팀프로젝트 제안서

- 모의기획안 -

2025 년 5 월 2 일

이름: 최재영, 황선영, 김성민, 오원창, 운드랄

소속 학과: 컴퓨터공학과

학번: 21622137, 22412140, 22112110, 22111138, 22411924

팀명: 8 조

제안서

제안명	**인공지능 기반 보이스피싱 탐지 앱(PhishingBlock)**
제안내용	국내 보이스피싱 피해는 전 연령층에서 발생하고 있지만, 특히 디지털 취약계층이 주요 피해 대상이 되고 있습니다. **50 대 이상 중장년층 및 노년층이 전체 피해자의 약 49.9%**를 넘어서고 있으며, 이외에도 장애인, 디지털 기기에 익숙하지 않은 사람들까지 취약계층의 피해 사례가 꾸준히 증가하고 있습니다. **이러한 상황에서 에이닷, 후스콜과 같은 어플의 등장으로 보이스피싱 피해는 한동안 감소** 추세를 보였으나, 최근 **스푸핑(Spoofing)**과 같은 첨단 기술을 활용한 수법이 더욱 악랄해지면서 피해가 다시 급증하고 있습니다. **스푸핑은 발신자 정보를 조작**하여 금융기관이나 공공기관의 전화번호로 위장하는 기술로, 피해자들이 실제 기관의 연락으로 오인하게 만들어 피해를 키우고 있습니다. 그 결과 **지난해 1 인당 평균 피해액은 4,100 만원, 전체 피해액은 8,545 억 원으로 모두 역대 최고치를 기록**했습니다.  피싱블락은 이러한 문제를 해결하기 위해 직관적인 인터페이스로 디지털 취약계층도 손쉽게 보이스피싱 위험을 감지하고 대응할 수 있도록 설계되었습니다. 피싱블락은 **직관적인 인터페이스**로 디지털 취약계층도 손쉽게 보이스피싱 위험을 감지하고 대응할 수 있도록 설계되었습니다. 디지털 취약계층을 보호하는 것은 사회적 책임이며, 이앱은 디지털 취약계층의 디지털 금융 안전을 지키는 필수적인 보안 솔루션이 될 것입니다.

목적 및 핵심기술	**보이스피싱 실시간 탐지** : 통화 중 의심스러운 대화 패턴과 내용을 실시간으로 감지하여 사용자에게 경고 **고령층 디지털 취약계층 보호** : 특히 디지털 기기 활용에 취약한 고령층 사용자를 보이스피싱으로부터 보호 **사용자 친화적 인터페이스** : 고령층도 쉽게 사용할 수 있는 직관적인 사용자 인터페이스 제공 **1. STT(Speech-to-Text) 기술** 통화 내용을 실시간으로 텍스트로 변환하는 음성 인식 기술 한국어에 최적화된 음성 인식 알고리즘 사용[1]) **2. 의심 키워드 및 패턴 탐지 기술** 보이스피싱에 자주 사용되는 의심 단어 및 문장 패턴 데이터베이스 구축 실시간 대화 내용과 의심 패턴 매칭을 통한 위험 신호 감지 금융 기관 사칭, 공공기관 사칭, 가족 사칭 등 사기 유형별 **3. 위험도 분석 및 경고 시스템** 의심 키워드 출현 빈도와 문맥 분석을 통한 위험도 점수화[2]) 위험도에 따른 단계별 경고(진동, 시각 알림, 음성 알림) 제공 고위험 상황 감지 시 사전 설정된 보호자에게 자동 알림 발송 **4. 사용자 친화적 인터페이스** 고령층을 위한 큰 글씨, 단순한 메뉴 구조, 높은 색상 대비의 UI 설계 음성 명령으로 앱 실행 및 제어 가능한 음성 인터페이스 시각 및 청각 장애를 가진 사용자를 위한 접근성 기능 강화 **5. 학습 및 개선 시스템** 새로운 보이스피싱 수법에 대한 정기적인 데이터베이스 업데이트 사용자의 개인 통화 패턴 학습을 통한 맞춤형 보호 제공
독창성 및 혁신성	최근 보이스피싱 수법은 점점 더 정교해지고 있으며, 특히 전화번호 스푸핑(Caller ID Spoofing), 즉 전화번호를 조작하여 공식기관이나 지인처럼 가장하는 수법이 급증하고 있습니다. 이러한 방식은 기존의 전화번호 기반 탐지 시스템(예: 후후, 후스콜)으로는 탐지가 어렵습니다. 실제로 이들 앱은 과거 신고 이력에 기반한 데이터베이스 중심의 방식이기 때문에, 새롭게 등장한 번호나 스푸핑된 번호를 실시간으로 식별하는 데 한계가 있습니다. 이에 비해 저희가 개발하는 인공지능 기반 보이스 탐지 앱은 통화 내용의 음성 데이터를 실시간으로 분석하여, 화자의 말투, 패턴, 키워드, 감정 상태 등을 종합적으로 판단합니다. 이를 통해 전화번호가 아닌 음성 기반의 AI 탐지 기술로 사기 가능성을 탐지하며, 기존 방식으로는 잡아내기

