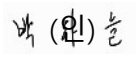
2025년 전남대학교 소프트웨어중심대학사업
소·중·대 산학협력프로젝트(캡스톤디자인) 결과보고서

<table>
<tr><td colspan="2">프로젝트명</td><td colspan="6">영상 분석을 통한 거짓말 탐지기</td></tr>
<tr><td colspan="2">Github url 주소</td><td colspan="6">https://github.com/Luapad/HCoder</td></tr>
<tr><td colspan="2">팀 명</td><td colspan="2">H:CODER</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>컴퓨터정보통신공학과</td><td>성 명</td><td colspan="2">박수형</td></tr>
<tr><td rowspan="8">프로젝트 수행인원
(※팀장은 첫줄에 기입)</td><td colspan="2">이 름</td><td>학과(부·복수전공)</td><td>학년</td><td>학번</td><td>연락처(HP)</td><td>E-Mail</td></tr>
<tr><td>팀장</td><td>박하늘</td><td>컴퓨터정보통신공학과</td><td>4</td><td>221832</td><td>010-4017-5875</td><td>phn0987@naver.com</td></tr>
<tr><td rowspan="5">팀원</td><td>유주형</td><td>컴퓨터정보통신공학과</td><td>3</td><td>214857</td><td>010-2989-5971</td><td>kum7683@naver.com</td></tr>
<tr><td>허성민</td><td>컴퓨터정보통신공학과</td><td>3</td><td>214570</td><td>010-7767-4592</td><td>214570@jnu.ac.kr</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></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">애드시티</td><td colspan="2">조세근</td><td>대표</td><td>010-5583-5004</td><td>adct@adct.co.kr</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 18 일

신청자명(대표학생) : 박하늘 

지도교수 : 박수형 

전남대학교 소프트웨어중심대학사업단장 귀하

산학협력프로젝트(캡스톤디자인) 결과보고서(요약)

프로젝트명	영상 분석을 통한 거짓말 탐지기		
수행기간	2025. 9. 24. ~ 12. 19.	소요예산	396,709원
소요예산 세부내역	-회의비 299,900원 -(추가지원금)SW활용비: 96,809원		

참여인원	구분	인원수	성명(모두 기재)
	교수	1	박수형
	석박사과정	0	
	학부생	3	박하늘, 허성민, 유주형
	기업체	1	조세근
	계	5	

추진배경

배경 및 필요성: 기존 폴리그래프는 고가(2~3억원)의 장비와 물리적 접촉이 필요하여 일반인의 접근이 불가능했음. 이에 스마트폰/웹캠 등 범용 기기를 활용한 저비용·비접촉식 거짓말 탐지 솔루션이 시급함.
최종 목표: 영상 및 음성 데이터를 분석하는 멀티모달 모델을 개발하여, 하드웨어 비용 없이 누구나 활용 가능한 신뢰성 검증 AI 소프트웨어 구축.

목표 및 내용

초기 목표 AUC 0.85 대비 최종 AUC 0.82를 달성(달성률 96.4%).

보완 사항: 초기 0.7 후반대 정체 현상을 극복하기 위해 DBSCAN 군집화로 특정 인물 과적합을 방지하고 , 음성 특징을 정적(Static) 값에서 동적(Delta) 변화량 중심으로 재설계하여 성능을 최적화함. 잔여 격차는 데이터셋의 절대적 양 부족에 기인하며 추후 데이터 확충으로 해결 가능함.

기술적 구현 (Architecture)
1. 경량화 모델 설계: 고비용 3D-CNN을 데이터 규모에 맞게 경량화하여 독자적인 아키텍처를 구축하고 연산 효율성을 확보함.
2. 멀티모달 퓨전(Fusion): 영상·음성 모델을 분리 학습 후 특징 벡터를 결합(Concatenation)하는 방식을 적용, 정보 손실을 최소화함.
3. 최적화: 긴 영상 분석 시 발생하는 지연 문제를 해결하기 위해 문장 단위 파싱(Parsing) 기능을 구현하여 분석 속도를 단축함.

기대효과

1. 사회적: 라이브 커머스 사기 예방, 허위 영상 판독 등을 통해 정보 비대칭(Digital Literacy)을 해소하고 범죄율 감소에 기여함.
2. 경제적: 플랫폼 기업의 정크 데이터 필터링을 통해 서버 운영 비용을 절감하고 기업 신뢰도를 제고함.

1. 프로젝트 개요

프로젝트명	영상 분석을 통한 거짓말 탐지기
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

본 프로젝트는 영상 및 음성 데이터 분석을 통한 비접촉식 거짓말 탐지 AI 모델 개발을 목표로 기존 거짓말 탐지 기술의 근본적인 문제를 해결하고자 한다. 현재 널리 사용되는 폴리그래프(Polygraph) 방식은 피험자의 심박수, 호흡, 피부 전도 반응 등 생체 신호를 측정하여 진실 여부를 판단한다. 그러나 이 방식은 단순히 정확도나 심리적 한계를 넘어, 극심한 비용 문제라는 치명적인 단점을 안고 있다. 이러한 문제점은 기술의 보급과 활용에 있어 심각한 제약으로 작용하며, 거짓말 탐지 기술이 가진 잠재력을 충분히 발휘하지 못하게 하는 주된 원인이 된다.

기존 거짓말 탐지기는 전문적인 장비로 분류되어 고가의 하드웨어 비용이 발생한다. 최신 폴리그래프 장비는 수백만 원에서 수천만 원에 이르는 높은 가격대를 형성하며, 이는 일반 개인은 물론 중소기업이나 비전문 기관의 접근을 원천적으로 차단한다. 심지어 대기업이나 정부기관에서도 도입을 주저하게 만드는 요인으로 작용한다. 또한, 장비의 유지보수, 소모품(전극, 젤 등) 교체 비용, 그리고 전문 분석가 양성을 위한 교육 비용까지 추가로 발생하여 전체적인 비용이 기하급수적으로 늘어난다. 이러한 높은 비용 구조는 거짓말 탐지 기술의 활용 범위를 사법 기관이나 대기업의 특정 부서 등으로 매우 제한시킨다. 기술이 특정 계층과 기관에만 독점적으로 활용되는 불평등한 상황을 초래하며, 기술의 사회적 효용성을 크게 낮춘다.

높은 진입 장벽은 사회 전반에 걸쳐 신뢰성 검증 기술의 대중화를 가로막는 주요 원인이 된다. 예를 들어, 개인 간의 중고 거래나 소규모 계약, 심지어는 엔터테인먼트 목적으로도 사용되기 어렵다. 중요한 사회적 이슈가 발생했을 때 진실 규명에 대한 대중의 요구가 커짐에도 불구하고, 기술적·비용적 한계로 인해 활용되지 못하는 안타까운 상황이 반복되는 것이다. 거짓말 탐지 기술이 극소수의 전문 분야에서만 사용되는 '고급 도구'로 남아 있게 되는 것이다. 이런 현실은 급변하는 정보화 시대에 필요한 빠르고 효율적인 신뢰성 검증 시스템의 부재로 이어지며, 사회적 불신을 심화시킬 수도 있다.

