

2024년 경산이노베이션아카데미 미래 융합 인재 발굴 SW경진대회  
아 이 디 어 소 개 서

팀명

인사이드아웃

아이디어명

SER(Speech Emotion Recognition)을 이용한 CAPTCHA

참가 동기,  
활동 배경

## I 현재 상황

인터넷 사용량이 급증함에 따라 봇(bot)에 의한 악성 활동이 증가하고 있습니다. Imperva Bad Bot Report 2024에 따르면, 전 세계 웹 트래픽의 약 49.6%가 봇에 의해 생성되며 이 중 약 32%가 악성 봇에 해당한다고 보고되었습니다[1].

봇으로 인한 대표적인 피해로는 계정 탈취, 스팸 게시물 작성, DDoS 공격 등이 있고, 최근 생성형 AI(Artificial Intelligence)의 폭발적인 성장은 상황을 더욱 악화시키고 있습니다. 실제로 계정 탈취 공격은 2022년 대비 2023년 10% 증가했으며, API(Application Programming Interface) 엔드포인트를 노린 공격은 2023년 44%로 2022년 대비 35% 증가했습니다[1]. 봇에 의한 공격을 방지하기 위해 다양한 CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart)가 개발 및 사용되고 있습니다.

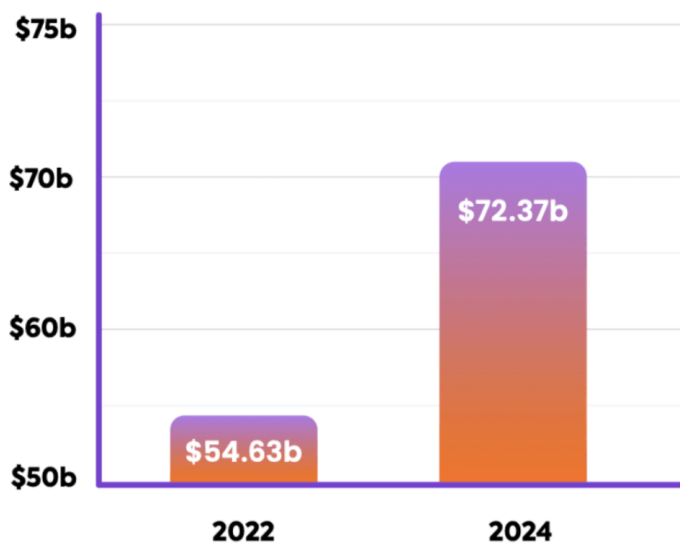
## II 문제 도출

현재는 주로 텍스트, 이미지, 체크박스, 상호작용 기반 CAPTCHA가 사용되고 있습니다. 그러나 AI의 이미지 인식 및 자연어 처리 능력이 급격히 향상됨에 따라 기존의 텍스트나 이미지 기반 CAPTCHA는 더이상 효과적으로 봇을 막을 수 없게 되었습니다. 실제로 2014년 구글에서 개발한 'reCAPTCHA'는 딥러닝 기반의 봇에 의해 99.8%의 정확도로 우회되어 현재 사용이 중지되었습니다[2]. 2021년에는 구글에서 개발한 'reCAPTCHA v3'마저 딥러닝 기반 봇에 의해 우회될 수 있음이 확인되었습니다[3]. 이는 현재의 CAPTCHA 시스템이 더이상 사람과 봇을 효과적으로 구분하지 못함을 보여줍니다. 따라서 봇으로 인한 피해를 최소화하기 위해서는 현재의 AI기술로도 쉽게 극복할 수 없는 새로운 형태의 CAPTCHA가 필요하다고 판단하였습니다.

### Ⅲ 고객 분석

본 프로젝트의 주요 고객은 악성 봇으로 피해를 입는 광고주 및 플랫폼 기업입니다.

LUNIO의 보고서에 따르면, 2024년 Google 에서만 광고주가 광고 사기 봇으로 인한 IVT(무효트래픽)으로 인해 165억 9천만 달러의 손실을 입을 것으로 예상되며 소셜 미디어 전체에서는 547억 8천만 달러의 손실을 입을 것으로 추정됩니다. 전반적으로 2022년 대비 2024년에 낭비되는 광고 지출은 72.37억 달러(한화 약 98조 원)로 예상된다고 합니다[4].



새로운 CAPTCHA의 개발은 자동화된 트래픽 감소를 통해 불필요한 사회적 비용 낭비를 막을 수 있으며 추가적으로 시각장애인의 접근성을 개선하여 인터넷 사용에 따른 사용성 개선에 큰 도움이 될 것이라고 생각됩니다[5].

