2025년 전남대학교 소프트웨어중심대학사업
소·중·대 산학협력프로젝트(캡스톤디자인) 결과보고서

<table>
<tr><td colspan="2">프로젝트명</td><td colspan="6">AI 기반 음성 진위 판별 시스템 – VoiceGuard –</td></tr>
<tr><td colspan="2">Github url 주소</td><td colspan="6">https://github.com/shaddo82/AI-.git</td></tr>
<tr><td colspan="2">팀 명</td><td colspan="3">deepfakeR</td><td>과제수행기간</td><td colspan="2">2025. 9. 24. ~ 12. 19.</td></tr>
<tr><td colspan="2">지도교수</td><td>학 과</td><td colspan="2">인공지능학부</td><td>성 명</td><td colspan="2">조영준</td></tr>
<tr><td rowspan="8">프로젝트 수행인원
(※팀장은 첫줄에 기입)</td><td colspan="2">이 름</td><td colspan="2">학과(부·복수전공)</td><td>학년</td><td>학번</td><td>연락처(HP)</td><td>E-Mail</td></tr>
<tr><td>팀장</td><td>황재민</td><td colspan="2">인공지능전공</td><td>3</td><td>214258</td><td>010-2641-1791</td><td>ghkd1791@gmail.com</td></tr>
<tr><td rowspan="5">팀원</td><td>김병훈</td><td colspan="2">소프트웨어전공</td><td>3</td><td>214606</td><td>010-9562-5440</td><td>byounghun02@gmail.com</td></tr>
<tr><td></td><td colspan="2"></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="2"></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="2"></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="2"></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">참여 기업</td><td colspan="2">기업명</td><td colspan="2">멘토명</td><td>직위</td><td>연락처(HP)</td><td>E-Mail</td></tr>
<tr><td colspan="2">인사이트에듀(주)</td><td colspan="2">주현웅</td><td>대표</td><td>010-9747-1427</td><td>insighteducorp@naver.com</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td><td></td><td></td><td></td></tr>
</table>

위와 같이 2025년 전남대학교 소프트웨어중심대학사업
산학협력프로젝트 지원 프로그램 결과보고서를 제출합니다.

2025년 12월 19 일

신청자명(대표학생) : 황재민 (인)

지도교수 : 조영준

전남대학교 소프트웨어중심대학사업단장 귀하

산학협력프로젝트(캡스톤디자인) 결과보고서(요약)

프로젝트명	AI 기반 음성 진위 판별 시스템 – VoiceGuard –		
수행기간	2025. 9. 24. ~ 12. 19.	소요예산	349,318원
소요예산 세부내역	회의비 296,000원, SW활용비 53,318원		

참여인원	구분	인원수	성명(모두 기재)
	교수	1	조영준
	석박사과정	0	
	학부생	2	황재민, 김병훈
	기업체	1	주현웅
	계	4	

추진배경

최근 AI 음성 합성(Text-to-Speech, TTS) 기술의 발전으로 실제 사람의 음성과 구분이 어려운 합성 음성이 대량 생성되고 있으며, 이를 악용한 보이스피싱 및 로보콜 범죄가 증가하고 있음

기존 보이스피싱 대응 방식은 사후 대응에 집중되어 있어, 통화 중 음성의 진위를 자동으로 판별할 수 있는 예방 중심 기술의 필요성이 제기됨

목표 및 내용

실제 음성과 AI 기반 합성 음성을 구분할 수 있는 음성 진위 판별 모델을 구축하는 것을 목표로 하였으며, Wav2Vec2 기반 딥러닝 모델을 활용하여 음성 데이터를 학습하고 분석하였다. 학습된 모델은 실제 음성(orig)과 인공지능 기반 합성 음성(tts), 통신 환경을 고려한 합성 음성(tts_gsm)을 구분할 수 있도록 3-class 분류 구조로 설계되었다. 또한 실제 통화 환경에서 발생할 수 있는 음질 저하와 전송 특성을 반영하기 위해 음성 데이터 전처리 과정을 적용하였다.