본 연구는 이러한 비용 문제를 해결하기 위한 혁신적인 대안을 제시한다. 바로, 스마트폰, 웹캠과 같은 일상적인 기기를 활용하는 것이다. 이는 개인이 이미 소유하고 있는 보편적인 장치들을 최대한 활용하여, 추가적인 하드웨어 구매에 대한 부담을 없애는 것을 핵심으로 한다. 이를 통해 누구나 쉽고 저렴하게 접근할 수 있는 솔루션을 제공하고자 한다.

기존의 하드웨어 기반 방식과 달리, 본 프로젝트는 소프트웨어 중심의 AI 모델을 개발함으로써 하드웨어 구매 비용을 제로화한다. 사용자들은 이미 보유하고 있는 기기를 활용하여 즉시 분석을 시작할 수 있다. 예를 들어, 스마트폰 앱이나 웹 기반 서비스 형태로 제공되어 별도의 장비 설치 없이 바로 이용 가능하게 하는 것이다. 이는 거짓말 탐지 기술을 누구나 손쉽게 사용할 수 있게 전환하는 중요한 첫걸음이 될 것이다. 사회 구성원 모두가 공정하게 신뢰성을 검증할 수 있는 환경을 조성하는 데 기여한다.

본 연구는 고가의 장비 없이도 작동하는 AI 모델을 통해 기존 시장의 독점적 구조를 깨고, 거짓말 탐지 기술의 접근성을 획기적으로 향상시켜 새로운 시장을 창출하고 사회 전반의 신뢰성 증진에 기여할 것이다.

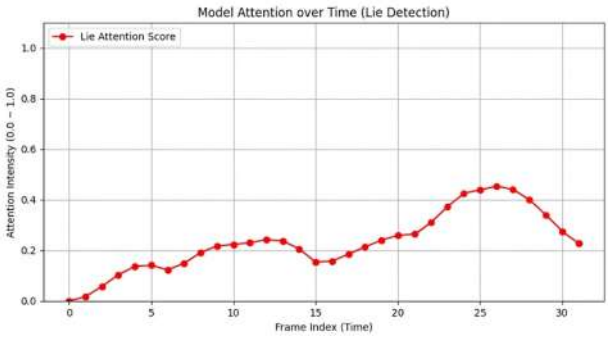
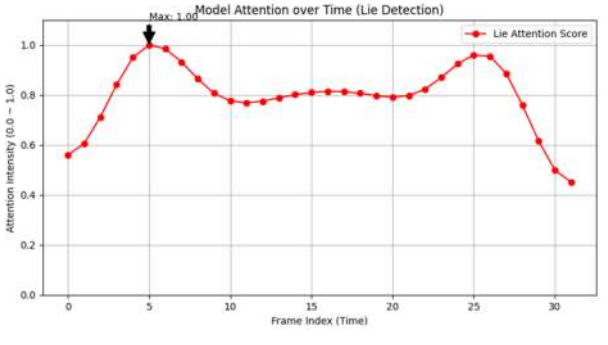
AI 모델은 사람의 미세한 표정 변화, 목소리 톤, 음성의 높낮이, 발화 속도 등 비언어적 단서를 정밀하게 분석하여 거짓 여부를 판단하는 방식으로 작동한다. 이러한 비접촉식 방식은 피험자에게 심리적 부담을 주지 않으면서도 객관적인 데이터를 확보할 수 있다는 장점이 있다.

결과적으로 본 프로젝트는 기술적 진보를 넘어, 사회적·경제적 가치를 창출하는 중요한 전환점이 될 것이다. 궁극적으로는 투명하고 신뢰할 수 있는 사회를 구축하는 데 필수적인 도구로 자리매김할 것으로 기대된다.

3. 프로젝트(주제) 목표 및 내용

1) 정성적 연구개발성과(연구개발결과)

1. 영상 모델 분석 과정

<table>
<tr>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td>진실 영상의 score 그래프</td>
<td>거짓 영상의 score 그래프</td>
</tr>
</table>

(score: 거짓말 특징의 강도)

거짓으로 라벨링 된 영상의 분석 결과, score 그래프가 뚜렷한 피크(Peak)를 형성하며 최댓

값 1.00에 도달하는 고강도 활성화 패턴이 관찰되었다. 이는 발화자가 특정 시점(Frame)에서 거짓말과 관련된 결정적인 행동적 징후가 발현되었음을 의미한다.

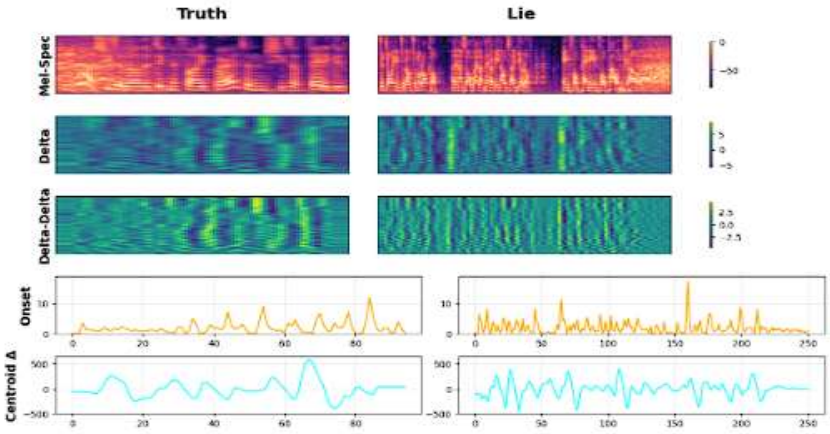
반면, 진실 영상의 경우 전체 프레임에 걸쳐 Score 0.5 미만의 낮은 수치를 유지하는 안정적인 패턴을 보였다. 거짓 영상에서 나타났던 급격한 스파이크(Spike)나 지속적인 고강도 활성화 구간은 관찰되지 않았다. 이는 모델이 해당 영상 내에서 기만행위로 의심할 만한 유의미한 특징을 발견하지 못했음을 시각적으로 보여준다.

Grad-CAM 기법을 통한 시각화 분석 결과, 모델은 진실 데이터와 거짓 데이터 간에 명확한 활성화 패턴의 차이를 보였다.

우선 거짓 영상의 경우, 모델은 발화자의 이마와 안면 중심부에서 붉은색의 강한 활성화 반응을 나타냈다. 이는 모델이 해당 영역의 근육 움직임을 거짓 판별의 핵심 특징으로 포착했음을 의미한다. 해부학적으로 이마 근육의 움직임은 눈을 크게 뜨거나 눈가에 힘을 줄 때 눈썹이 함께 움직이며 주름이 형성되는 과정과 직결된다. 실제 해당 영상에서도 발화자가 눈에 힘을 주는 부자연스러운 모습이 직관적으로 관찰되는데, 모델 역시 이러한 눈과 이마 주변의 긴장 및 움직임을 강력한 거짓말 징후로 인식하여 Score 1.00의 높은 확신도로 거짓을 판별하였다.

