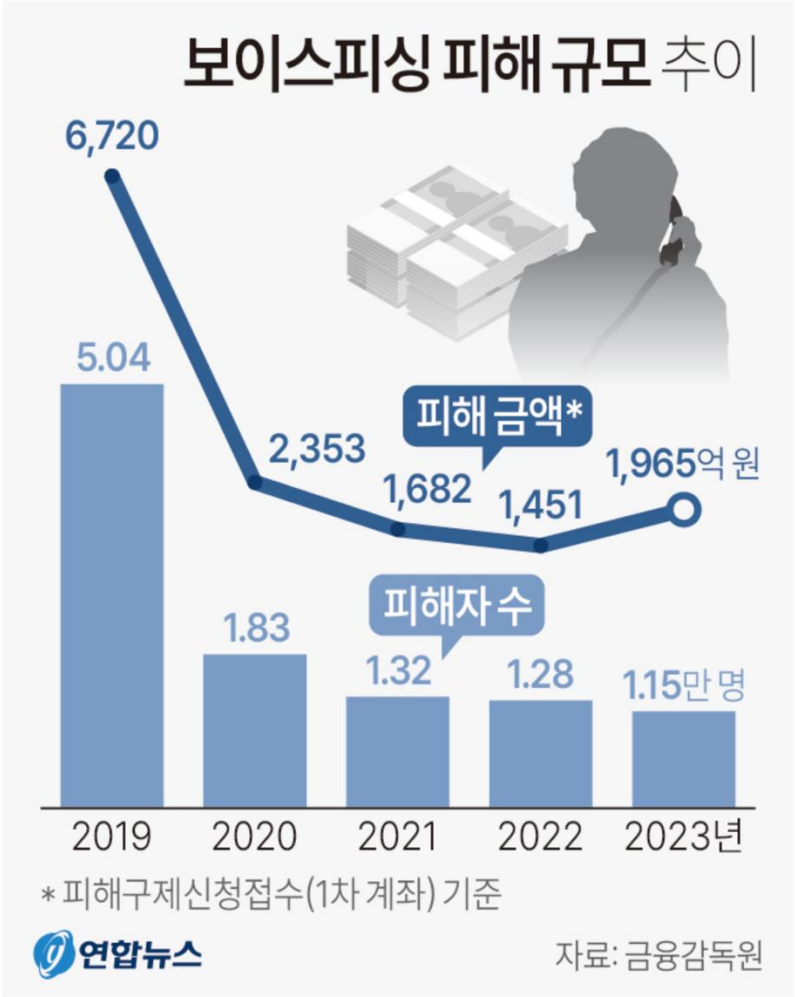
Phishing Guard
(종합 피싱 탐지 서비스)