아울러 음성 진위 판별 모델을 서버 환경에 배포하여 추론이 가능하도록 구성하고, 서버 기반 추론 시스템과 모바일 애플리케이션을 연동하는 구조를 구현하였다. 이를 통해 모바일 환경에서 수집된 음성 데이터를 서버로 전송하여 분석한 뒤, 판별 결과를 다시 애플리케이션으로 전달하는 전체 동작 흐름을 완성하였다. 최종적으로 사용자가 음성 진위 여부를 직관적으로 확인할 수 있도록 Android 애플리케이션을 개발하였으며, 음성 기반 보이스피싱 및 로보콜과 같은 범죄 상황에서 활용 가능한 실용적인 음성 진위 판별 시스템을 구현하였다.

기대효과

보이스피싱 및 로보콜과 같은 음성 기반 범죄를 사전에 인지하고 예방할 수 있는 기술적 기반 마련
모바일 환경에서 활용 가능한 음성 보안 서비스로 확장 가능

향후 금융, 공공, 통신 등 다양한 분야에 적용 가능한 음성 진위 판별 기술 확보
AI 음성 합성 악용에 대응하는 실용적 보안 솔루션으로의 활용 기대

1. 프로젝트 개요

프로젝트명	**AI 기반 음성 진위 판별 시스템 – VoiceGuard –**
주제영역	□ 생활 □ 업무 □ 공공/교통 □ 금융/핀테크 □ 의료 □ 교육 □ 유통/쇼핑 □ 엔터테인먼트
기술분야	□ IoT ☑ 모바일 □ 데스트톱 SW ☑ 인공지능 □ 보안 □ 가상현실 □ 빅데이터 □ 자동제어기술 □ 블록체인 □ 영상처리 □ 기타( )
성과목표	□ 논문게재 및 포스터발표 □ 앱등록 □ 프로그램등록 □ 특허 □ 기술이전 □ 실용화 □ 공모전(*공모전명*) □ 기타( )

2. 프로젝트 추진배경

최근 보이스피싱 범죄는 인공지능 음성 합성 기술을 활용한 범죄 형태로 진화하고 있다. 특히 AI 기반 음성 합성(Text-to-Speech, TTS) 기술의 발전으로 인해 실제 사람의 음성과 구분하기 어려운 합성 음성이 생성 가능해지면서, 음성 기반 범죄의 위협성은 더욱 증가하고 있다.

이러한 기술이 범죄에 악용된 대표적인 사례가 로보콜(Robocall)이다. 로보콜은 사전에 생성된 합성 음성을 이용해 다수의 사용자에게 자동으로 전화를 발신하는 방식으로, 짧은 시간 내에 대규모 피해를 유발할 수 있다. 로보콜에 사용되는 음성은 실제 사람의 음성과 유사한 특성을 가지기 때문에, 일반 사용자가 통화 중 이를 즉각적으로 판별하는 것은 현실적으로 어렵다.

이와 같은 배경에서, 실제 통화 환경을 고려한 음성 진위 판별 기술의 필요성이 점차 대두되고 있다.

본 프로젝트는 이러한 문제의식에서 출발하여, 보이스피싱 유형 중 하나인 로보콜을 사전에 탐지하고 예방할 수 있는 모바일 애플리케이션 VoiceGuard를 개발하는 것을 최종 목표로 한다. 이를 위해 통화 중 수집되는 음성을 분석하여 해당 음성이 실제 사람의 음성인지, 또는 인공지능 기반으로 생성된 합성 음성인지를 자동으로 판별하는 시스템을 설계하였다.

이를 바탕으로 Wav2Vec2 기반 음성 진위 판별 모델과 서버 기반 추론 시스템, Android 모바일 애플리케이션을 연동한 구조를 구현함으로써, 사용자에게 보이스피싱 위험을 사전에 인지시킬 수 있는 예방 중심의 보안 애플리케이션을 제안한다.

