

현장 검증단

2차년도
1차

보이스피싱 정보수집·가공 및 빅데이터 기반
수사지원시스템 개발



경찰대학 치안정책연구소

스마트치안지능센터

content



연구 내용

사업 개요, 연구개발 기관
보이스피싱 실제 데이터 구축 및 실행



연구개발 경과

음성 분석, 대화내용 분석, 수사기법 개발, 핵심기술 통합 및 API 개발, 수사지원 시스템 개발
연구 개발물 활용 구성도



KONAN 화면 정의

voice Fake / voice Note / voice Matching / case Matching
음성비교 국과수

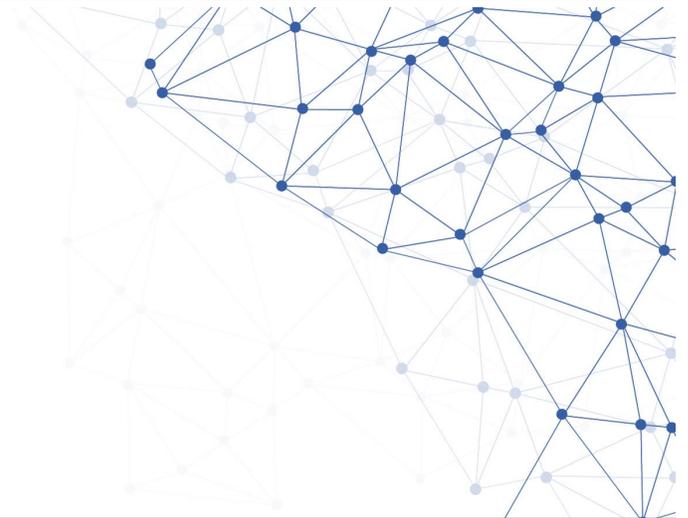
연구 내용

62억원 사업비 지출

8개 기관 참여

경찰대학이 주관하는 보이스피싱 수사 지원 연구개발

/01



보이스피싱 정보 수집·가공 및 빅데이터 기반 수사지원시스템 개발

사업 개요

과제기간

2022.04.01~2024.12.31

총 사업비

6,239,166,000원

총괄 책임자

스마트치안지능센터장 장광호



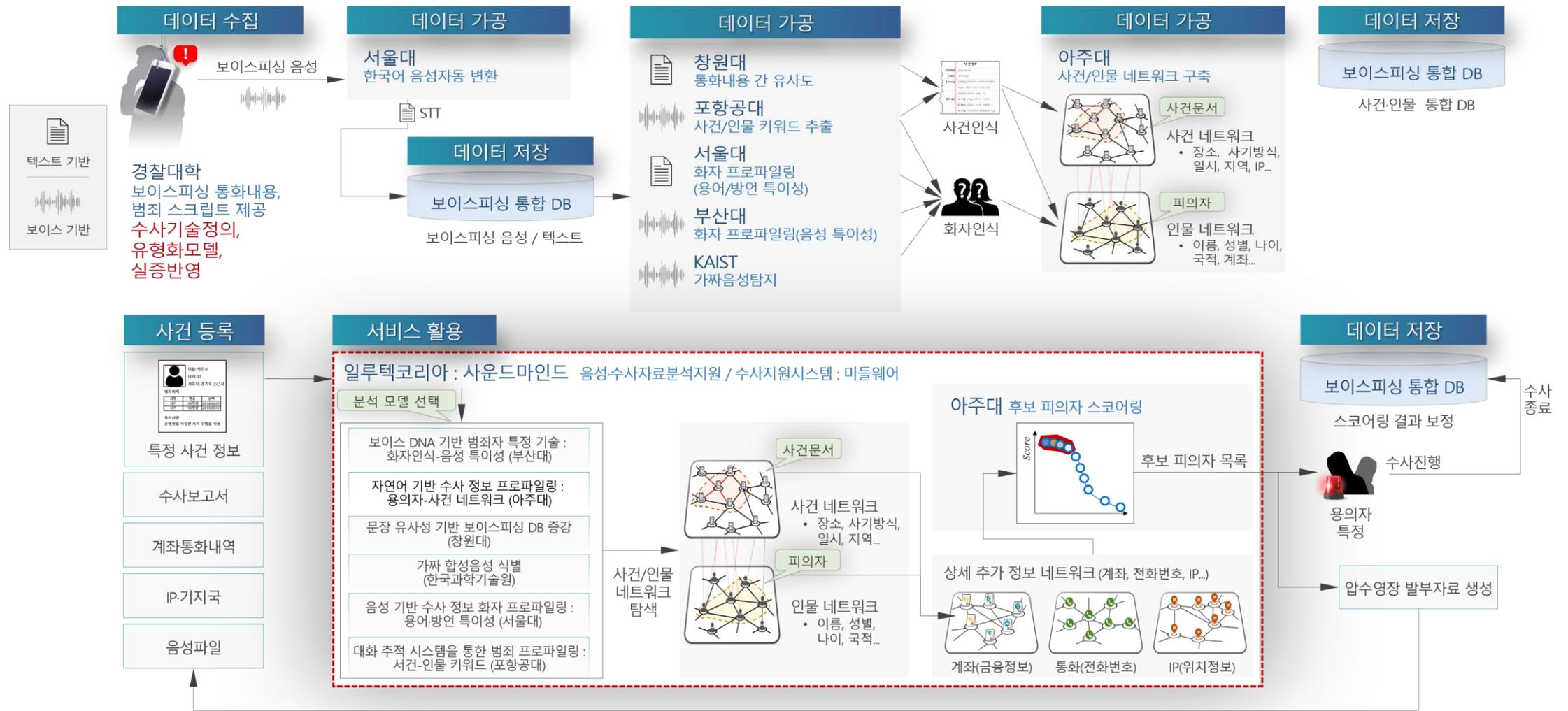
연구개발기관

주관연구개발기관
경찰대학

공동연구개발기관
서울대, 부산대, 카이스트, 창원대,
(주)일루텍코리아, (주)사운드마인드

위탁연구개발기관
포스텍

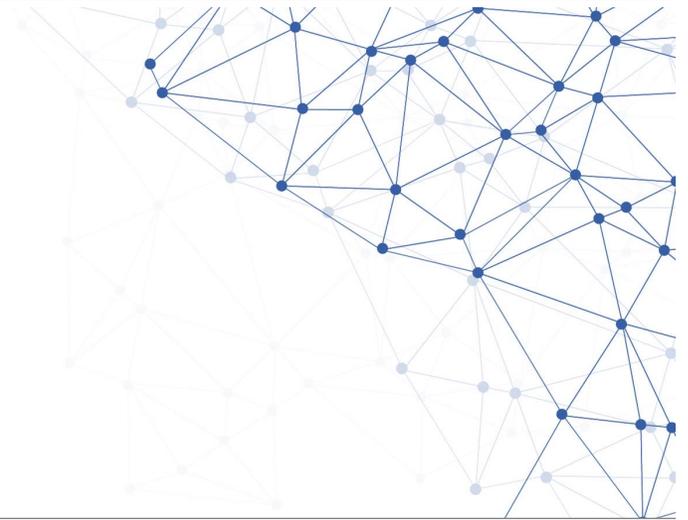
보이스피싱 실제 데이터 구축(수집, 가공, 저장) 및 실행(분석, 활용, 제공)