CLAY
안준성 조문선 홍요한

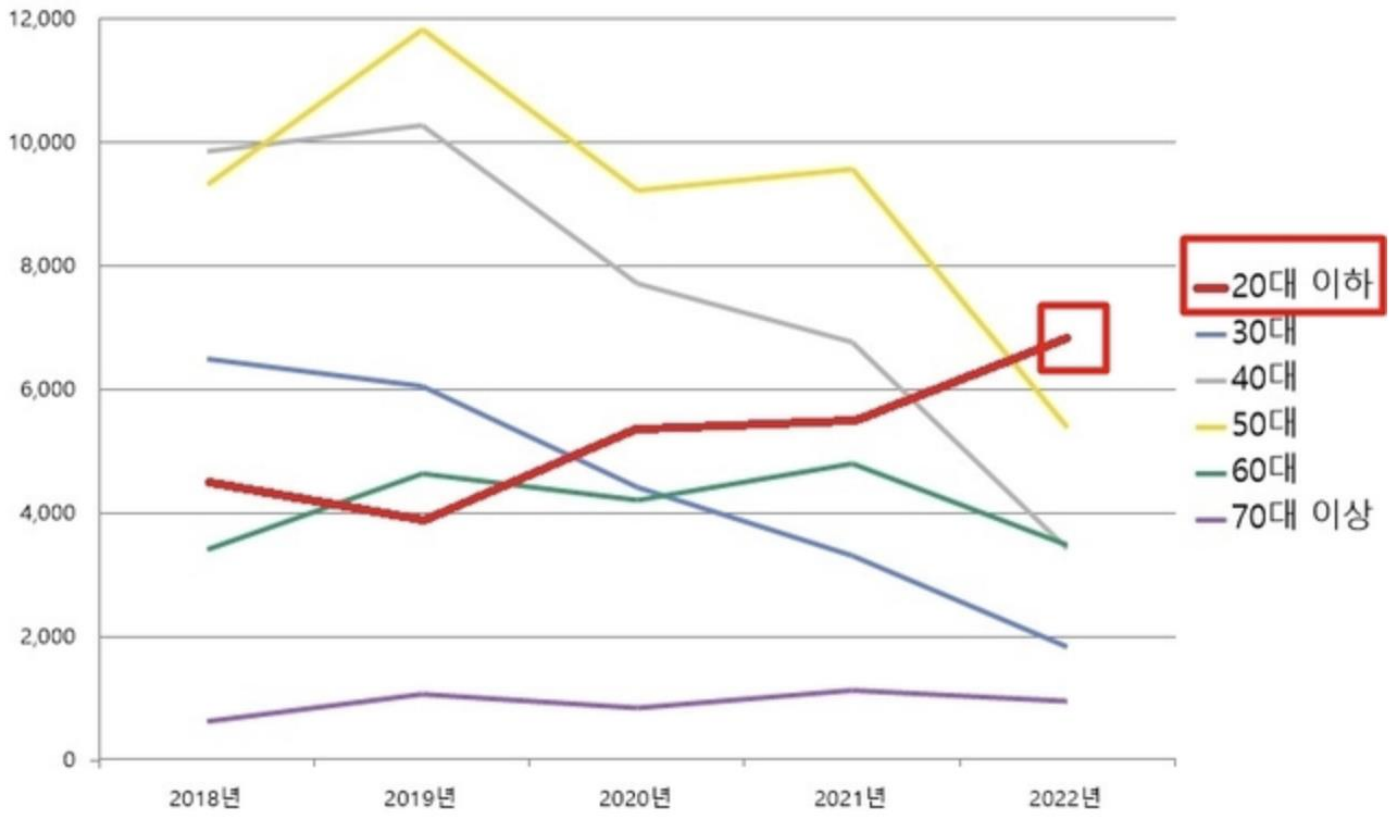
Part 1,

문제점

후후(whowho) 쓰면 되던데?



출처: 연합뉴스



최근 5년간 보이스피싱 피해자 연령별 현황. 경찰청 제공

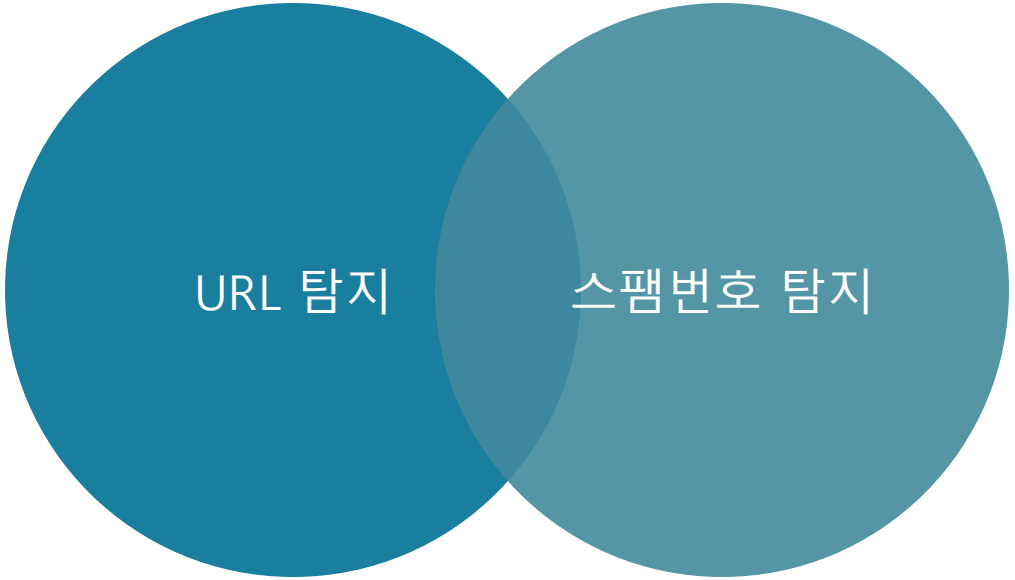
출처: 국민일보 "MZ가 보이스피싱에 속는다고? ..."