3. 프로젝트(주제) 목표 및 내용

본 프로젝트의 최종 목표는 AI 기반 음성 합성(TTS)과 실제 음성을 자동으로 구분할 수 있는 음성 진위 판별 시스템을 구축하고, 이를 모바일 환경에서 활용 가능한 형태로 구현하는 것이다.
이를 위해 다음과 같은 세부 목표를 설정하였다.
첫째, 실제 음성(orig), TTS 음성(tts), 통신 환경을 고려한 TTS 음성(tts_gsm)을 구분하는 3-class 음성 분류 모델을 학습하였다.
둘째, Wav2Vec2 기반의 딥러닝 모델을 활용하여 높은 정확도의 음성 진위 판별 성능을 확보하였다.
셋째, 학습된 모델을 서버 환경에 배포하고, Android 애플리케이션과 연동하여 사용자에게 판별 결과를 제공하는 시스템을 구현하였다.
최종적으로 본 프로젝트를 통해 음성 기반 보이스피싱을 사전에 탐지하고 사용자에게 경고할 수 있는 VoiceGuard 시스템을 완성하였다.

4. 시스템 구성 및 내용

본 프로젝트에서 제안하는 **VoiceGuard 앱**은 보이스피싱 유형 중 하나인 로보콜을 탐지하고 예방하기 위한 음성 진위 판별 시스템이다. 시스템은 크게 모바일 애플리케이션, 서버 기반 추론 시스템, 그리고 딥러닝 음성 판별 모델의 세 가지 구성 요소로 이루어져 있다.

사용자의 음성은 모바일 애플리케이션을 통해 수집되며, 해당 음성은 서버로 전송되어 학습된 Wav2Vec2 기반 모델을 통해 분석된다. 이후 판별 결과는 다시 모바일 애플리케이션으로 전달되어, 사용자가 음성 진위 여부를 확인할 수 있도록 화면에 표시된다.

시스템 동작 흐름

1. **음성 수집 단계**
   Android 모바일 애플리케이션을 통해 사용자의 음성 입력을 실시간으로 수집한다. 수집된 음성은 일정 시간 단위로 처리되어, 음성 진위 판별을 위한 입력 데이터로 사용된다.

2. **서버 전송 단계**
   수집된 음성 데이터는 파일 또는 스트림 형태로 서버에 전송된다.

3. **추론 단계**
   서버는 수신한 음성 데이터를 전처리한 후, Wav2Vec2 기반 음성 진위 판별 모델을 통해 추론을 수행한다.

4. **결과 반환 단계**
   모델의 출력 결과(orig, tts, tts_gsm)는 서버에서 모바일 애플리케이션으로 전달된다.

5. **사용자 알림 단계**
   애플리케이션은 판별 결과에 따라 사용자에게 보이스피싱 위험 여부를 시각적 또는 텍스트 형태로 제공한다.

5. 프로젝트 결과물에 대한 기술

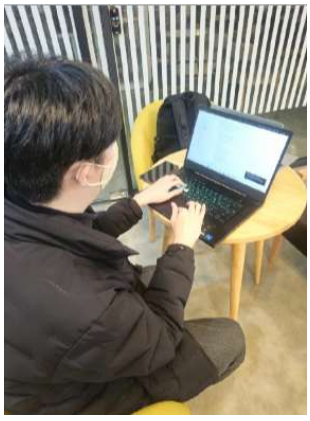
구분	기능정의	세부기능 설명
음성 판별 모델	음성 진위 판별	Wav2Vec2 기반 3-class 분류(orig / tts / tts_gsm)
서버 시스템	음성 분석 및 추론	Flask 기반 API 서버에서 음성 전처리 및 추론 수행
모바일 앱	사용자 인터페이스	통화 중 음성 분석 결과 시각적 표시
데이터 처리	음성 전처리	샘플링 통일, 길이 정규화, 통신 환경(GSM) 반영
결과 알림	보안 경고 제공	합성 음성 판별 시 사용자에게 메시지 제공