연구개발 경과

음성분석 / 대화내용 분석 / 수사기법 개발
핵심기술 통합 및 API 개발 / 수사지원 시스템 개발
연구 개발물 활용 구성도

/02

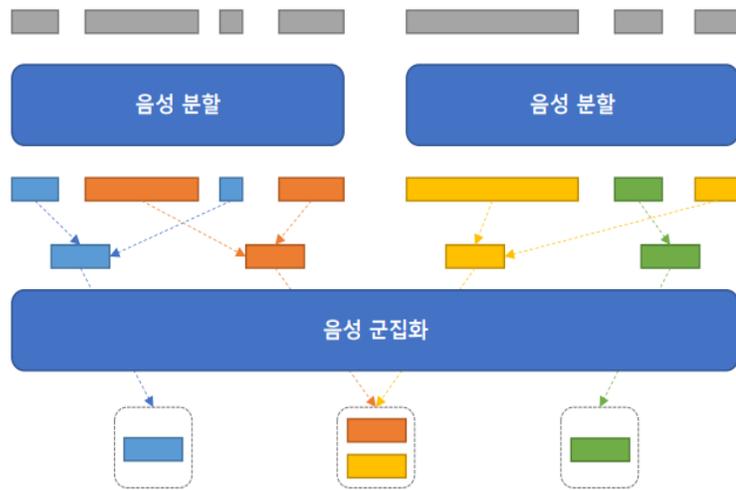


음성분석 [부산대학교]



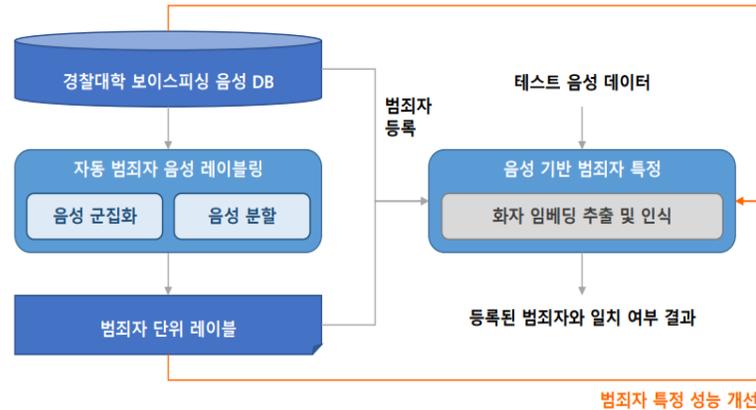
음성기반 범죄자 특정 기술 개발

자동 범죄자 음성 레이블 기술 고도화



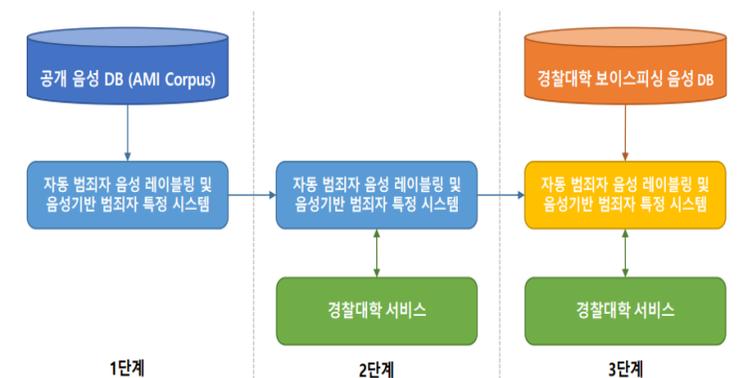
- 딥러닝 기반 음성 분할 및 음성 균집화 기술 적용
- 음성 분할 및 균집화 모듈 통합 최적화

음성기반 범죄자 특정 개선



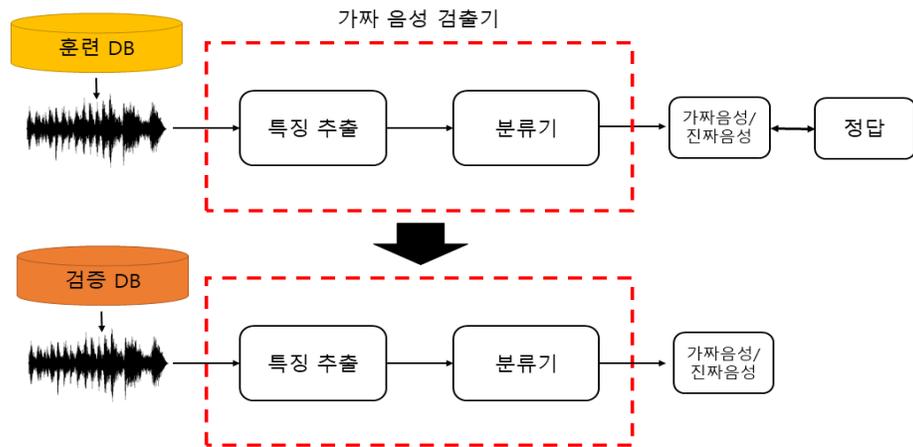
- 생성된 음성 레이블을 범죄자 등록 뿐만 아니라, 음성기반 범죄자 특정 성능을 개선하는데 활용
- 모델 미세조정, 도메인 적응, 유사도 정규화 적용

서비스 연동 및 최적화



- 1단계: 부산대에서 공개 데이터를 활용하여 기술 개발
- 2단계: 경찰대학 서비스와 연동
- 3단계: 경찰대학 음성 DB를 바탕으로 시스템 최적화

통화 환경의 AI 합성 음성 탐지 기술



한국어에서 최신 음성 알고리즘(TTS, VC)에 추가적으로 대응할 수 있는 합성 음성 탐지 기술 개발

한국어 최신 합성 알고리즘 기반 음성 데이터셋 구축

- 단일 여성·남성 화자 합성기(Single-speaker TTS)
- 다화자 합성기(Multi-speaker TTS)
- 음성변환(VS)