Part 1,

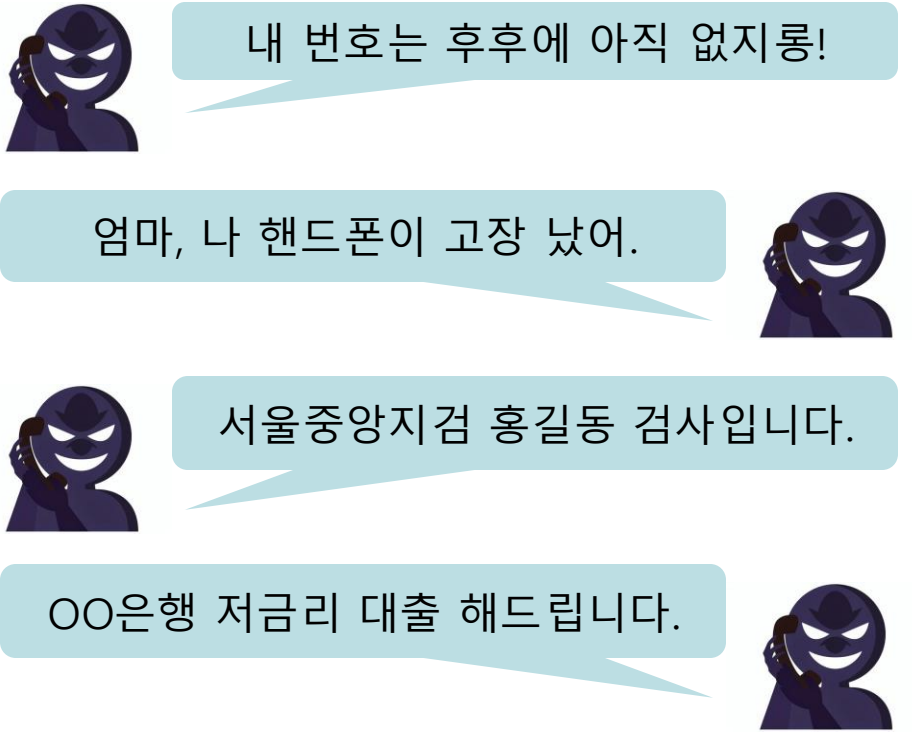
문제점

후후(whowho)만으로는 안돼!

후후(whowho)가 뭐 해주는데?



그럼 얘네는 어떡해?



Part 1,

문제점

어.. 이거 내 목소린데?

■ 대포폰 적발 횟수 (단위 : 건)

■ 국내외 주요 딥보이스 피해 사례

- 2019. 3 영국 기업, 고위 임원 목소리에 속아 22만 유로(약 3억2369만원) 송금
- 2020 UAE 은행 지점장, 딥보이스에 속아 3500만 달러(약 487억원) 송금
- 2022. 3 질렌스키 우크라이나 대통령 항복선언 가짜 영상 논란

출처: 더스쿠프 "AI로 내 목소리 훔치는 '보이스피싱 사기꾼'..."

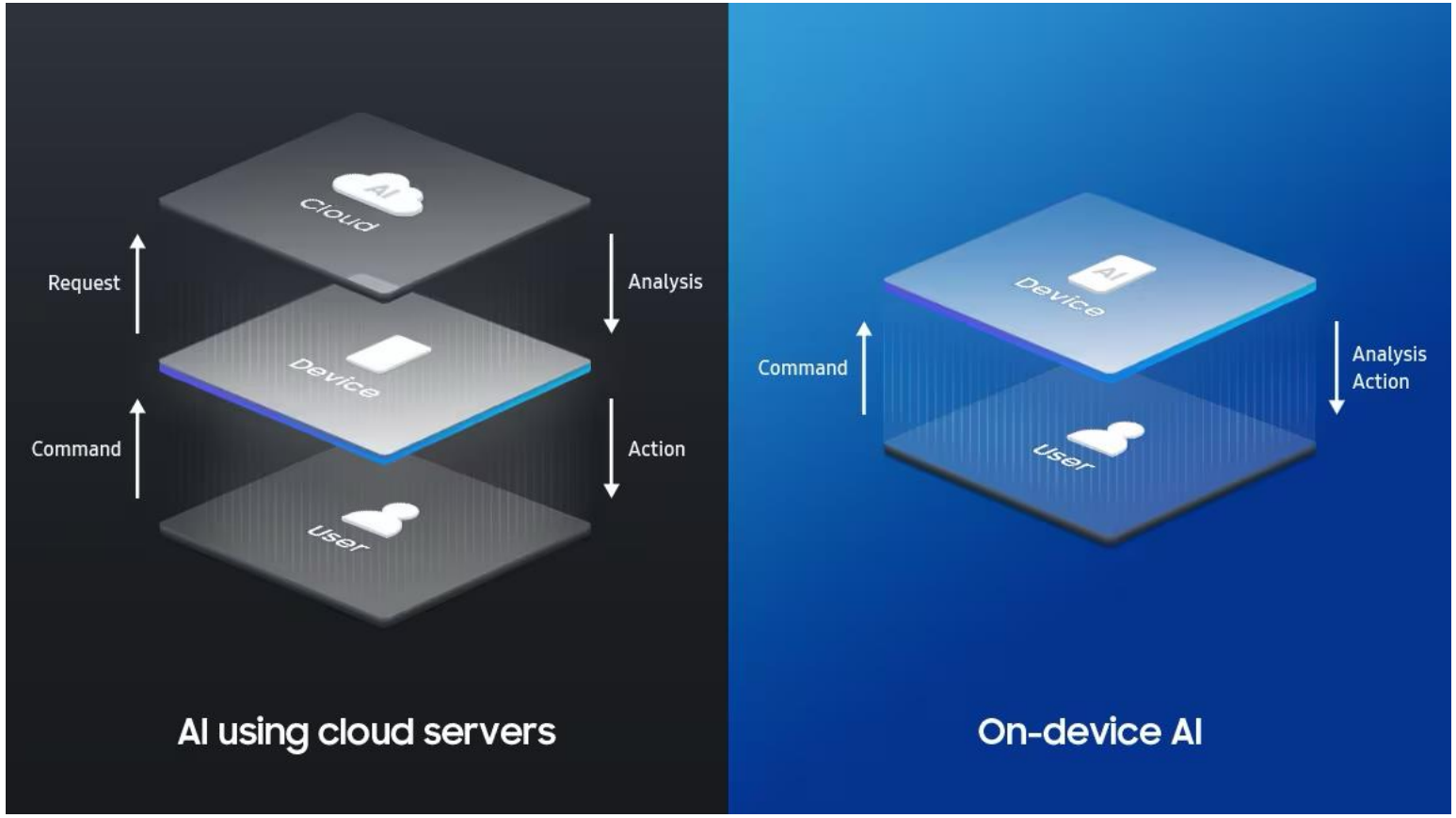
대포폰 + 감정까지 표현하는 딥보이스 조합은 알고도 당할 걸?

