체와 상황에 대한 이해를 요구하는 질문들로 구성.
- 대부분의 데이터가 이미지의 객체에 대해 물체의 재질, 색상, 개수 그리고 위치를 묻는 질문과 답변으로 구성

training image/question pairs : 224,499

test image/question pairs : 15,128



1. google translation의 문서 번역 시스템으로 한->영 번역
2. train data/validation data/test data 무작위 추출
3. 파인튜닝을 위해 데이터 수 조절

이미지에 대한 영어 질문-답 쌍으로 구성

training image/question pairs : 20,000

validation image/question pairs : 2,000

test image/question pairs : 100

Solution |

PEFT* method로 Fine-tuning 진행

*Parameter-Efficient Fine-Tuning of Billion-Scale Models on Low-Resource Hardware

all params:
3,749,922,816



TRAINABLE: 0.1398%

PEFT

- 사전 학습된 모델의 대부분의 파라미터는 고정(freeze)한 채로 일부 파라미터만을 학습.
- 적은 수의 파라미터만을 학습시킴으로써 전체 모델을 파인튜닝하는 것과 유사한 효과.

Solution |

Fine-tuned Visual QA Model

파인튜닝된 시각적 질의응답 모델 결과 예시



Q : What is written here?

A : Hello World



Q : What's on display?

A : pots, pottery and utensils



Q : What is he doing?

A : playing computer games

Fine-tuned Visual QA Model

파인튜닝된 시각적 질의응답 모델 결과 성능

VizWiz Baseline (기본 모델의 성능 수치)

BASELINE	accuracy
VizWiz Paper*	0.475

Fine-tuned Model (서비스 모델의 성능 수치)

	BLEU	accuracy
Fine-tuned VQA model	0.16	0.54

TEST DATA

- VizWiz : 100
- 시각장애인 시스템 개발을 위한 VQA 데이터셋 : 100
-> TOTAL : 200

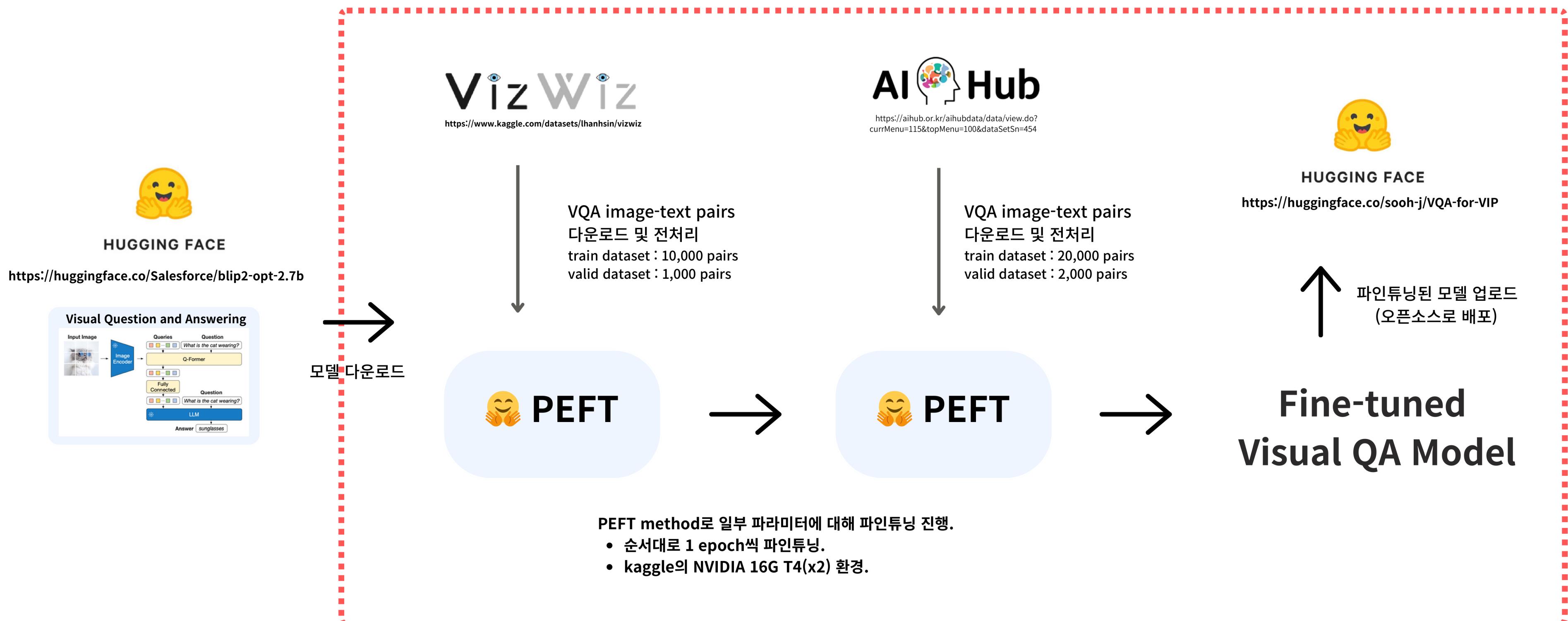
*Gurari, D., Li, Q., Stangl, A. J., Guo, A., Lin, C., Grauman, K., ... & Bigham, J. P. (2018).