실생활 노이즈에 강건한 합성 음성 탐지 기술 개발

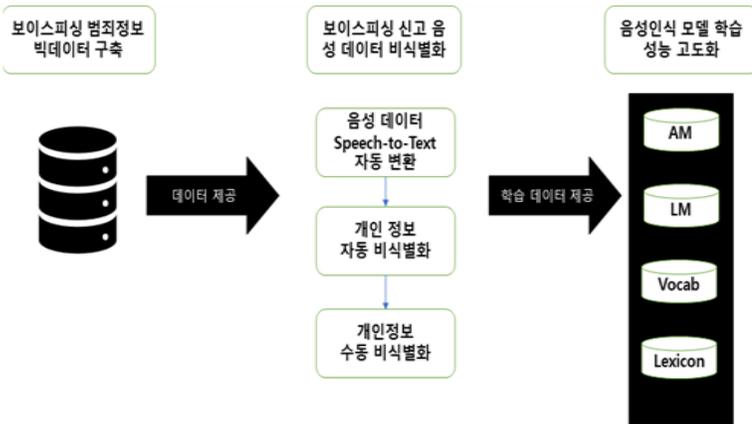
- 백그라운드 노이즈는 음성신호의 품질을 저하시켜 탐지 알고리즘에 영향을 미칠 수 있음
- 실생활에서 발생할 수 있는 다양한 노이즈에 강인한 합성 음성 탐지기 필요
- NOISX-92, DEMAND 데이터셋 활용하여 노이즈를 첨가한 한국어 음성 데이터셋 구축

음성분석 [서울대학교]

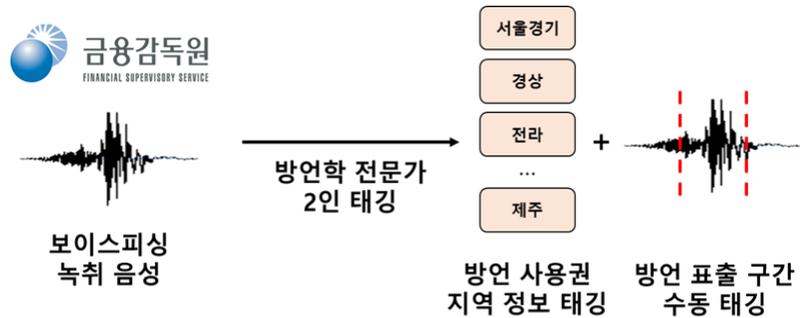


보이스피싱 음성 자동전사 구축 및 음성 특징 분석

보이스피싱 신고 전화 자동전사 성능 고도화



금융감독원 보이스피싱 음성 범죄자 방언 표출 구간 태깅



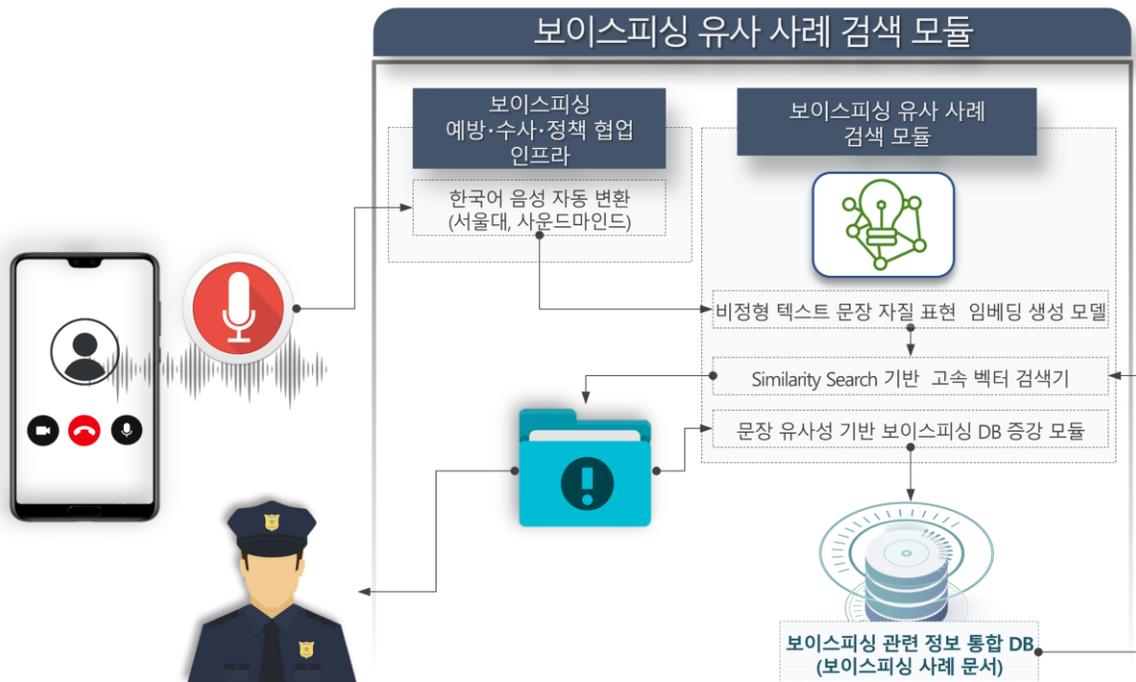
Self-supervised learning 자동전사모델 성능 측정

- Hubert와 xlsr의 최신 동향 반영
- 보유중인 Narrow Band(8Khz) 음성 DB 3,700시간을 활용하여 xlsr 모델 학습
- 보유 레이블 데이터를 이용한 fine tuning 모델 학습
- DNN-HMM chain 모델과 성능 비교

대화내용 분석 [창원대학교]



보이스피싱 유사 사례 검색 기술



문장 유사성 분석 기술 성능 고도화

- 딥러닝 기반 보이스피싱 도메인 한국어 문장 유사도 판단 기술 개발
- 보이스피싱 사례 데이터 추가하여 도메인 범위 확장 및 최적화

Similarity Search를 통한 유사 문장 탐색 및 추천

- 보이스피싱 도메인을 기반으로 성능 고도화
- 범죄 관련 주요 키워드를 활용하여 데이터에 대한 표현력 개선

도메인에 기반한 모델 학습

- 사전 언어 모델에 학습되지 않은 토큰 표현력 보완
- 보이스피싱에 주요한 키워드를 반영한 인코더 학습

유사 문장 탐색 고도화

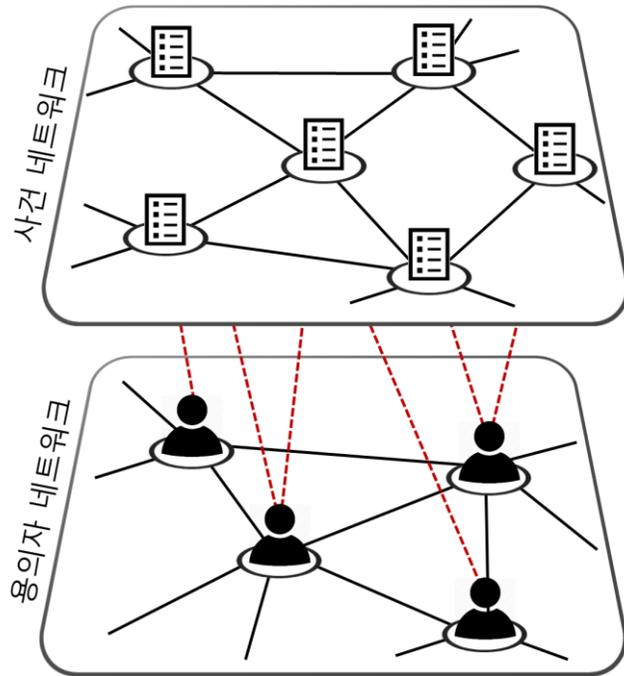
- 유사도 결과 분석을 통한 탐색 기술 고도화
- 탐색 속도 및 성능개선