*"어떻게 하면 피싱을 **근절**시킬 수 있을까?"*

Part 2,

Phishing Guard

우리는 실시간으로 알려줄게!



출처: 삼성 공식 홈페이지

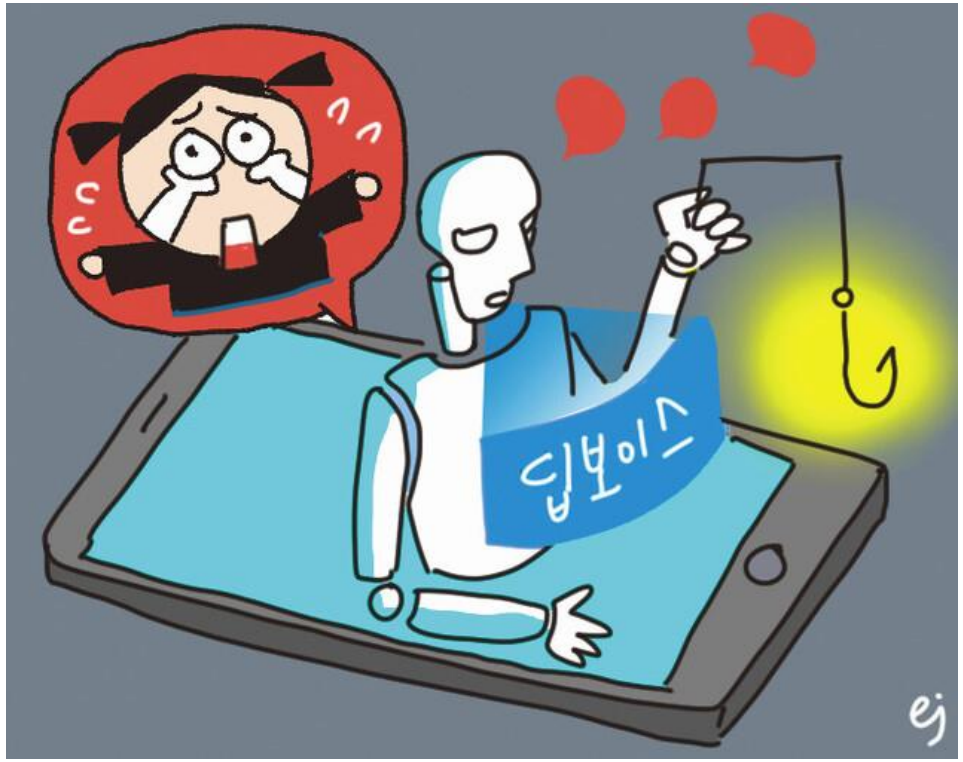
스미싱, 보이스피싱 탐지 AI를 On-device형태로 탑재한다면?

↳ 보이스피싱을 당하고 있는 사용자에게 실시간으로 경고 가능!

Part 2,

Phishing Guard

우리는 이런 것도 해 준다?



딥보이스 탐지 AI도 On-device형태로 탑재,

상대방의 음성이 딥보이스로 생성된 음성이라면 사용자에게 실시간으로 경고!