Vizwiz grand challenge: Answering visual questions from blind people. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3608-3617).

Solution |

[SUMMARY]

VisWiz QA dataset으로 Visual QA 모델 Fine-tuning 후 업로드



[프로젝트 진행 내용]

Frontend |

Tool

Language & build : Kotlin

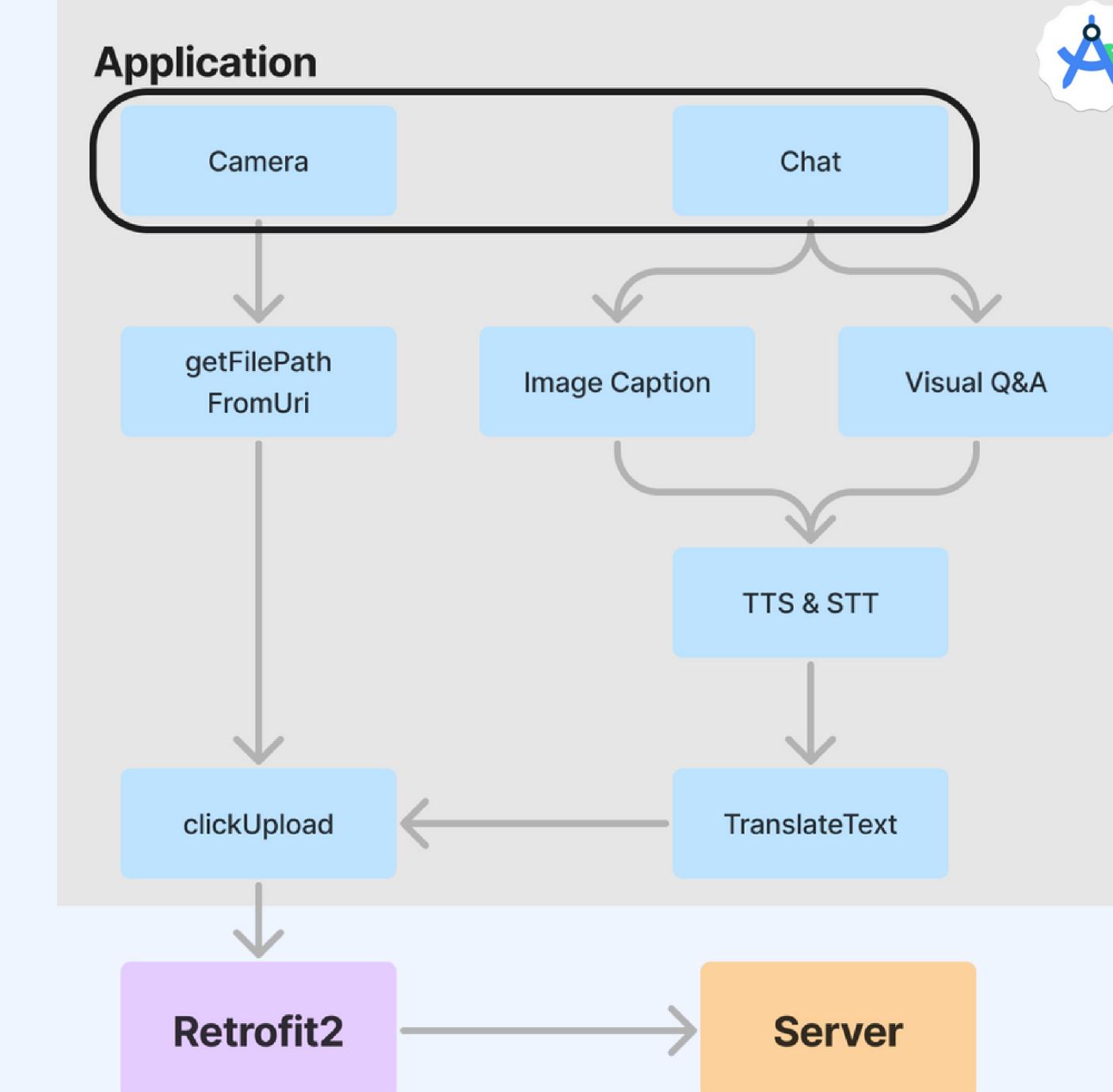
Library : CameraX, Retrofit2, STT

IDE : Android Studio

Target Device: Samsung Galaxy A31

Collection : Git, Github

Frontend Architecture



Frontend |

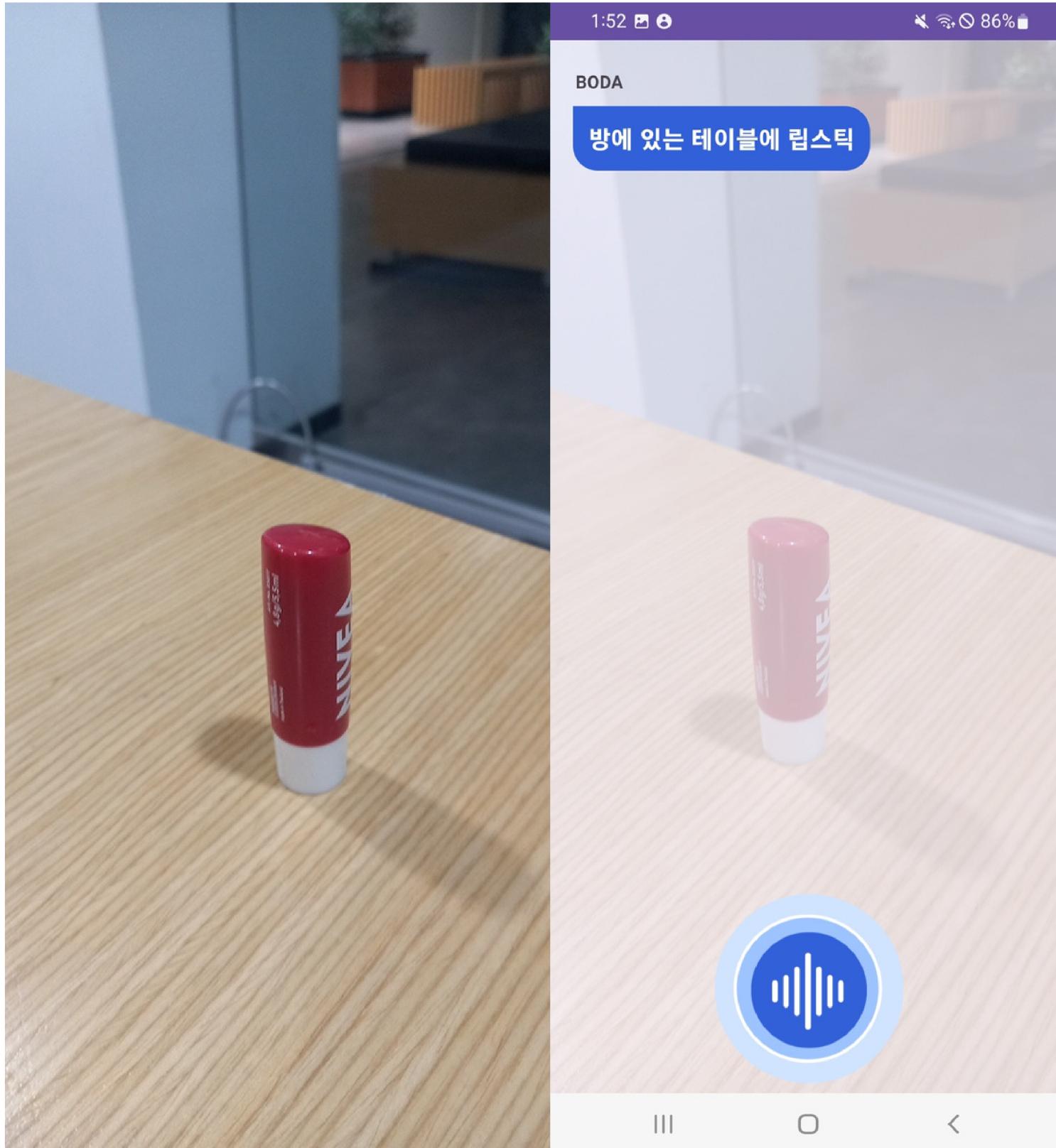


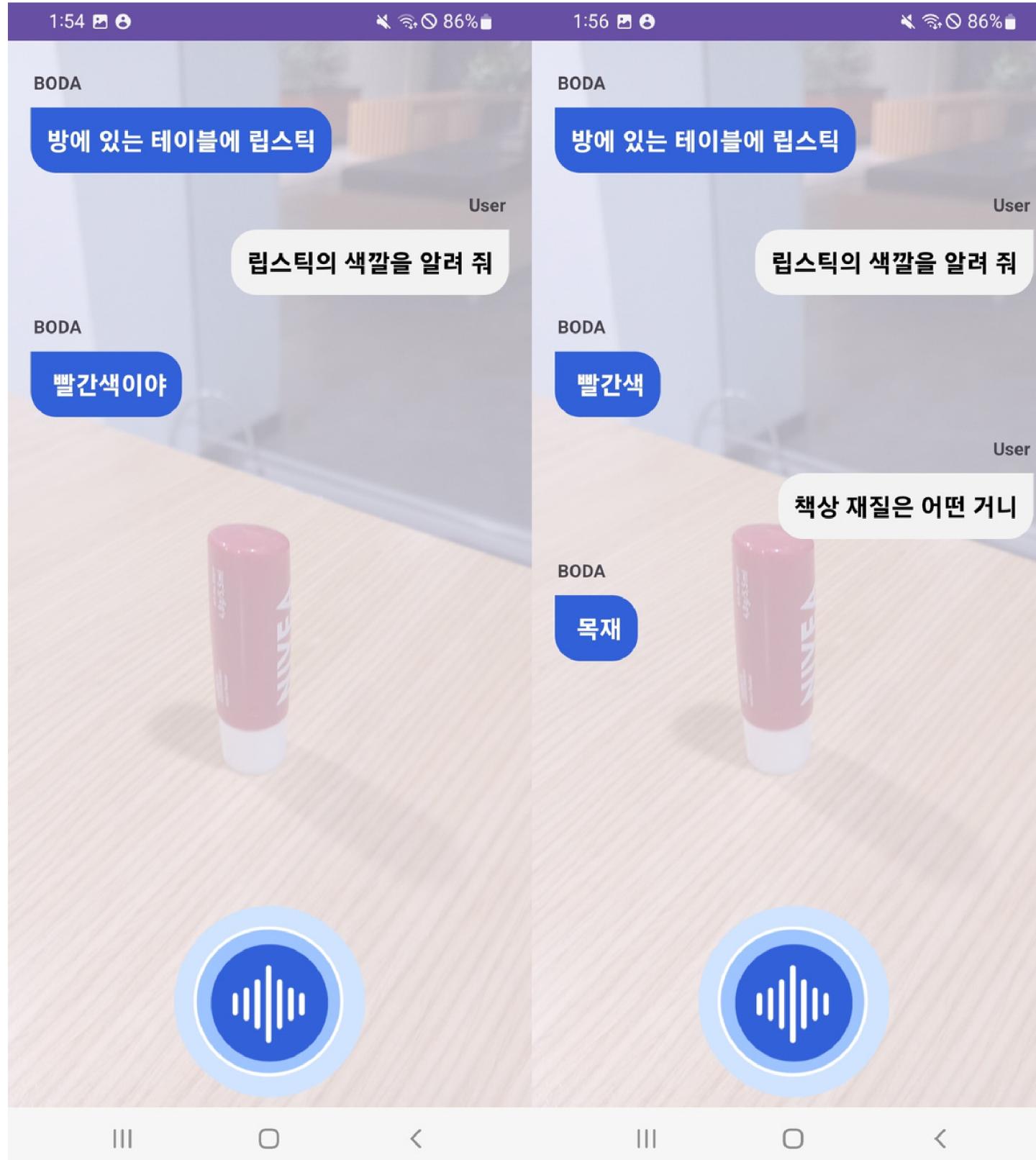
Image Captioning

라이브러리