대화내용 분석 [아주대학교]

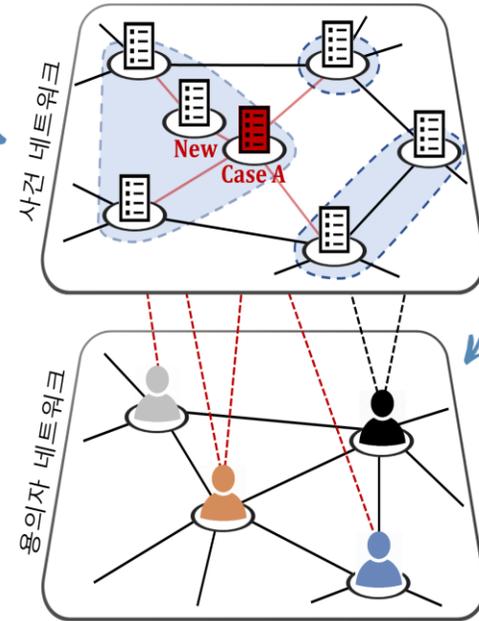


신고 및 사건 네트워크 설계 및 구성 방법론 개발

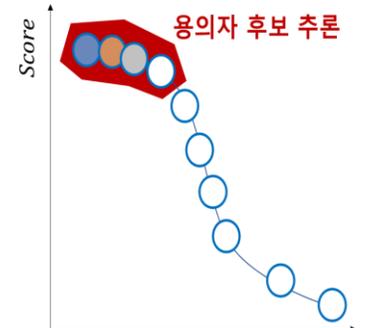
용의자-사건 연결 네트워크 구축



용의자 후보 스코어링을 위한 추론 알고리즘



Machine Learning Algorithms



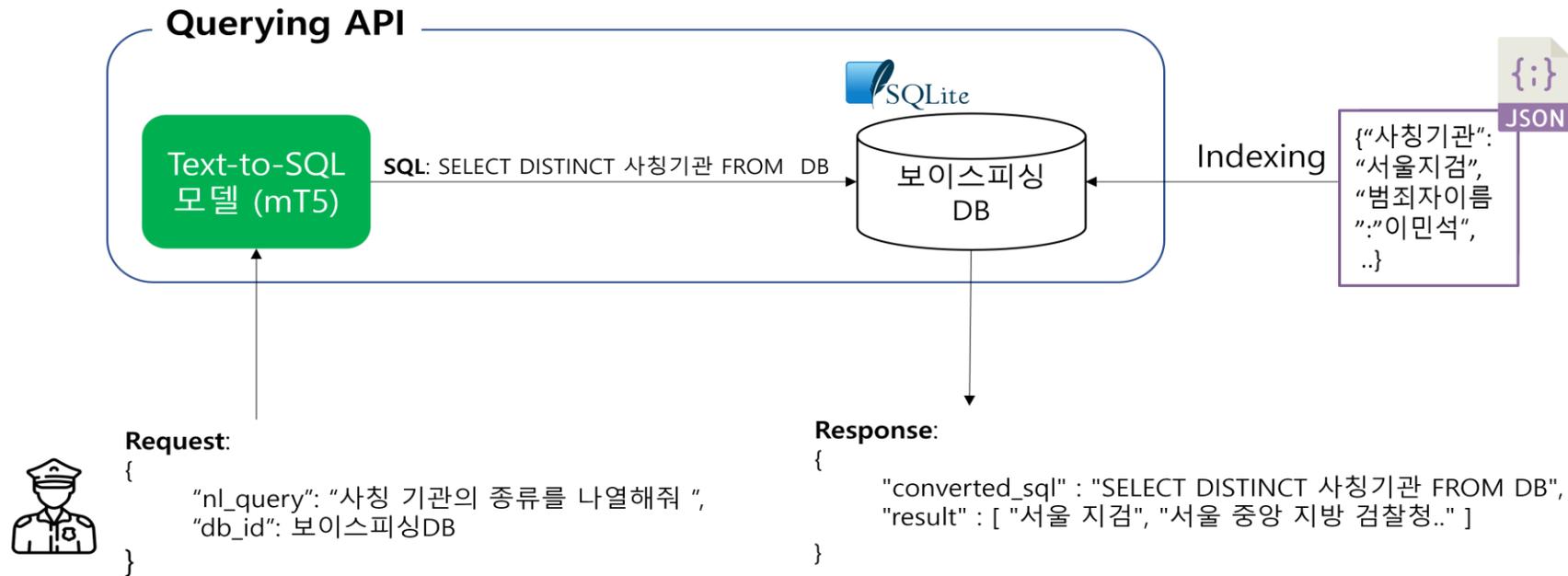
대화내용 분석 [포항공과대학]



보이스피싱 DB 구축 및 검색 시스템 개발

보이스피싱 DB 구축 파이프라인, DB 검색시스템(Text-to-SQL 모델 + Querying API)

- "질문-이름", "질문-생년월일", "질문-전화번호", "질문-나이", "질문-직업", "질문-위치", "사칭기관", "사칭직급", "사칭이름"



수사기법 개발 [경찰대학]