개발 과정과 발전가능성

개발 과정

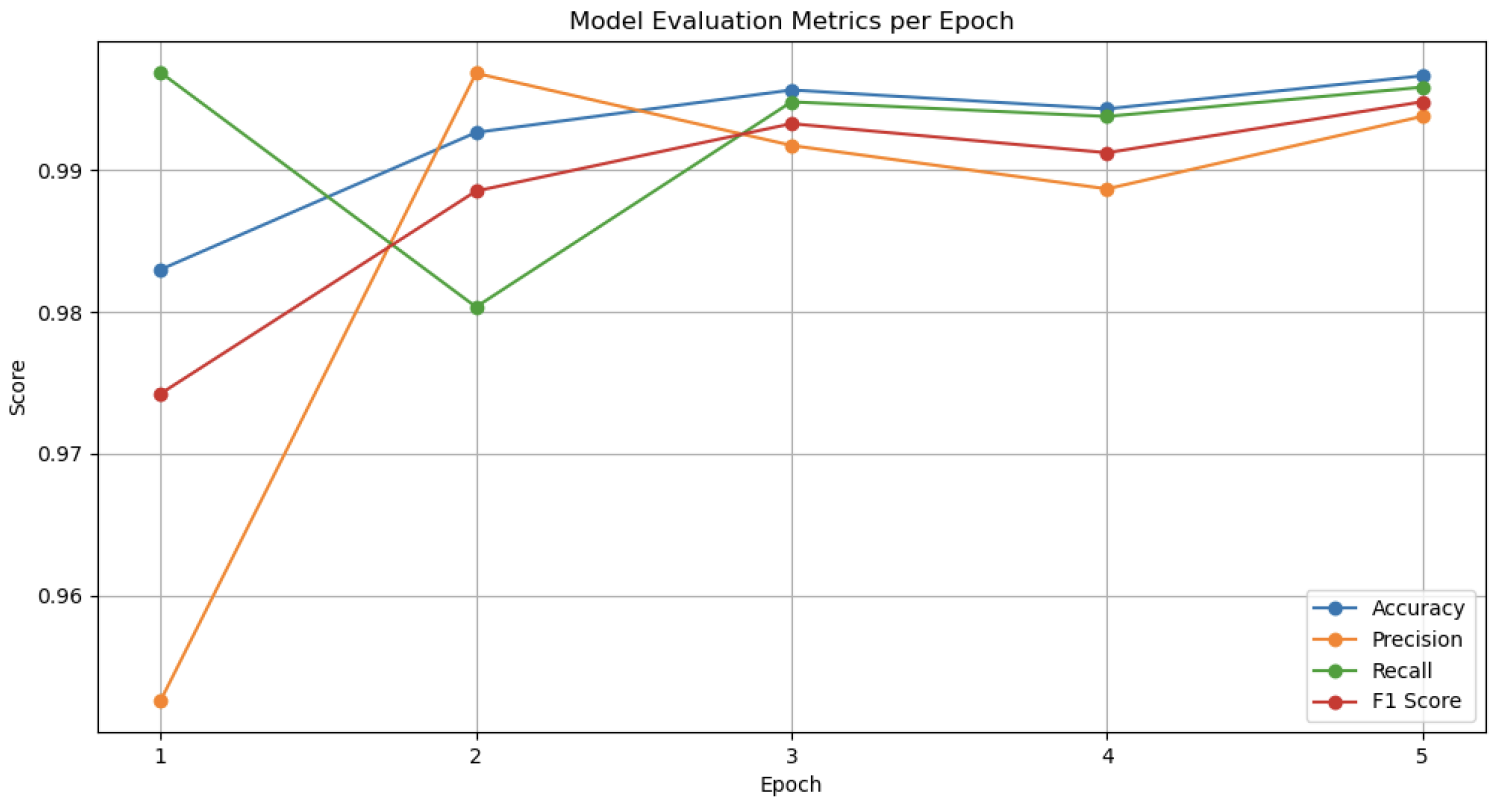
AI를 만들라고요..? 제가요?

보이스피싱, 스미싱 탐지 AI

1. K-means Clustering을 활용한 유사도 측정
   - 학습데이터셋 부족으로 인한 낮은 정확도

2. 벡터화를 통한 Cosine Similarity 계산
   - 학습 데이터와 실제 데이터 간의 데이터량 차이로 인한 낮은 유사도

3. GRU모델 KoBERT 기반 지도 학습
   - 라벨링된 데이터로 반복 학습하여 높은 정확도의 모델 구현
   - GRU: 인공 신경망의 일종, 사람의 뇌 속 뉴런의 작용을 본떠 패턴 구성



Part 3,

개발 과정

밤샘의 결과물..!