Camera X, TTS, Google Cloud Translation

사용 방법
화면을 한 번 터치 시 “사진 촬영” 음성 출력
화면을 이중 탭 할 시 Image Captioning 진행

=> 기존 시각 장애인들에게
익숙한 어플 사용 방식 도입

Frontend |



Visual Q&A

라이브러리
STT, TTS, Google Cloud Translation

사용 방법

하단 음성 입력 버튼 한 번 터치 시 “질문하기” 음성 출력
하단 음성 입력 버튼 이중 탭 할 시 Visual Q&A 진행

=> 사진에 대한 추가 질문 가능

Backend

Tool

Language & build : Typescript

Library & Framework : Nest.js

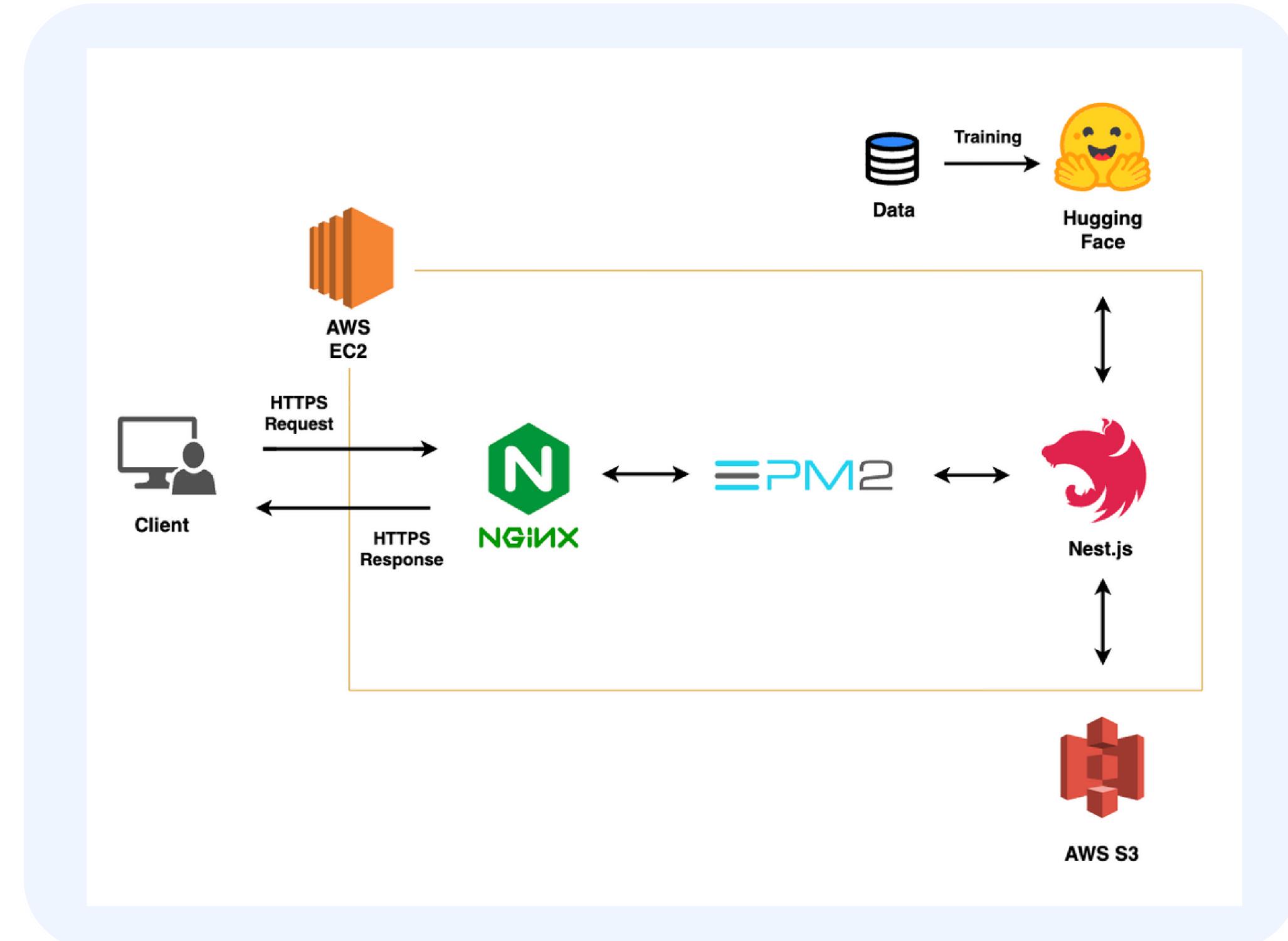
IDE : Visual Studio Code

Deploy : AWS EC2, Nginx, PM2

Additional : AWS S3, Hugging Face

Collection : Git, Github

Backend Architecture



API 설계 |

'/upload/file'