범죄 분석 시나리오를 반영한 보이스피싱 유형화 AI 개발

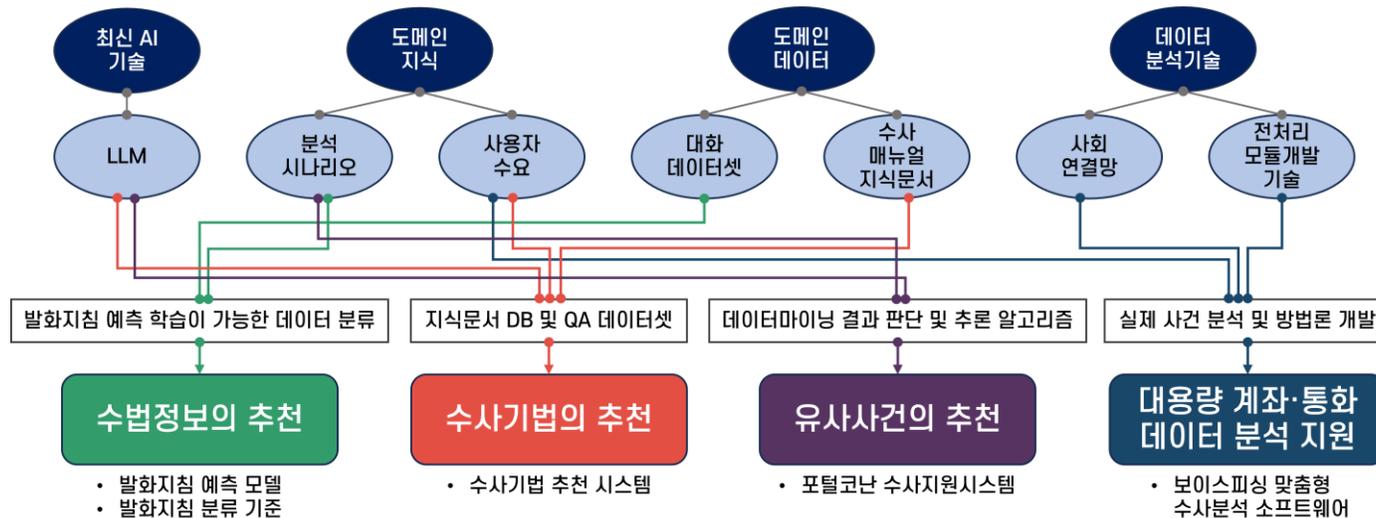
- 금융감독원 데이터 전사본으로 발화지침 예측 학습이 가능한 데이터 분류 → 생성형 언어모델을 활용 {상세 기망 유형 + 발화 의도 + 다음 발화 예측}
- 수사 매뉴얼/법률 매뉴얼에 대한 지식 문서 DB 및 QA 데이터셋 구축 → Retriever 기반 오픈 도메인 QA 태스크를 통해 적절한 보이스피싱 수사기법을 추천해 주는 기술

사용자 시나리오에 기초한 데이터마이닝 결과 판단 및 추론 알고리즘 개발

- 음성 데이터 확보로부터 피해자 및 범인 특징을 위한 영장 작성에 이르는 수사 과정을 반영한 음성/텍스트 AI 기술의 보이스피싱 유사 사건 추정 알고리즘 설계완료 및 이에 따른 화면 구상 완료

보이스피싱 범죄 수사 맞춤형 대용량 다종 테이블 데이터 분석 기술 개발

- 사회연결망 분석 알고리즘 + 다종 수사단서 반자동 전처리 프로세스 + 시각화 기능



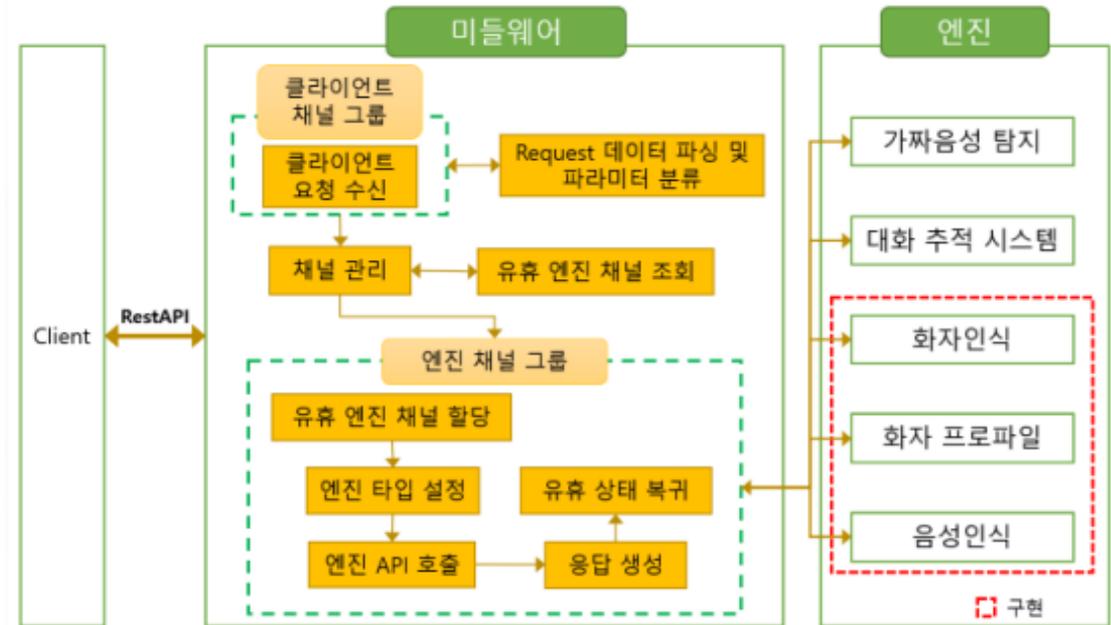
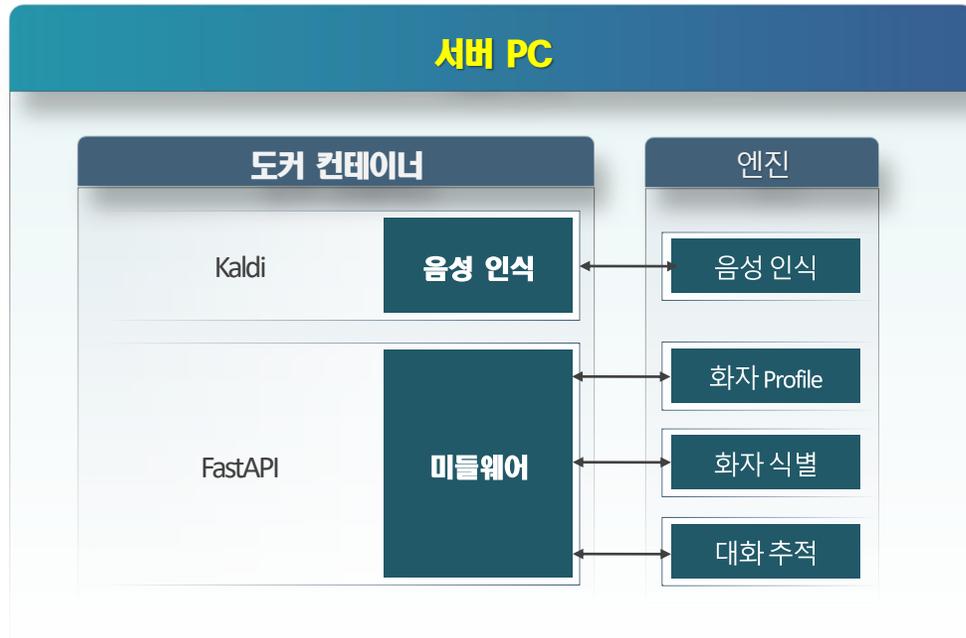
핵심기술 통합 및 API 개발 [㈜사운드마인드]