	힘든 새로운 보이스피싱 시도까지도 조기에 경고할 수 있습니다. 또한 대화 중 실시간으로 위험도를 점수화하거나, 사용자가 말한 문장에 대해 AI 가 즉시 판단하고 안내해주는 기능을 통해 보이스피싱 피해를 사전에 예방하는 능동적인 보안 솔루션을 제공합니다. 이는 전화번호 목록에 의존하는 기존 서비스와는 근본적으로 다른 접근 방식으로, 진화하는 범죄에 맞서는 새로운 패러다임이라 할 수 있습니다. **1. 실시간 통화 모니터링 및 분석 기술** 통화 중 실시간으로 음성을 텍스트로 변환하여 즉시 분석 의심 키워드나 패턴이 감지되면 통화 도중 즉시 경고 제공 사기꾼이 정보를 요구하거나 이체를 유도하는 순간 즉시 경고하여 피해 예방 **2. 고령층 특화 사용자 경험 설계** 고령층 사용자 테스트를 통해 개발된 직관적이고 단순한 인터페이스 한 번의 설치로 자동 작동하는 시스템으로 사용자 부담 최소화 경고 메시지를 쉬운 언어와 큰 글씨로 표시하여 즉각적인 이해 가능 **3. 진화하는 보이스피싱 패턴 대응 시스템** 사용자 피드백 시스템으로 오탐지 개선 및 새로운 패턴 학습 정기적인 보안 업데이트로 최신 사기 수법에 대한 대응력 유지 **4. 맞춤형 위험도 분석 기술** 통화 맥락과 패턴을 종합적으로 분석하는 고급 알고리즘 개인별 통화 패턴 학습을 통한 맞춤형 위험도 측정 단계별 경고 시스템으로 경고 피로도 감소 및 중요 경고 주목도 향상 **5. 전화번호 스푸핑 대응 능력** -Caller ID Spoofing(전화번호 위조) 방식에도 효과적인 탐지 -기존 시스템이 인식하지 못하는 신규 수법이나 낯선 번호에 대한 선제적 대응력
기대효과	**1. 고령층 금융 피해 예방** 디지털 취약계층의 금융 보안 강화로 사회안전망 구축

	후후, 후스쿨	피싱블락
탐지 방식	전화번호 기반 (신고 DB 매칭)	음성 기반 AI 실시간 분석
스푸핑 대응력전	전화번호가 조작되면 탐지 불가	전화번호가 위조되어도 통화 내용으로 탐지 가능
실시간 반응	없음(전화 수신 시 안내만)	통화 중 사기 의심 시 즉시 경고 제공
신규 보이스피싱 대응	신고된 번호만 인식 (신규 번호 탐지 어려움)	신고 이력 없어도 화자 내용으로 탐지 가능