반면, 진실 영상에서는 특정 부위에 대한 집중적인 활성화 없이 전체적으로 푸른색(비활성화) 분포를 보였다. 이는 거짓으로 의심할 만한 특이 패턴이나 근육의 과도한 움직임이 발견되지 않았음을 시사하며, 모델은 이를 근거로 해당 영상을 진실로 올바르게 분류하였다.

2. 음성 모델 분석 과정



Mel-Spectrogram (에너지 흐름): 진실 데이터는 배음(Harmonic) 구조가 선명하고 시간 축에 따른 에너지의 흐름이 끊김 없이 연속적인 스펙트럼 연속성(Spectral Continuity)을 보인다. 반면, 거짓 데이터는 발화의 불확실성으로 인해 주파수 대역 간의 에너지가 흐릿하게 흩어지는 스미어링(Smearing) 현상이 관찰되거나, 구간별 에너지가 급격히 단절되는 등 전반적으로 불안정한 발성 패턴이 나타난다.

Delta (변화 속도): 진실 데이터는 조음 기관의 움직임이 규칙적이고 리듬감 있게 변화하여 시각화된 패턴이 안정적이다. 이와 달리 거짓 데이터는 심리적 압박으로 인한 머뭇거림(Hesitation)과 부정확한 발음(Slurring)으로 인해 불규칙한 노이즈 패턴이 다수 검출된다. 특히 시각화 자료에서 노란색으로 표시되는 고에너지 변화 구간이 산발적으로 나타나는데, 이는 발화 속도와 세기가 비정상적으로 급변함을 시사한다.

Delta-Delta (변화 가속도): 진실 데이터는 발화의 가속도가 매끄럽게 정돈되어 있어 음성 신호의 동적 복잡성이 낮다. 그러나 거짓 데이터는 인지적 부하(Cognitive Load)로 인해 발화 제어가 불안정해지면서 패턴이 난잡하고 복잡한 양상을 띤다. 시각화 그래프 상에서 급격한 변화를 의미하는 노란색 영역으로 알 수 있다.

Onset Strength (발화 리듬): 진실 데이터는 발화의 강세(Stress)와 어조가 주기적이고 일정한 리듬을 유지하는 안정적인 운율(Prosody)을 보인다. 반면, 거짓 데이터는 전반적으로 발화가 위축(Suppressed)되어 있다가 특정 구간에서 급격하게 에너지가 튀는 충동적인 스파이크(Spike) 형태의 변화를 보인다. 이는 거짓말을 꾸며내는 과정에서 발생하는 긴장과 이완의 불규칙한 반복으로 인해, 진실 데이터 대비 변화의 빈도가 잦고 진폭의 편차가 크게 나타나는 것으로 분석된다

Centroid Delta (음색 안정성): 진실 데이터는 소리의 밝기를 나타내는 스펙트럼 중심(Spectral Centroid)이 발화의 흐름에 따라 완만한 곡선을 그리며 이동한다. 이에 반해 거짓 데이터는 심

리적 긴장 상태가 유발하는 성대의 미세 떨림(Micro-tremor)과 호흡 불안정으로 인해, 중심 주파수의 변화 그래프가 뾰족하고 격렬하게 요동치는 고주파수 변동성을 보인다.

(※발화 리듬과 음색 안정성 그래프의 가로축은 전체 영상의 시간 축을 의미하므로, 특정 시점이 아닌 영상 전체에 걸친 변화의 거시적 흐름(Global Trend)을 중심으로 해석하였음.)

2) 목표 달성 수준

목표 AUC 점수: 0.85

달성 AUC 점수: 0.82

3) 목표 미달 시 원인 분석(해당 시)

3-1) 목표 미달 원인(사유) 자체분석 내용

초기 모델 성능은 꾸준히 향상되었으나, 0.7 후반대에서 수렴하며 정체되는 양상을 보였다. 이에 대한 원인 분석 결과, 영상 모델에서는 얼굴 인식 실패율이 높았고, 음성 모델의 경우 고유값(Eigenvalue) 기반 특징의 기준점이 모호하여 성능 저하의 원인이 되었다. 또한, 영상과 음성 데이터 간의 입력 차원 불일치로 인한 정보 손실 문제도 식별되었다.

식별된 문제점들을 개선하기 위해 가능한 모든 기술적 조치를 적용하였고, 그 결과 최종 성능을 0.82까지 끌어올렸다. 현 단계에서 기술적 최적화는 충분히 이루어졌다고 판단되며, 추가적인 성능 향상을 가로막는 주된 요인은 데이터셋의 절대적인 양 부족에 있는 것으로 생각된다.

3-2) 자체 보완활동

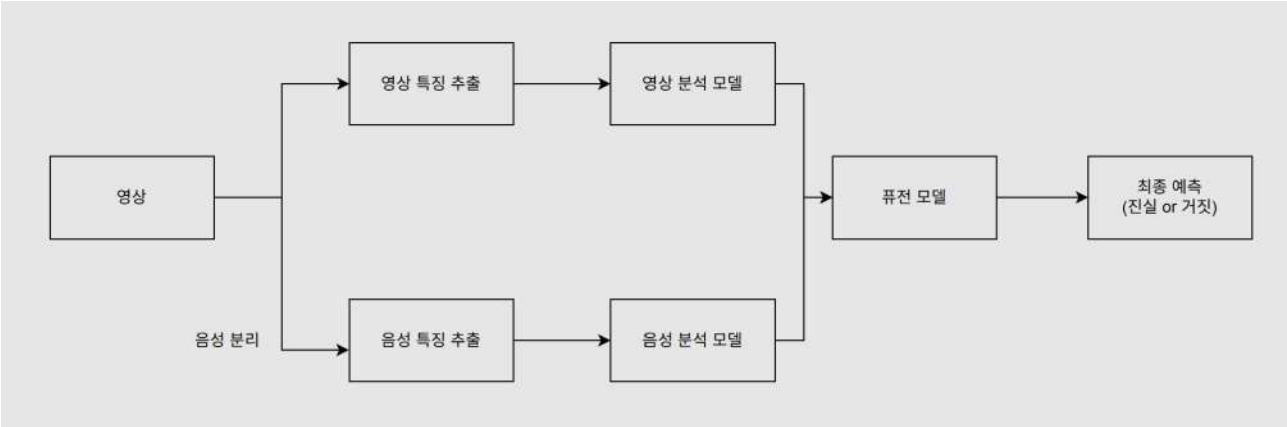
[그림 1]

원본 영상을 그대로 사용할 경우 불필요한 배경 보정으로 인해 모델의 주의가 분산되거나 얼굴 인식 실패율이 증가하는 문제가 발생하였다. 이러한 한계를 보완하고 탐지 성능을 극대화하기 위해, 영상 내에서 얼굴 영역만 따로 자르는 전처리 과정을 거쳤다.