참여기관이 연구한 기술을 통합한 엔진 개발



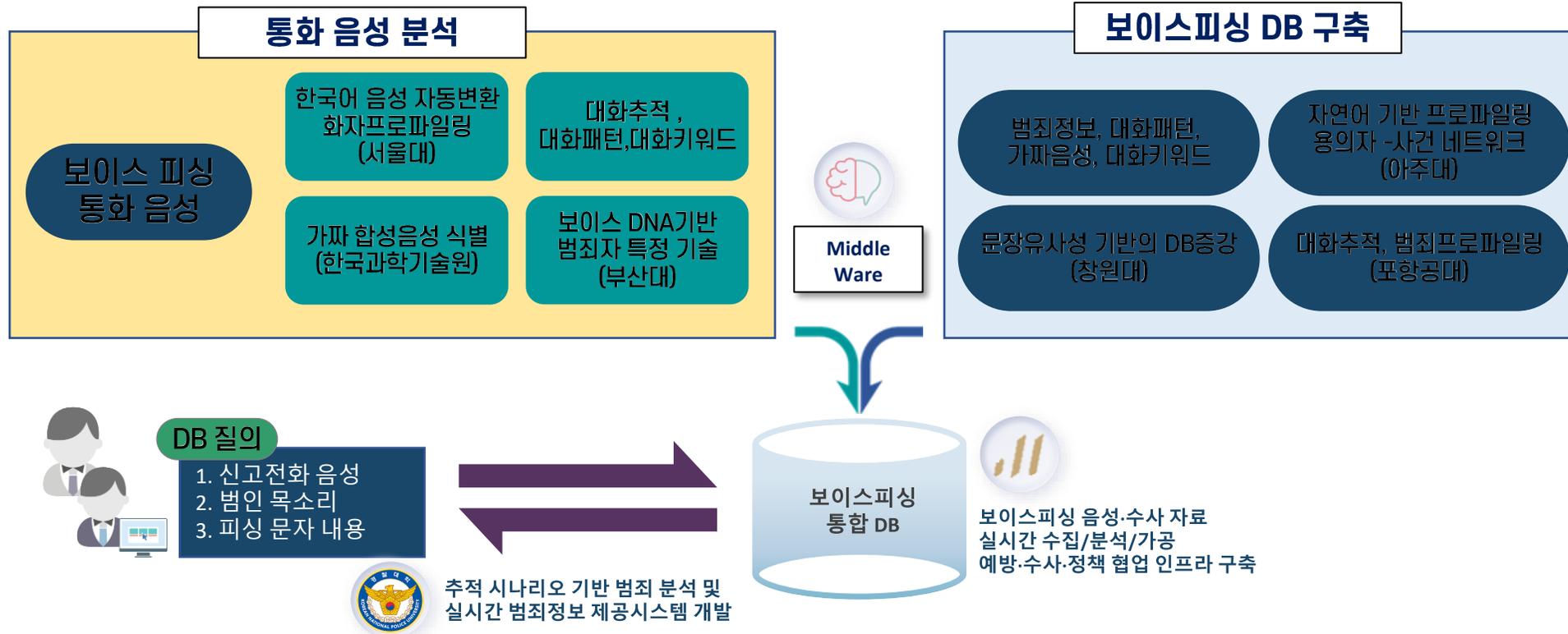
연구기관별 개발 환경을 위한 Docker 사용

미들웨어 프로토타입 개발

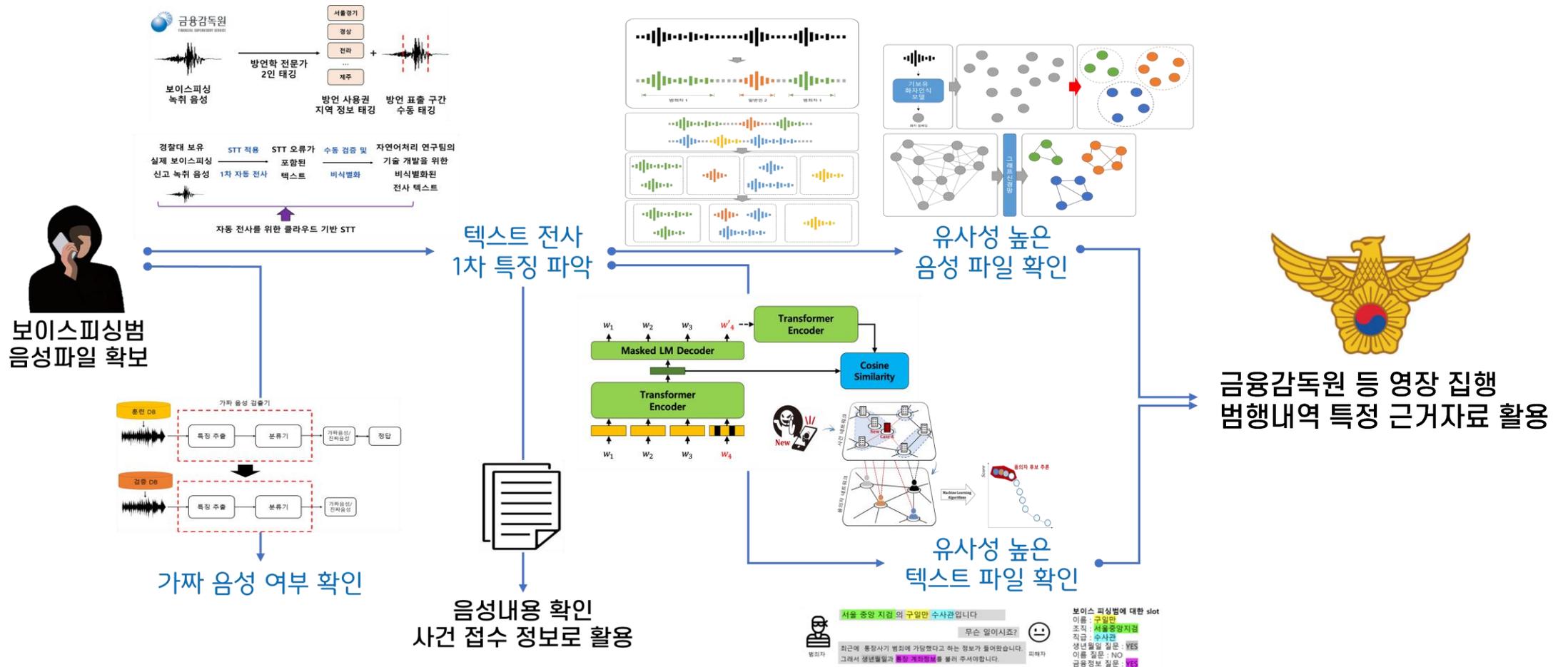


수사 지원 시스템 개발 [㈜일루텍코리아]