6. 프로젝트 진행내용

1) 참여인원 및 담당 역할

연번	소속학과	성명	수행역할 분담내용
1	인공지능전공	황재민	모델 학습 및 서버 구축
2	소프트웨어전공	김병훈	앱 개발

2) 회의 및 SW멘토링 진행

번호	일시/장소	회의/멘토링 내용(상세히 작성)	관련 사진
1	2025. 11. 24. (15:30~17:00) / 공대7호관 학생 라운지	Flask 기반 서버 구동 환경을 구축하여 모델 연동 테스트를 수행함. ● Wav2Vec2 기반 3-class 음성 진위 판별 모델을 서버 측에 배포하고 API 형태로 호출 가능하도록 구성. ● 서버에서 음성 파일을 업로드받아 전처리(16kHz 리샘플링, 고정 길이 변환)를 수행한 뒤 모델을 통해 orig / tts / tts_gsm 분류가 정상적으로 수행됨. ● 모델 inference 결과가 JSON 형태로 서버에서 정상 반환되는 것을 확인함.	
2	2025. 12. 03 (13:00~14:00)/ 진리관 203호	● 서버에서 수신한 판별 결과(원본/TTS/TTS_GSM 및 확률)를 앱 내부 UI에 실시간으로 표시하는 구조로 설계함. ● 게이지바, 색상 변화, 텍스트 표시 등 결과 시각화 방식 중 사용자 이해도가 높은 형태를 우선 적용하기로 결정함. ● 추후 로그 저장 기능(판별 결과 및 시간 기록)을 추가하여 사용자 편의성 개선 방향도 논의함.	

3	2025. 12. 11 (10:00~11:30)/ AI융합대학 3층 라운지	최종 보고서 구성 방향 정리 • 보고서는 문제 정의 → 데이터 구성 → 모델 설계 → 실험 결과 → 앱 구현 → 결론 및 한계 흐름으로 구성하기로 함. • 중간 발표 피드백을 반영하여, 결과 중심 설명을 강화하고 청중이 이해하기 쉬운 구조를 우선 적용하기로 결정함. • 음성 비교(원본/TTS/TTS_GSM), 스펙트로그램 시각화, 모델 성능 지표 등 시각 자료를 보고서에 적극 포함하기로 함. 모델 구조 및 학습 과정은 기술적 상세 대신 핵심 원리와 적용 의의 중심으로 간결하게 정리하기로 함.	

7. 프로젝트 세부일정 및 내용

No.	작업 내용	9월				10월				11월				12월				담당자	비고
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4		
1	프로젝트 계획서 작성																	김병훈	
2	데이터 수집 및 모델 학습																	황재민	
3	서버 구축 및 모델 연동																	황재민	
4	앱 UI 설계 및 구현																	김병훈	
5	통합 테스트 및 보완																	전원	
6	결과보고서 작성 및 제출																	전원	

8. 결과물에 대한 향후 활용계획

본 프로젝트에서 확인된 한계점을 바탕으로, 향후 다음과 같은 방향으로의 확장 및 개선이 가능하다.

첫째, 잡음 및 음질 저하 환경에 대한 데이터 확장과 전처리 기법 개선을 통해 실제 환경에서도 안정적인 음성 진위 판별 성능을 확보할 필요가 있다.

둘째, Android 플랫폼의 통화 서비스 권한 제약을 고려한 시스템 연계 방식에 대한 연구가 필요하며, 향후 운영체제 정책 변화나 통신사 API 연계를 통해 실제 통화 중 음성 판별 기능으로 확장할 수 있을 것으로 기대된다.

9. 참고자료 및 문헌