또한, 동일 인물의 영상이 학습 데이터와 평가 데이터에 혼재될 경우, 모델이 거짓말의 특징이 아닌 특정 인물의 생김새(Identity)를 암기하여 성능이 과대평가되는 과적합(Overfitting) 문제가 발생할 수 있다. 본 연구에서는 이러한 문제를 사전에 차단하기 위해, 비지도 학습 기반의 밀도 기반 군집화 알고리즘인 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)을 도입하였다. 이를 통해 영상 속 등장인물들의 얼굴 특징 벡터를 추출 및 군집화하여 각 발화자에게 고유 식별자를 부여하는 전처리 과정을 수행하였다.

기존 음성 모델은 특징값의 절대적인 크기에 의존하다 보니, '특정 데시벨(dB) 이상은 거짓'과 같은 편향된 규칙을 학습하는 문제가 있었습니다. 이는 영상 처리에 비유하자면, 상황의 전후 흐름을 무시한 채 정지된 한 장면만을 보고 판단하는 것과 비슷한 결과였습니다. 이를 개선하기 위해, 영상에서 물체의 움직임을 추적하듯 음성 특징값을 미분하여 그 변화량을 학습시켰습니다. 결과적으로 모델이 단순한 수치 값에 매몰되지 않고 소리가 변화하는 동적인 흐름 자체를 파악하게 되어, 성능의 안정성을 높일 수 있었습니다.

4. 시스템 구성 및 내용



5. 프로젝트 결과물에 대한 기술

1) 연구개발과제의 수행 과정 및 수행 내용

1.1 이론적·실험적 접근 방법

본 연구는 비접촉 환경에서 인간의 기만 행위를 판별하기 위해 멀티모달 데이터 분석(Multimodal Data Analysis) 기법을 핵심 방법론으로 채택하였다. 인간은 거짓말을 할 때 인지 부하가 증가하며, 이는 의도적으로 통제하기 어려운 미세한 신체적 변화를 동반한다. 본 팀은 이러한 변화가 시각적 요소(얼굴의 미세 움직임)와 청각적 요소(발화 패턴의 변동)에 동시에 투영된다는 가설을 세우고 실험적 접근을 시도하였다.

먼저, 연구의 실효성을 검토하기 위해 머신러닝 기반의 통계적 접근을 우선 실시하였다. MediaPipe와 Librosa를 활용하여 기초 특징 파이프라인을 구축하고, 저사양 환경에서도 신속한 검증이 가능한 Random Forest 모델을 최종 목표 정확도를 잡는 기준으로 설정하였다. 초기 테스트 결과 AUC 0.70~0.73의 유의미한 수치를 확보하였으며, 이를 통해 비접촉식 데이터만으로도 거짓말 탐지가 가능하다는 타당성을 확인하였다. 이후 성능 고도화를 위해 딥러닝 기반의 3D-CNN 및 멀티모달 퓨전(Fusion) 구조로 연구 범위를 확장하였다.

1.2 연구개발과제 수행 과정 및 상세 내용

프로젝트 초기에는 최적의 안면 인식 및 특징 추출 모델을 선정하기 위한 비교 분석을 진행하였다. dlib 기반의 CNN 구조를 설계하여 1차 테스트를 수행하였으나, 측면 얼굴 인식률 저하와 라이브러리 간 버전 충돌로 인한 배포의 어려움을 확인하였다. 이를 해결하기 위해 DeepFace를 활용한 인물 식별과 OpenCV 기반의 정밀 전처리 과정을 도입하여 입력 데이터의 신뢰도를 확보하였다.

음성 데이터 처리 단계에서는 초기 추출된 정적 특징(Static Feature)들이 화자의 감정 변화나 심리적 동요에 따른 시계열적 특성을 충분히 반영하지 못하는 한계가 있었다. 이를 보완하

고자 주파수의 변화량에 집중한 동적 특징(Delta Feature) 중심의 재설계 과정을 거쳤으며, 이를 통해 발화 속도와 무음 구간(Pause) 등 거짓말 탐지에 유효한 핵심 지표들을 정량화하였다.

초기 도입한 3D-CNN은 시공간 특징 추출에는 탁월하나, 제한된 데이터셋 대비 파라미터 수가 과도하여 과적합(Overfitting) 발생 가능성이 높았다. 이를 해결하기 위해 본 팀은 독자적인 경량화 3D-CNN 아키텍처를 설계하였다. 연산 비용을 대폭 절감하면서도 핵심적인 안면 움직임 데이터만을 선별적으로 학습하도록 구조를 개선하였다. 또한, 영상과 음성 데이터의 입력 규격(Input Size) 차이로 인해 발생하는 정보 손실을 막기 위해, 통합 모델 대신 영상 및 음성 특화 모델을 분리하여 개별 학습시키는 전략을 채택하였다. 이 과정에서 각 모델의 성능을 독립적으로 최적화하여 전체 파이프라인의 견고함을 높였다.

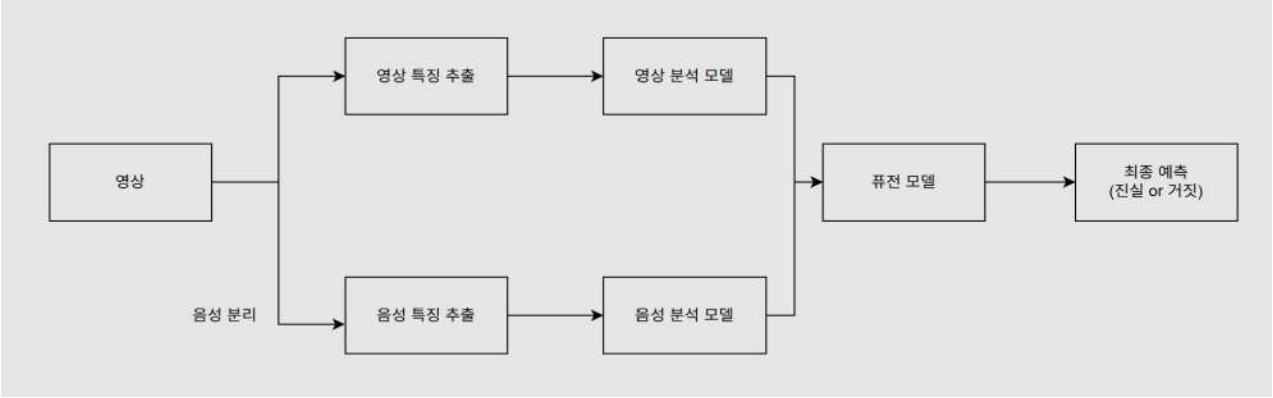
분리된 두 모델의 출력을 하나로 결합하여 최종 판별을 내리는 퓨전 단계는 본 프로젝트의 핵심 기술적 성과이다. 각 모달리티(영상, 음성)의 고유한 정보를 손실 없이 보존하기 위해 결합(Concatenation) 방식의 퓨전을 적용하였다.

각 모델에서 추출된 고차원 특징 벡터(Feature Vector)들을 병합한 후, 완전 연결 계층(Fully Connected Layer)을 통과시켜 최종적인 진실/거짓 확률을 도출한다. 이 방식은 단순 평균 방식에 비해 시청각 데이터 간의 상관관계를 더욱 정밀하게 학습할 수 있음을 실험적으로 증명하였다.