통합 엔진을 적용한 클라우드 기반 분석 시스템 구축



연구 개발물 활용 구성도



금융감독원 등 영장 집행
범행내역 특정 근거자료 활용

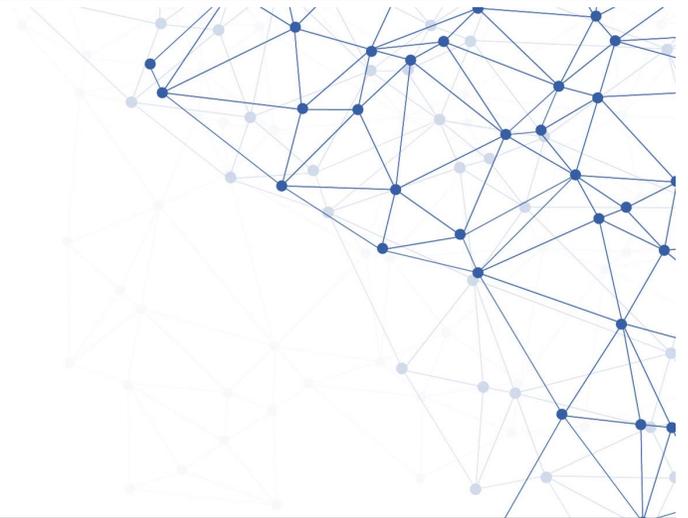
KONAN 화면 정의

voice Fake / voice Note

voice Matching / case Matching

음성비교 국과수

/03



수사지원

통합조회

음성분석

voice
Note

voice
Matching

case
Matching

voice
Fake

음성비교
국과수

Dashboard

새로운 녹음 2

파일선택

김희두

음성 기록

🔒 점유율

편집

- 참석자 1 00:13

죽음을 맞이해야 되는 거 아닌가 많이 늦은 거 아닌가 제가 여기 인터넷이 나와 자려고 했는데
- 참석자 1 00:45

강의실이 인터넷이 안 돼서 다들 기대하셨어요.
- 참석자 1 01:33

지난주 후방은 부원재 교수님 저희 학생 분을 제가 정하겠습니다.
다행히 제가 계산을 해봤을 때 지난 시간에 말씀드렸던 것처럼 시험은 14일이면 통과하기 한 주 전
- 참석자 1 01:58

말씀드렸던 것처럼 치우고 통전을 한 주 여전히 일찍 할 수 있을 것 같고요.
대신에 오늘 좀 스피디하게 좀 나갈 계획이고 제 계산상으로는 오늘 이제 외오 하이더 시스템에 대해서 다시 하고 나면 다
중 주에 멀티 프로세스와 관련된 캐시 포리언시 하고 그런 내용을 다루면서 마무리를 할 수 있을 것 같아요.
- 참석자 1 02:23

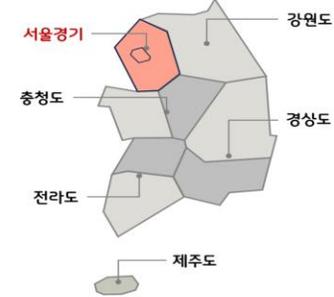
그래서 예정대로 14일에 기말고사를 치우도록 하겠다.
그리고 이제 지난주에 시발을 했으니까 뭐가 해야 되는지? 네 제가 말씀드렸던 것처럼 어쨌든 시험에서 다루는 내용을 다
커버가 가능한 상황이잖아요.
그래서 따로 반드시 풀어야 되는 보강 일정을 잡지는 않고 대신에 시험 전 금요일 9일
- 참석자 1 02:49

제가 좀 세션을 한 오후에 계속 열어놓을게요.
그래서 좀 링크 같은 걸 드리고 만약에 시험 그 범위와 관련해서 궁금한 사항이 있으시거나 아니 경우에 그 중에 접속해서
서 자유롭게 물어보실 수 있도록 지난 중간고사 기간에도 생각보다 질문을 하시는 분들이 많아요.

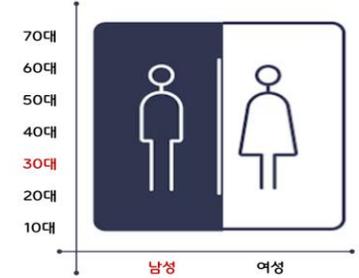
내사건 등록

Next Analysis

지역(방언)



나이 | 성별



검색결과 화면

01:06:16

1x



01:21:07

voice Matching

수사지원

통합조회

음성분석

voice
Note

voice
Matching

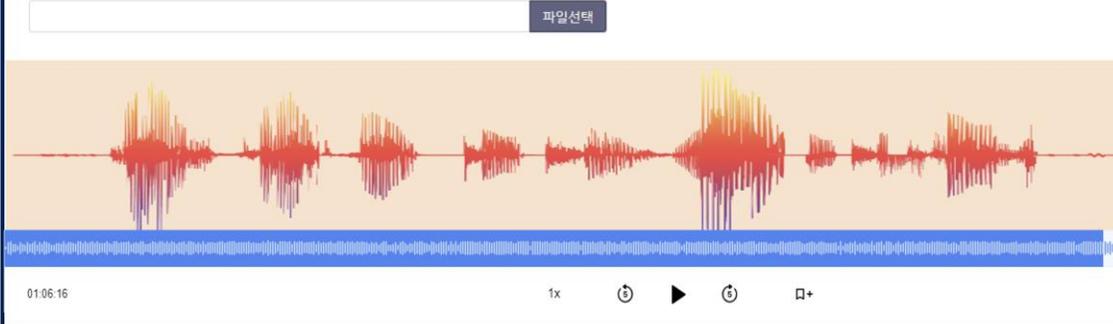
case
Matching

voice
Fake

음성비교
국과수

Dashboard

새로운 녹음 2



화자 구분
[3] 명 보이스피싱 통화 음성에서 화자 구분

음성 전사 내용

- 참석자 1 01:58
말씀드렸던 것처럼 치우고 동전을 한 주 여전히 일찍 할 수 있을 것 같고요.
대신에 오늘 좀 스피디하게 좀 나갈 계획이고 제 계산상으로는 오늘 이제 외오 하이터 시스템에 대해서 다시 하고 나면 다음 주에 멀티 프로세스와 관련된 캐시 포리언시 하고 그런 내용을 다루면서 마무리를 할 수 있을 것 같아요.
- 참석자 1 02:23
그래서 예정대로 14일에 기말고사를 치우도록 하겠다.
그리고 이제 지난주에 시발을 했으니 뭐가 해야 되는데? 네 제가 말씀드렸던 것처럼 어쨌든 시험에서 다루는 내용을 다 커버가 가능한 상황이야요.
그래서 따로 반드시 출석해야 되는 보장 일정을 잡지는 않고 대신에 시험 전 금요일 9일
- 참석자 1 02:49
제가 좀 세션을 한 오후에 계속 열어놓을게요.
그래서 좀 링크 같은 걸 드리고 만약에 시험 그 범위와 관련해서 궁금한 사항이 있으시거나 아니 경우에 그 쪽에 접속해서 자유롭게 물어볼 수 있도록 지난 중간고사 기간에도 생각보다 질문을 하시는 분들이 많아요.