	가족 간 금전적 갈등 및 스트레스 감소로 가족 관계 개선 **2. 사회적 비용 절감** 보이스피싱 범죄 수사 및 피해 복구를 위한 사회적 비용 절감 피해 발생 후 치료가 아닌 예방적 접근으로 효율적인 자원 배분 범죄 피해로 인한 고령층 정신적 트라우마 및 의료비용 감소 **3. 디지털 접근성 향상** 고령층의 디지털 기기 활용 자신감 상승으로 디지털 격차 완화 접근성 높은 인터페이스로 고령층의 모바일 서비스 활용도 증가 디지털 금융 서비스 이용률 증가로 고령층 경제 활동 촉진 **4. 범죄 예방 및 대응력 강화** 최신 보이스피싱 수법에 대한 빠른 정보 공유 및 대응 체계 구축 보이스피싱 성공률 저하로 범죄 시도 자체의 감소 유도 **기존 서비스는 통화 후 분석 또는 통화 중 녹음 데이터를 기반으로 한 번호 필터링 중심의 서비스** 입니다. 반면, 우리가 개발하는 서비스는 **통화 중 음성을 실시간으로 STT 변환하여 위험 키워드를 분석하고, 즉각적으로 경고를 제공하**는 방식입니다. 이를 통해 기존 방식으로는 탐지하기 어려웠던 지인 사칭, 공공기관 사칭, 악성 앱 설치 유도와 같은 보이스피싱 수법까지 실시간으로 파악할 수 있으며, **기존에 30~40%**에 달하는 전체 탐지율은 **기존 대비 약 1.5 배에서 최대 2.5 배까지 향상**될 것으로 기대됩니다.

[추가자료 #1] 주요 알고리즘

1. 한국어 음성인식 알고리즘

1) 형태소 기반 분석
한국어는 교착어로, 단어 하나가 어근 + 조사 + 어미 등으로 복잡하게 결합됨.
예: 먹었습니다→ 먹 (어간)+ 었 (과거)+ 습니다 (존대)
▶ **영어는 대부분 단어 단위(띄어쓰기 기준)**로 처리하지만, **한국어는 형태소(Morpheme) 단위**

2) 단어 경계 인식의 어려움
한국어는 띄어쓰기가 애매하고, 실제 말할 때 띄어쓰지 않음 → **음절 연속이 모호함.**
▶ 알고리즘이 **음향 모델외**에도 **언어 모델(Language Model)을 통해 문맥을 파악**해야 함.

3) 음운 변화 (연음, 축약, 탈락 등)
실생활 한국어 발음에는 많은 변화가 있음.
예: "같이 가자" → "가치 가자"처럼 발음됨.
▶ 이런 **음운 현상을 반영한 발음 사전(pronunciation dictionary)과 딥러닝 기반 음향 모델**

4) 언어 모델의 특화
한국어 문장 구조는 SOV (주어-목적어-동사) 구조이며, **동사가 마지막**
▶ 영어(주어-동사-목적어, SVO)와 달라서 **후행 맥락 기반의 언어 모델**이 더 중요함.

외부 API 사용시

기술	장점	단점
Azure Speech-to-Text	전화번호 기반 (신고 DB 매칭)	1.사용량 많으면 비용 발생 2.정확도는 중상급
Google Cloud Speech-to-Text	1.한국어 지원 2.다양한 입력(마이크,파일 등)지원	1.서버리스 환경에 연동하려면 추가 작업 필요 2.비용 주의