학습된 모델을 실제 서비스 환경에 적용하기 위한 웹 아키텍처 설계와 최적화 작업을 병행하였다. 로컬 서버와 웹 인터페이스를 연동하는 과정에서, 고해상도 영상의 경우 분석 시간이 비정상적으로 길어지는 병목 현상을 확인하였다.

이를 해결하기 위해 문장 단위 파싱(Parsing) 기능을 설계하고 구현하였다. 전체 영상을 의미 있는 최소 단위로 분할하여 병렬 처리함으로써 체감 분석 시간을 획기적으로 단축하였다.

본 프로젝트는 데이터 수집부터 모델링, 시스템 통합에 이르기까지 팀원 간의 유기적인 협업을 통해, 영상과 음성 데이터가 각각 독립적인 처리 과정을 거쳐 최종적으로 융합되는 퓨전 모델을 구축하였다.



dlib의 한계를 DeepFace와 OpenCV로 극복하고, 정적 음성 특징을 동적 특징으로 고도화하며, 무거운 모델을 경량화 3D-CNN으로 대체하는 등 수많은 기술적 난제를 실험적으로 해결

하였다. 이러한 과정은 단순한 기술 구현을 넘어, 실제 시장(중고거래, 라이브커머스 등)에서 즉시 활용 가능한 수준의 비접촉식 거짓말 탐지 솔루션을 도출하는 밑거름이 되었다.

2) 연구개발 과정의 성실성 (조원 각자의 역할)

[박하늘]

프로젝트 초기, 베이스 모델 선정을 위해 dlib과 CNN을 기반으로 한 모델 구조를 설계하고 조원들의 모델과 성능을 비교 분석하였다. 검증 과정에서 dlib의 한계점(측면 영상 인식 취약, 학습 환경 버전 충돌 등)을 확인하여, 최종적으로 프로젝트에 더 적합한 타 모델을 선정하였다.

모델 선정 이후에는 웹 아키텍처 설계와 구현을 전담하였으며, 임시 로컬 서버를 구축하여 웹과의 연동 및 테스트를 진행하였다. 테스트 중 영상 길이가 길어질수록 분석 시간이 과도하게 소요되는 문제를 발견하여, 이를 해결하고자 문장 단위 파싱 기능을 도입 및 설계하였다.

이후 파싱 모델을 구현해 서버와 연동시켰으며, 영상 분석 모델의 고도화 작업을 지원하였다. 특히 영상 처리 과정에서 얼굴 인식 실패(Detection Failure) 문제를 파악하고, 입력 데이터 전처리(Preprocessing) 과정을 개선하여 인식률을 높이는 데 기여하였다.

[허성민]

프로젝트 초기, 연구 타당성 검증 및 목표 성능 설정을 위해 MediaPipe와 Librosa를 활용한 특징 추출 파이프라인을 구축하였다. 초기 연구 가능성의 확인을 위해 신속한 결과확인과 저사양 환경에에서 가능한 random forset를 사용해서 초기 테스트를 진행했다. 테스트 결과 AUC 0.70~0.73을 달성하여 연구 가능성을 확인하고 이를 기준으로 최종 목표를 AUC 0.85로 설정하였다.

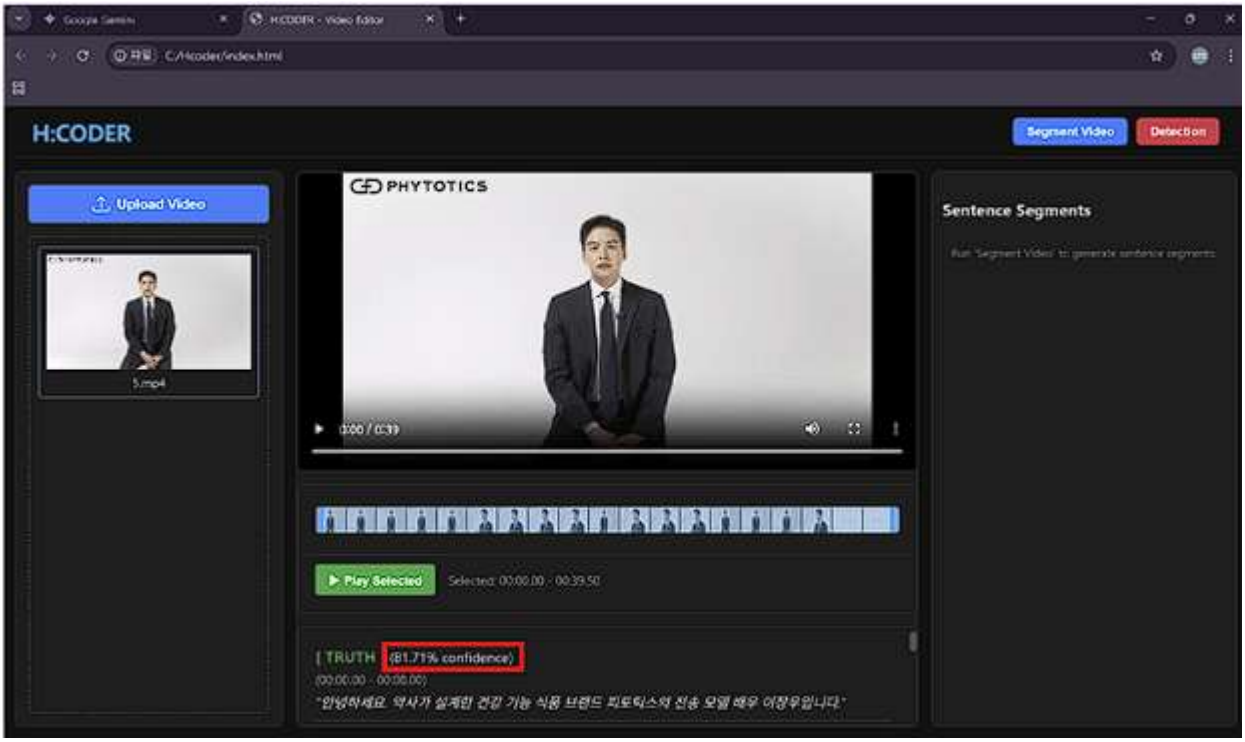
이후 성능 향상을 위해 시계열 데이터 처리에 강점이 있는 RNN, LSTM 도입을 검토했으나, 팀 내 논의를 거쳐 CNN 구조로 최종 확정하고 영상과 음성을 분리하여 학습하는 전략을 수립하였다. 특히 음성 모델 개발 과정에서는 초기 고유값 중심의 정적 특징(Static Feature)들이 시계열 변화를 반영하지 못하는 점을 보완하고자, 변화량(Delta) 중심의 동적 특징으로 재설계하여 학습 효율을 높였다.

최종적으로 영상 모델 담당자와 협업하여 멀티모달 퓨전(Fusion) 모델을 개발하였다. 퓨전 단계에서는 특징 벡터(Feature Vector)들을 평균(Averaging)하는 방식과 결합(Concatenation)하는 방식을 비교 실험하였으며, 각 모달리티의 고유 정보를 보존하며 성능을 극대화할 수 있는 결합 방식을 채택하여 파이프라인을 완성하였다.