내사건 등록 >>>

Next Analysis >

유사도 TOP5

- 1순위 [98%] 구일만, 서울중앙지검, 수사관 🔍

01.06.16 1x 🔊 ▶ 🔇 ⏸
- 2순위 [91%] 기본정보, 주요 키워드 표출 🔍

01.06.16 1x 🔊 ▶ 🔇 ⏸
- 3순위 [87%] 기본정보, 주요 키워드 표출 🔍

01.06.16 1x 🔊 ▶ 🔇 ⏸
- 4순위 [79%] 기본정보, 주요 키워드 표출 🔍

01.06.16 1x 🔊 ▶ 🔇 ⏸
- 5순위 [65%] 기본정보, 주요 키워드 표출 🔍

01.06.16 1x 🔊 ▶ 🔇 ⏸

Case Matching



내사건 등록 >>>

Next Analysis >

- 수사지원
- 통합조회
- 음성분석**
- voice Note
- voice Matching
- case Matching**
- voice Fake
- 음성비교 국과수
- Dashboard

새로운 녹음 2

파일선택

음성 기록 🔒 ✎

- 참석자 1 00:13

죽음을 맞이해야 되는 거 아닌가 많이 늦은 거 아닌가 제가 여기 인터넷이 나와 자려고 했는데
- 참석자 1 00:45

강의실이 인터넷이 안 돼서 다들 기대하셨어요.
- 참석자 1 01:33

지난주 후방은 부원재 교수님 저희 학생 분들 제가 정하겠습니다. 다행히 제가 계산을 해봤을 때 지난 시간에 말씀드렸던 것처럼 시험은 14일이면 통감하기 한 주 전
- 참석자 1 01:58

말씀드렸던 것처럼 치우고 통전을 한 주 여전히 일찍 할 수 있을 것 같고요. 대신에 오늘 좀 스피디하게 좀 나갈 계획이고 제 계산상으로는 오늘 이제 외오 하이더 시스템에 대해서 다시 하고 나면 다음 주에 멀티 프로세스와 관련된 캐시 프리언시 하고 그런 내용을 다루면서 마무리를 할 수 있을 것 같아요.
- 참석자 1 02:23

그래서 예정대로 14일에 기말고사를 치우도록 하겠다. 그리고 이제 지난주에 시발을 했으니까 뭐가 해야 되는데? 네 제가 말씀드렸던 것처럼 어쨌든 시험에서 다루는 내용을 다 커버가 가능한 상황이야요. 그래서 따로 반드시 들춰야 되는 보강 일정을 잡지는 않고 대신에 시험 전 금요일 9일
- 참석자 1 02:49

제가 좀 세션을 한 오후에 계속 열어놓을게요. 그래서 좀 링크 같은 걸 드리고 만약에 시험 그 범위와 관련해서 궁금한 사항이 있으시거나 아니 경우에 그 중에 접속해서 자유롭게 걸어보실 수 있도록 지난 중간고사 기간에도 생각보다 질문을 하시는 분들이 많아요.

입력으로 넣을 텍스트 드래그 & 끌어 놓기

유사도 TOP5

1순위 [98%]	구일만, 서울중앙지검, 수사관
<p>참석자 1 01:33</p> <p>지난주 후방은 부원재 교수님 저희 학생 분들 제가 정하겠습니다. 다행히 제가 계산을 해봤을 때 지난 시간에 말씀드렸던 것처럼 시험은 14일이면 통감하기 한 주 전</p>	
2순위 [92%]	기본정보, 주요 키워드 표출
유사도 계산된 텍스트	
3순위 [86%]	기본정보, 주요 키워드 표출
유사도 계산된 텍스트	
4순위 [79%]	기본정보, 주요 키워드 표출
유사도 계산된 텍스트	
5순위 [62%]	기본정보, 주요 키워드 표출
유사도 계산된 텍스트	

수사지원

통합조회

음성분석

voice
Note

voice
Matching

case
Matching

voice
Fake

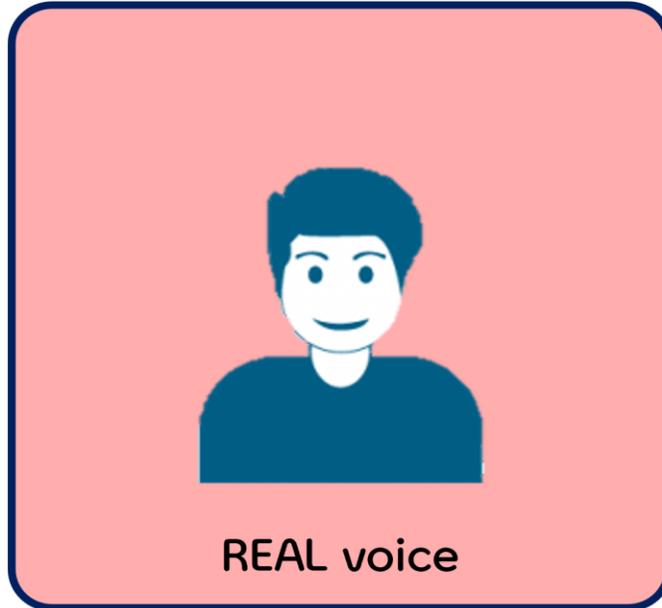
음성비교
국과수

Dashboard

새로운 녹음 2

파일선택

가짜 음성 탐지 기술



수사지원

통합조회

음성분석

voice
Note

voice
Matching

case
Matching

voice
Fake

음성비교
국과수

Dashboard

알림!

분석을 실행하기 전에 범인의 목소리만 20초 ~ 1분가량 편집한 파일을 준비하여야 합니다.

음성 편집 프로그램 다운로드 받기

매뉴얼 다운로드 받기

[1단계]

음성 파일

파일 경로

변환

★ 음성파일을 먼저 (.npy)로 변환 하여야 합니다.

[2단계]

입력 음성

파일 경로



비교 음성

파일 경로



분석 실행

[분석결과]

유사도 점수 [9.525]%

동일 화자일 확률 [81.902]%

동일화자가 아닐 확률 [0.000]%



Thanks.

경찰대학 치안정책연구소
스마트치안지능센터