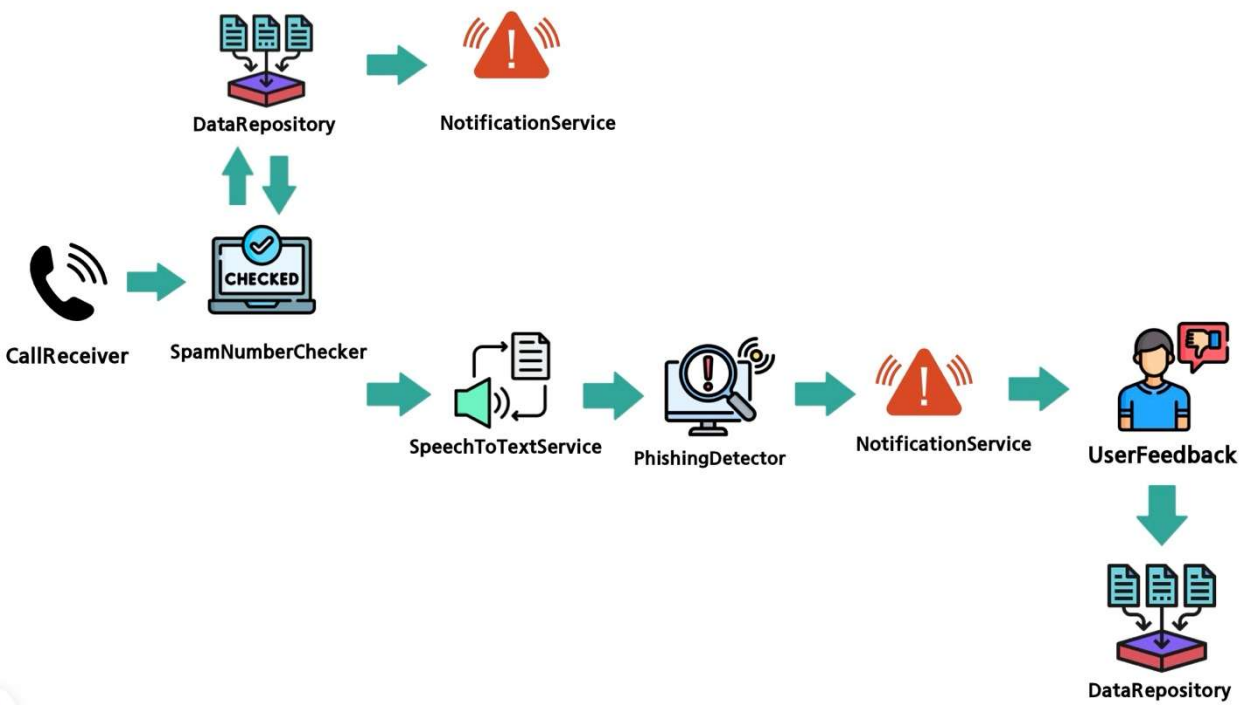
2. 위험도 점수화 알고리즘

1) 전처리 및 토크나이징 (형태소 기반)
예시 도구: [KoNLPy 의 Okt, Mecab] or [transformers 기반 BERT tokenizer]
-조사, 어미 제거
-명사, 동사, 고유명사 추출
-띄어쓰기 보정

2) 의심 키워드 탐지
미리 정의된 의심 단어 리스트 사용
예: 계좌, 입금, 공인인증서, 보이스피싱, 검찰, 금융, 압수수색등
각 단어에 가중치(weight)부여

3) 위험도 점수화
KoBERT / KoELECTRA 등의 모델로 문장을 벡터화 후
미리 학습된 "보이스피싱 문장"과의 유사도 비교 (Cosine Similarity)
특정 유사도 이상일 경우 가중치 점수 추가

[추가자료 #2] PhisingBlock 앱의 진행 개요도



1. user 에게 전화가 걸려온다.

→ **CallReceiver** 가 전화 수신 이벤트를 감지한다.

2. SYSTEM 은 수신된 번호를 DB 와 대조하여 스팸 여부를 확인한다.

→ **SpamNumberChecker** 가 **DataRepository** 와 통신해 스팸 여부를 조회한다.

3. 실시간 통화 중 음성을 분석하여 보이스피싱 여부를 판별한다.

→ **SpeechToTextService** 가 음성을 텍스트로 변환하고,

→ **PhishingDetector** 가 분석하여 위험 여부를 판단한다.

4. 스팸 또는 보이스피싱으로 판단되면 SYSTEM 이 사용자에게 경고를 보낸다.

→ **NotificationService** 가 오버레이 또는 알림으로 사용자에게 경고 메시지를 표시한다.

5. user 는 통화 종료 후 피드백을 남길 수 있다.

→ **UserFeedback** 이 신고 또는 평가 정보를 수집한다.

6. 피드백은 DB 에 저장되어 추후 판단에 활용된다.

→ **DataRepository** 가 피드백을 반영하고 데이터베이스를 업데이트한다.