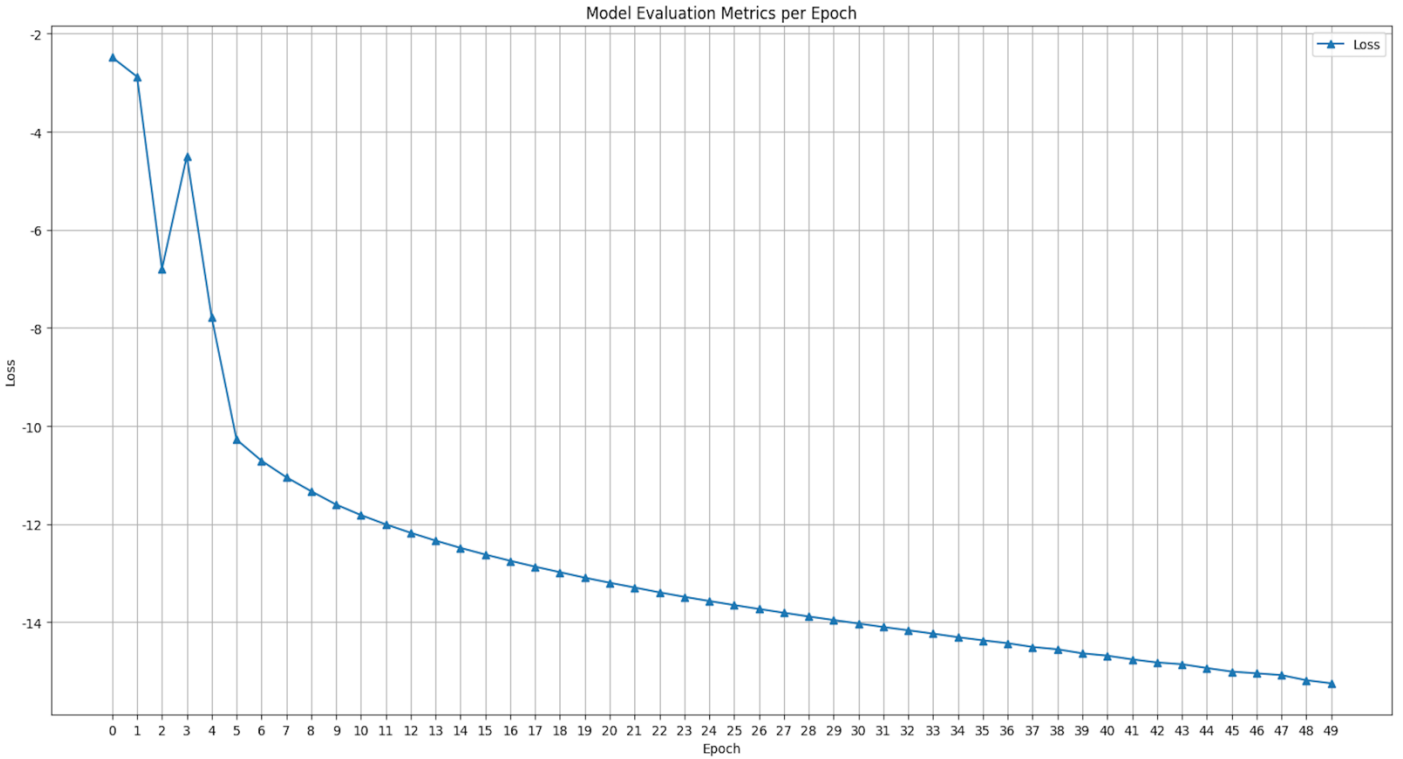
딥보이스 탐지 AI

1. 음성 특징 추출 기법 MFCC와 Mel-Spectrogram
   - MFCC: 음성 신호에서 추출된 주파수 특징을 수치화
   - Mel-Spectrogram: 사람의 음성 주파수 정보를 표현한 그래프