[유주형]

초기 단계에서 DeepFace로 인물을 식별하고 3D-CNN 베이스라인을 구축하였다. 데이터셋 규모 대비 연산 비용이 과도함을 파악하여 독자적인 구조의 경량화 3D-CNN을 설계해 효율성을 확보하였다. 이후 통합 모델 학습 과정에서 영상과 음성 데이터의 입력 크기 불일치로 인한 정보 손실 문제와, 정확도 70%대 후반에서 발생하는 과적합 현상을 직면하였다. 이를 해결하기 위해 영상과 음성 모델을 분리하기로 결정했으며, 이 중 영상 모델 개발을 전담하였다.

영상 파트에서는 OpenCV 기반의 정밀 얼굴 전처리와 앞서 설계한 경량화 모델을 적용하여 모델의 성능을 높였다. 최종적으로 분리된 영상 및 음성 모델의 출력을 결합하는 Fusion 모델을 구현하였고, 각 모델에서 추출된 특징을 이어 붙이는 방식을 적용하여 시청각 정보를 동시에 분석하는 멀티모달 거짓말 탐지 파이프라인을 완성하였다.



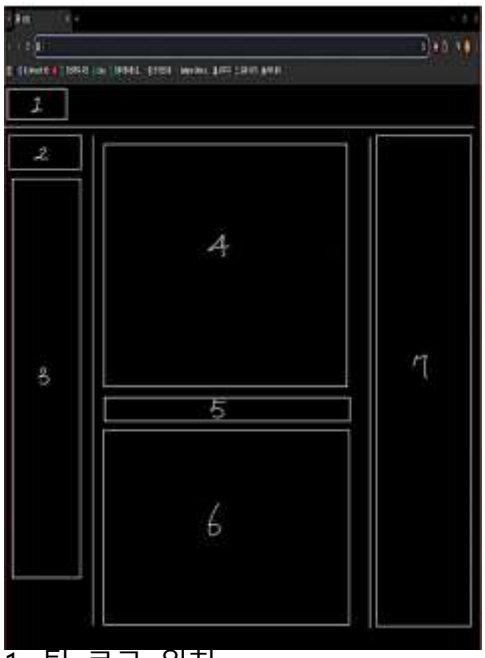
구분	기능정의	세부기능 설명
1	모델 - 거짓말 분석	영상을 분석하여 최종 거짓말 여부를 알려줌
2	웹 - 업로드	동영상을 여러개 업로드하여 선택작업이 가능
3	웹 - 파싱	선택한 영상을 파싱하여 문장별로 분석을 보낼 수 있음
4	웹 - 재생	선택한 영상을 전체, 일부분 재생이 가능

6. 프로젝트 진행내용

1) 참여인원 및 담당 역할

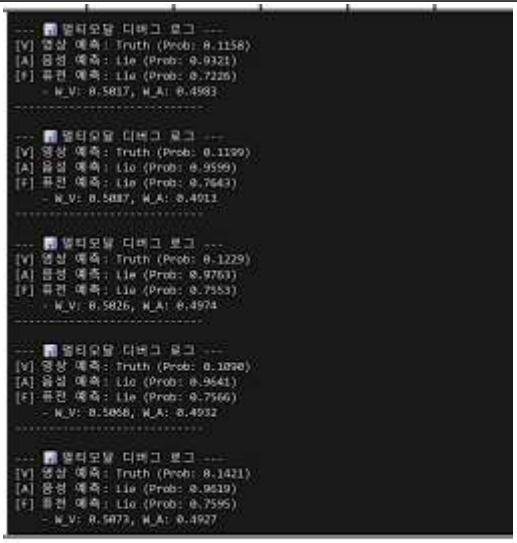
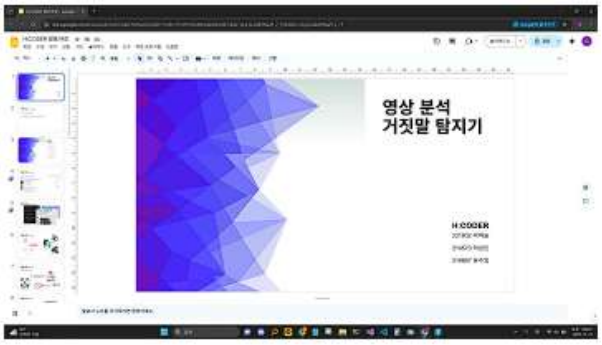
연번	소속학과	성명	수행역할 분담내용
1	컴퓨터정보통신공학과	박하늘	웹&서버 개발
2	컴퓨터정보통신공학과	유주형	영상분석모델 개발
3	컴퓨터정보통신공학과	허성민	음성분석모델 개발

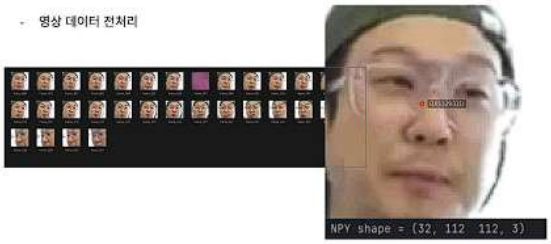
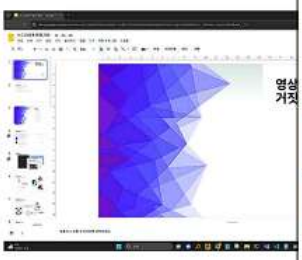
2) 회의 및 SW멘토링 진행

번호	일시/장소	회의/멘토링 내용(상세히 작성)	관련 사진
1	2025. 10. 1. (19:00~22:00) / 디지털도서관	 1. 팀 로고 위치 : 팀명 H:CODER를 로고화 하여 배치할 것 2. 업로드 버튼 : 동영상 아이콘과 함께 '업로드'라는 글자를 넣어서 버튼 생성 3. 업로드한 동영상 목록 : 각 동영상의 썸네일을 보여주고 이 위치에 동영상 파일을 드래그해서 올려도 업로드가 가능하게 기능구현 할 것. 4. 동영상 보여주기 : 3에 있는 업로드된 목록 중 하나를 클릭하면 해당 동영상이 틀어지는 구간. 7번에서 타임라인을 선택하면 해당구간으로 이동도 가능하게 구현5. '분석' 버튼을 넣을 것 6. 7번에서 선택한 타임라인 구간을 분석한 뒤 분석 결과를 보여주는 곳7. 사용자가 선택한 동영상을 파싱하여 타임라인으로 나눔. 각 타임라인마다 시작지점의 썸네일을 보여주고 구간을 시간으로도 보여줌. 선택가능하도록 설정함.	