request: 사진 file

response: 캡셔닝 결과

'/question'

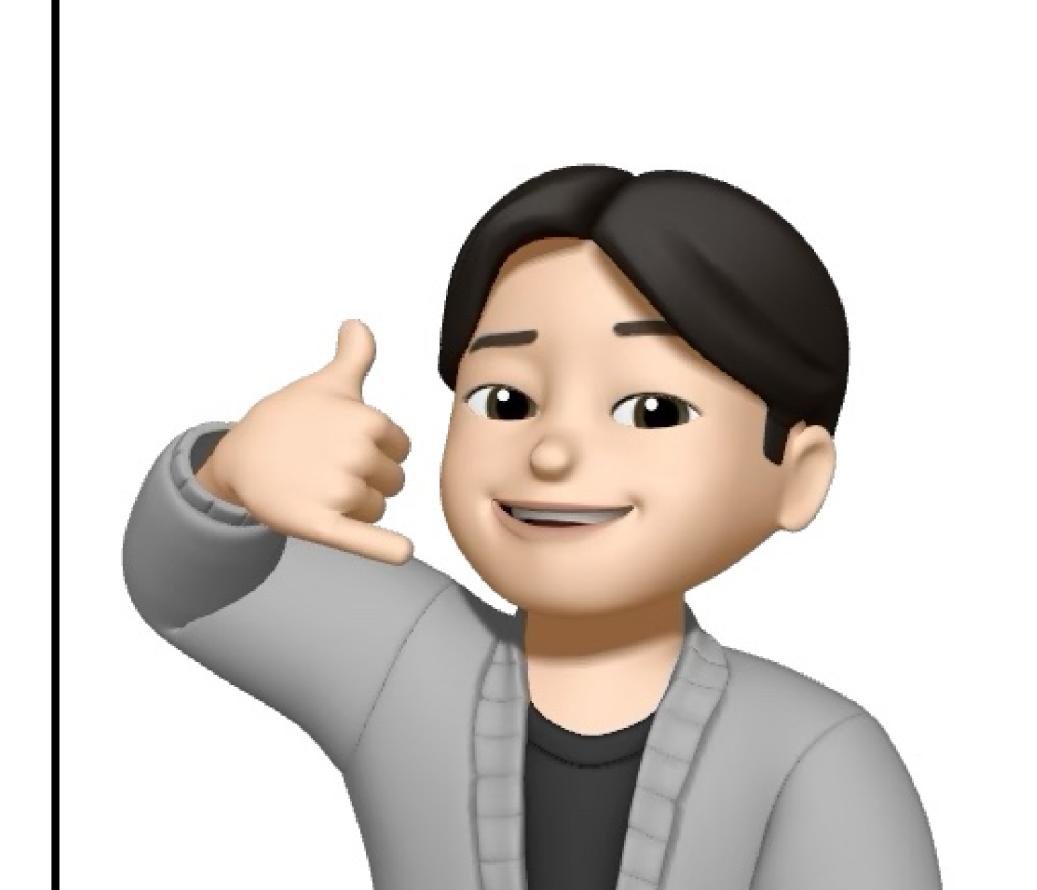
request: 사진URL, 질문

response: 사진에 대한 질의 결과

=> 사진 file을 받아서 AWS S3에 이미지를 업로드
후, 캡셔닝 결과와 사진 URL을 응답으로 리턴

=> 사진 URL과 질문을 받고 해당하는
visualQA의 결과를 응답으로 리턴

Role |

		
고형민	전수현	정도영
프론트엔드	AI	백엔드

감사합니다.