2. CNN, 넌 누구냐?
   - 인간의 시신경 구조를 모방, 이미지나 영상 데이터를 처리할 때 쓰이는 딥러닝 모델

3. Keras 활용하여 CNN 모델 구현
   - Keras: 파이썬 딥러닝 라이브러리

4. 라벨링된 음성파일을 Mel-Spectrogram으로 전처리하여 지도학습
   - 반복학습을 통해 낮은 손실률(높은 성능)의 모델 구현

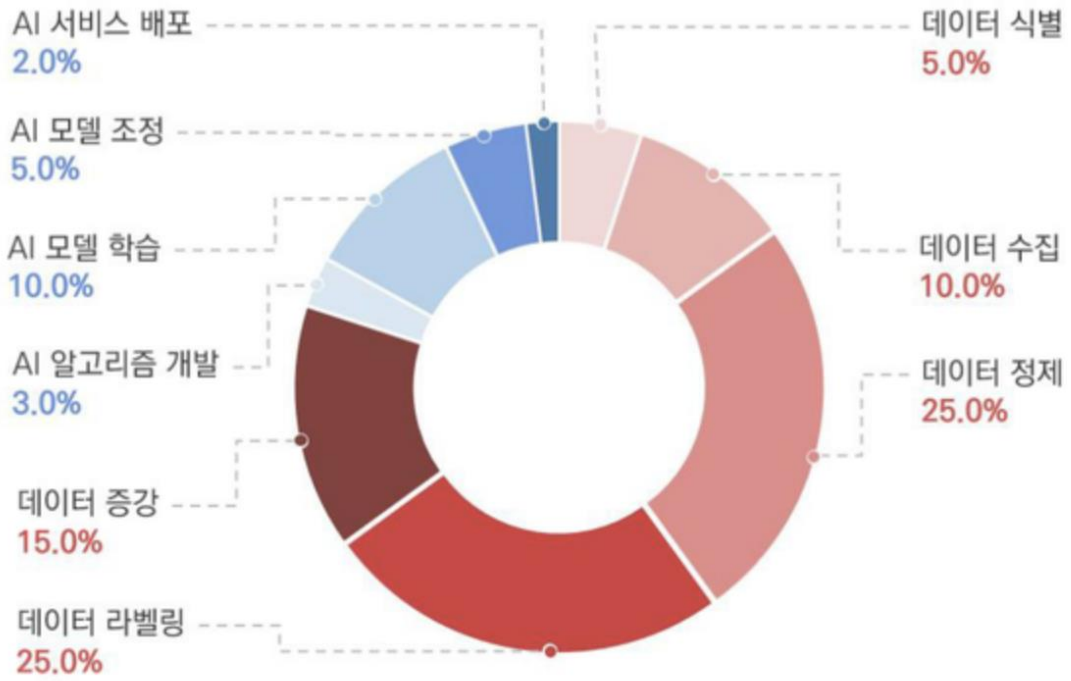


Part 4,

발전가능성

데이터가 부족해..

< AI 프로젝트에 소요되는 시간 비율>



※ 자료 : The Ultimate Guide to Data Labeling for ML, Cloudfactory, 재구성

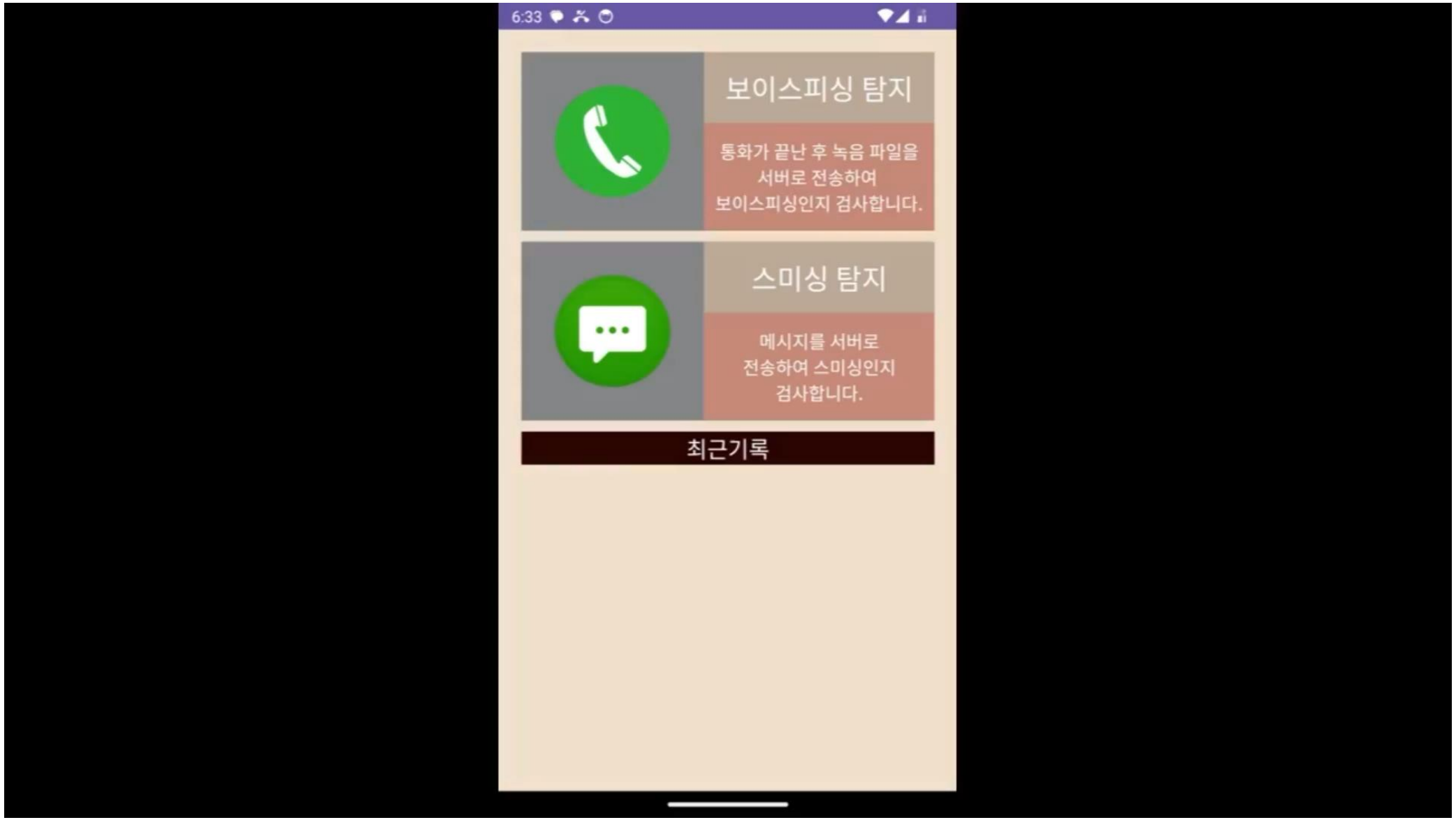
머신러닝의 핵심은 양질의 데이터!