2	2025. 10. 13. (19:00~22:00) / 디지털도서관	 1. Web UI 수정 - 분석 버튼 세분화(영상분석과 거짓말 분석) - 영상 타임라인에 선택표시(selected) 추가 2. 모델 개선 - LSTM, MPL 방식 비교 분석	
3	2025. 10. 29. (19:00~22:00) / 디지털도서관	 1. Web 수정 - 분석 모델에 맞춰 UI 개선 - 모델과 연동을 위한 백엔드 서버 구축 2. 모델 개선 - 하이퍼 파라미터 수정(히든 레이어 수, 뉴런 수 등)조정하며 정확도 개선 작업.	
4	2025. 11. 5. (19:00~22:00) / 디지털도서관		

		1-1 우측 타임라인 복구  1-2 임시 서버 개발  1. Web 수정 - 타임라인 복구(스크롤 바로만 조절하는 불편함이 있었음.) - 임시 서버 개발 2. 모델 개선 - 하이퍼 파라미터 수정(히든 레이어 수, 뉴런 수 등)조정하며 정확도 개선 작업.	
5	2025. 11. 12. (19:00~22:00) / 디지털도서관	 1. Web 수정 - 타임라인 강화: 파싱 모델과 연동하여 임시 타임라인이 아닌 실제 타임라인 생성 2. 분석 모델 개선 - 하이퍼 파라미터 수정(히든 레이어 수, 뉴런 수 등)조정하며 정확도 개선 작업. 3. 파싱 모델 개선 - 기존 모델의 GPU사용 제한으로 인해 모델 변경 시도.	
6	2025. 11. 18. (19:00~22:00) / 디지털도서관		

		 1. PPT 작성 - 다음주 멘토링 회의를 위해 미리 PPT 작업을 진행. 2. 분석 모델 문제점 파악 - 여러 가능성을 두고 소거법으로 문제점 찾기 -> 영상의 얼굴 추출 문제는 아닌 것을 파악 후 로그 분석을 통해 음성 모델에서 문제가 생긴 것을 찾음.	
7	2025. 11. 19. (14:00~17:00) / 공7-117	 1. PPT 작성 - 다음주 멘토링 회의를 위해 미리 PPT 작업을 진행. - 구글 드라이브 슬라이드로 동시 작업 2. 음성 분석 모델 문제점 개선 방법 의논 - 학습 데이터와 테스트 데이터의 언어가 다른 것에서 오는 문제. 3. 영상 모델의 얼굴 추출 성공률 개선 방법 의논	
8	2025. 11. 19. (19:00~22:00) / 디지털도서관		

		- 영상 데이터 전처리  1. 발표 대본 작성 - 다음주 멘토링 회의를 위해 미리 발표 대본 작업을 진행. 2. 모델 개선 - 영상 모델 얼굴 추출 성공 비율 향상을 위한 데이터 전처리 작업	
9	2025. 11. 24. (19:00~22:00) / 디지털도서관	  1.발표 준비 - 발표 대본 작성 - 모의 발표 2. 멘토링 준비 - 질문 사항 정리 - ppt 페이지별 간단 설명 정리	

7. 프로젝트 세부일정 및 내용

		9월				10월				11월				12월			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
박하늘	영상 파싱 모델 설계																
	dlib & CNN 기반 설계																
	웹 설계 및 개선																
	회의록 정리																
유주형	Deepface&CNN 기반 설계																
	Video CNN 모델 설계 및 개선																
	Fusion 모델 설계 및 개선																
허성민	MediaPipe&Random Forest 기반 설계																
	Audio CNN 모델 설계 및 개선																
	Fusion 모델 설계 및 개선																
공통	발표자료 수집 정리																

8. 결과물에 대한 향후 활용계획

(1) 성과 관리 추진체계

본 연구 프로젝트를 통해 산출된 소프트웨어 코드, 알고리즘, 데이터셋 등의 핵심 자산을 안정적으로 보존하고 지속 가능한 발전을 도모하기 위해 체계적인 관리 시스템을 구축하였다. 우선 소프트웨어 형상 관리(SCM) 측면에서는 GitHub를 활용하여 개발, 테스트, 배포 단계를 엄격히 구분하는 브랜치 전략을 도입함으로써 코드의 무결성을 확보하였다. 특히 영상 전처리, 음성 특징 추출, 모델 학습 등 각 기능 모듈을 독립적으로 관리하고, 실험 단계부터 최종 퓨전(Fusion) 모델에 이르는 가중치(Weight) 파일을 메타데이터와 함께 아카이빙하여 연구의 재현성을 보장하고 있다.

나아가 본 연구의 독자적 성과물인 '경량화 3D-CNN 구조'와 '멀티모달 퓨전 파이프라인'에 대한 특허 출원과 핵심 알고리즘의 소프트웨어 저작권 등록을 추진하여, 기술적 배타성을 확보하고 지식재산권을 보호할 계획이다.

(2) 예상되는 연구개발성과의 활용분야

본 연구에서 개발한 멀티모달 거짓말 탐지 시스템은 진위 여부 판별이 필수적인 다양한 산업군에서 핵심 솔루션으로 활용될 수 있다. <딥페이크 영상을 구분하는 기술을 넣어준다> 가장 일차적으로는 유튜브나 틱톡 등 뉴미디어 플랫폼에서 확산되는 딥페이크 영상이나 가짜 뉴스를 자동으로 필터링하는 미디어 포렌식(Media Forensics) 분야에 적용 가능하다. 또한, 급성장하는 라이브 커머스 시장에서는 판매자의 상품 설명에 대한 과장이나 허위 사실 여부를 실시간으로 모니터링하여 소비자에게 신뢰 지표를 제공하는 안전장치로 기능할 수 있다.

이외에도 수사기관에서는 피의자 심문 시 비접촉식 진술 분석 도구로 활용하여 수사의 효율성을 높일 수 있으며, 금융 및 보험 분야에서는 비대면 실명 인증이나 보험 청구 인터뷰 영상

의 진실성을 검증하여 금융 사기를 예방하는 인슈어테크(InsurTech) 기술로도 폭넓게 응용될 것으로 기대된다.