But, 개인이 이를 구하는 것은 너무나도 힘든 일..

Part 5,

시연영상

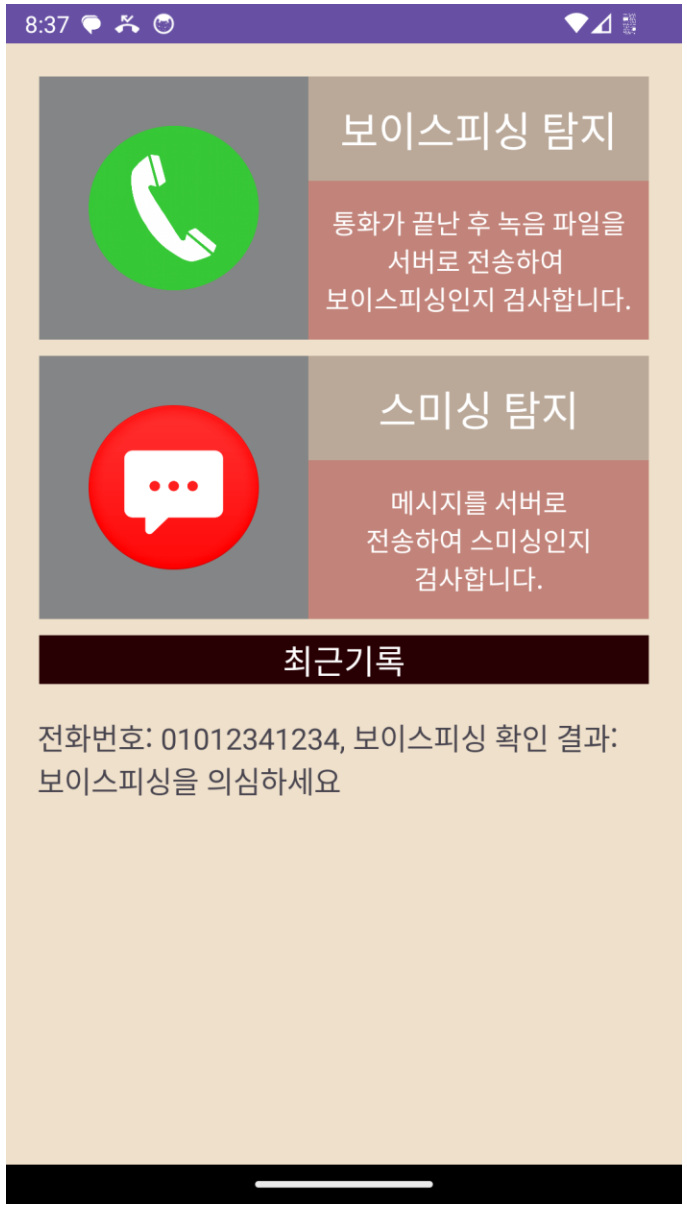
노력의 결실



Part 5,

시연영상

노력의 결실



" 안전한 대한민국을 만들기 위한 일환으로
저희를 초대해 주셔서 감사합니다."