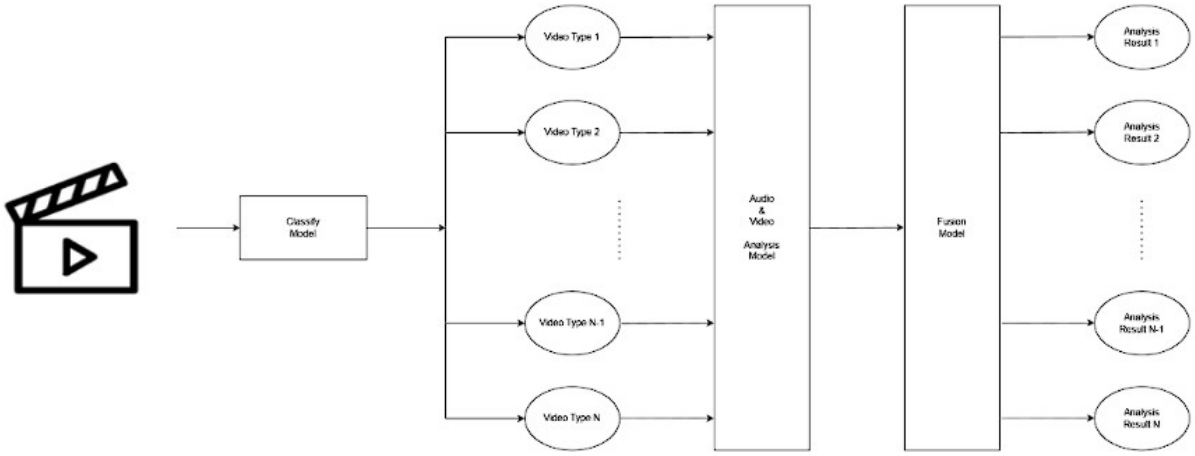
(3) 활용방안

개발된 기술을 실제 서비스 환경에 효과적으로 적용하기 위해 다각적인 활용 방안을 수립하였다. 먼저, 시스템의 접근성을 높이기 위해 기 구축된 로컬 서버를 클라우드 기반의 API 서비스(SaaS)로 확장할 계획이다. 이를 통해 외부 플랫폼들이 영상 파일이나 URL을 전송하면 즉각적으로 거짓말 확률과 분석 리포트를 제공받을 수 있도록 지원하며, 특히 긴 영상의 경우 문장 단위 파싱 기능을 통해 핵심 구간에 대한 타임스탬프 기반 분석 결과를 제공하여 사용자 편의성을 극대화한다.

보안이 중요시되는 수사기관이나 공공기관을 위해서는 경량화된 모델을 탑재한 독립형(On-Premise) 소프트웨어를 패키징하여 배포함으로써, 인터넷 연결이 제한된 환경에서도 실시간 분석이 가능하도록 지원한다. 아울러 일반 대중의 디지털 리터러시를 보조하기 위해 웹 브라우저 확장 프로그램을 개발하여, 사용자가 온라인 영상을 시청할 때 직관적인 신뢰도 신호(Signal)를 실시간으로 확인할 수 있는 B2C 서비스로도 활용 영역을 넓혀갈 것이다.

(4) 추가연구의 필요성

현재 달성한 기술적 성과를 바탕으로 상용화 수준의 완벽한 신뢰도를 확보하기 위해서는 다각적인 후속 연구가 필수적이다. 우선 현 단계에서는 문장별 파싱으로 분석 속도 문제를 해결하였으나, 라이브 스트리밍과 같은 초저지연(Ultra-low Latency) 환경에 대응하기 위해서는 엣지 컴퓨팅(Edge Computing) 기술을 도입하여 모델을 더욱 경량화하고 최적화하는 연구가 선행되어야 한다.



아울러 영상 콘텐츠의 장르(뉴스, 예능, 인터뷰 등)에 따라 화자의 발화 톤과 제스처 등 행동 양식이 현격히 달라지는 점을 고려하여, 거짓말 탐지 수행 전 영상의 종류를 자동으로 식별하는 사전 분류 모델(Pre-classification Model)의 도입이 필요하다. 이를 통해 영상의 분위

기와 맥락(Context)에 맞춰 최적화된 가중치를 적용함으로써, 다양한 영상 환경에서도 분석의 정확도를 획기적으로 높일 수 있을 것이다.

또한 AI의 판단 결과를 사용자가 신뢰할 수 있도록 설명 가능한 AI(XAI) 기술을 도입해야 한다. 단순히 거짓 여부를 확률로 제시하는 것을 넘어, 눈 깜빡임의 변화나 음성 피치의 불안정성 등 구체적인 판단 근거를 시각화하여 제공함으로써 시스템의 설득력을 높여야 한다. 더불어 한국어 데이터뿐만 아니라 다양한 언어와 인종, 연령대별 데이터를 추가로 학습시켜 모델의 편향을 줄이고 일반화(Generalization) 성능을 강화하는 작업도 지속적으로 이루어져야 한다.

(5) 타 연구에의 응용

본 프로젝트를 통해 확보한 요소 기술들은 거짓말 탐지 분야를 넘어 다양한 연관 분야로 기술 전이가 가능하다. 영상과 음성의 미세한 변화를 감지하는 멀티모달 분석 기술은 우울증이나 불안 장애를 조기에 진단하는 비대면 정신 건강 케어(Mental Health Care) 서비스로 응용될 수 있으며, 온라인 교육 환경에서는 학생의 표정과 시선 처리를 분석하여 집중도를 측정하는 에듀테크(EduTech) 기술로도 활용 가치가 높다. 또한, 본 모델의 판별자(Discriminator) 기능은 생성형 AI가 만들어낸 결과물의 자연스러움을 평가하고 튜닝하는 품질 평가 지표로 역이용될 수 있어, AI 생성 콘텐츠 산업의 발전에도 기여할 수 있다.

(6) 기업화 추진방안

본 연구 성과의 사업화 전략은 영리 추구보다는 '디지털 범죄 예방'과 '정보 신뢰성 회복'이라는 사회적 공헌에 최우선 가치를 둔다. 따라서 경제적 여건과 관계없이 누구나 안전한 디지털 환경을 누릴 수 있도록, 개발된 솔루션을 일반 대중에게 무료로 배포하는 것을 기본 원칙으로 한다. 이를 통해 서비스의 접근성을 극대화하여 딥페이크나 피싱 범죄에 취약한 계층까지 포괄하는 보편적 사회 안전망으로서의 기능을 수행하고자 한다.

서비스 유지보수 및 트래픽 처리에 소요되는 운영 비용은 사용자에게 과금하는 대신, 웹 인터페이스나 애플리케이션 내 광고(Ad-supported Model) 및 팝업 배너 등을 활용하여 충당함으로써 지속 가능한 운영 구조를 확립할 계획이다. 초기에는 이러한 광고 기반의 민간 공익 서비스 형태로 운영하여 실제 범죄 예방 사례와 사용자 데이터를 축적하고, 향후 시스템의 실효성이 입증되는 단계에서는 이를 국가적 차원의 핵심 사업으로 발전시킬 가능성을 모색한다. 디지털 성범죄 및 금융 사기가 심각한 사회 문제로 대두되는 만큼, 정부(과학기술정보통신부, 경찰청 등) 주도의 '대국민 디지털 안심 서비스' 또는 공공 인프라 사업으로 제안하여 국가 예산 지원을 통한 안정적인 공공 서비스로의 전환을 최종 목표로 추진한다.

(7) 기술이전 계획

직접적인 사업화 외에도 보유 기술의 이전을 통해 수익 구조를 다각화할 계획이다. 핵심 기술인 '경량화 3D-CNN 모델 아키텍처'는 고성능 장비 도입이 어려운 보안 업체 등에 라이선싱 형태로 제공하고, 영상과 음성 데이터를 정교하게 결합하는 '멀티모달 퓨전 전처리 모듈'은 AI 솔루션 전문 기업에 기술 이전을 추진한다. 이를 위해 대학 산학협력단이나 기술보증기금 등 전문 중개 기관을 통해 기술 가치 평가를 공인받고, 관련 기업을 대상으로 기술 설명회를 개최하여 수요처를 적극적으로 발굴할 예정이다.

9. 참고자료 및 문헌

Audio-Visual Deception Detection: DOLOS Dataset and Parameter-Efficient Crossmodal Learning. International Conference on Computer Vision (ICCV). - Guo, X., Selvaraj, N.M., Yu, Z., Kong, A., Shen, B. and Kot, A.