### Ⅳ 아이디어 도출

본 프로젝트의 핵심 아이디어는 AI 기술의 한계를 이용하는 것입니다. 현재 딥러닝 기반의 음성 감정 인식 기술은 70% 내외의 정확도를 보이는 것으로 알려져 있습니다[6]. 반면 사람은 음성만으로도 화자의 감정을 90% 이상의 정확도로 파악할 수 있습니다. 이러한 차이를 이용하여, 음성을 통해 사람과 봇을 구분하는 새로운 형태의 E-CAPTCHA(Emotion CAPTCHA)를 개발하고자 합니다. 이는 사람에게는 쉬운 작업이지만, 현재의

	<p>AI 기술로는 정확하게 수행하기 어려운 과제가 될 것입니다.</p> <p><b>[참고문헌]</b></p> <p>[1] "Bad Bot Report", Imperva, 2024.</p> <p>[2] "I'm not a human: Breaking the Google reCAPTCHA", S. Sivakorn et al., 2016.</p> <p>[3] <a href="https://m.boannews.com/html/detail.html?idx=93947">https://m.boannews.com/html/detail.html?idx=93947</a></p> <p>[4] <a href="https://lunio.ai/blog/ad-fraud/traffic-bots/">https://lunio.ai/blog/ad-fraud/traffic-bots/</a></p> <p>[5] "로봇이 아닙니다"...캡차 인증, 시각장애인은 할 수 없다?, SBS, 2019.</p> <p>[6] "In-depth investigation of speech emotion recognition studies from past to present –The importance of emotion recognition from speech signal for AI–", Intelligent Systems with Applications</p>
<p><b>주요기능과 특징, 기대효과</b></p>	<p><b>V 주요 기능</b></p> <p>E-CAPTCHA의 기능은 다음과 같습니다. 캐글의 RAVDESS 데이터셋(감정이 태깅되어 있음)을 딥 러닝한다면 해당 감정의 TTS(Text To Speech)를 수행할 수 있습니다. 학습된 TTS를 통해 감정(예: 기쁨, 슬픔, 분노 등)이 표현된 음성 데이터를 생성한 후 로그인 사용자에게 음성을 제시합니다. 그리고 사용자가 해당 음성을 듣고, 그 감정의 일치 여부를 판단합니다.</p> <ol style="list-style-type: none"> <li>판단 결과 일치 <ul style="list-style-type: none"> <li>추가적인 더미 질문을 통해 감정 인식률이 낮은 음성에 대한 사용자들의 라벨링 유도</li> </ul> </li> <li>판단 결과 불일치 <ul style="list-style-type: none"> <li>봇으로 판단하여 접근 거부</li> </ul> </li> </ol> <p><b>VI 기대 효과</b></p> <p>Emotion CAPTCHA는 크게 세 가지 측면에서 기대 효과를 가집니다.</p> <p>첫째, 봇에 의한 악성 활동을 효과적으로 방지할 수 있습니다. 이를 통해 웹 사이트 운영자는 봇으로 인한 피해를 크게 줄일 수 있습니다. 앞으로의 광고 시장이 커짐에 따라 IVT 및 , SIVT(정교한 무효트래픽)을 판단하</p>

는 것이 더욱 중요해질 것입니다. 실제로 2009년 카네기 멜론 대학교에서 연구 개발된 reCAPTCHA의 경우 2009년 구글에 인수 되었으며 추정가격은 330억입니다. E-CAPTCHA가 적용되어 봇의 트래픽 공격의 정확도가 99% 정도에서 70% 정도로 개선된다면 98조원의 트래픽 시장에서 약 25조원의 금액 손해를 예방하여 플랫폼 기업의 광고 시장에 대한 신뢰도가 개선될 것으로 예상합니다.

둘째, 시각장애인의 사용자 편의성이 높아집니다. 텍스트를 입력하거나 알아보기 힘든 이미지를 판독하는 번거로움 없이, 음성만으로 간편하게 인증을 완료할 수 있습니다.

셋째, 음성 감정 인식 AI 기술 발전에도 기여할 수 있습니다. Emotion CAPTCHA를 통해 라벨링된 음성 데이터는 ML 학습의 기반이 될 수 있으며, 이는 더 정교한 감정 인식 모델 개발로 이어질 수 있습니다. 나아가 AI 나레이션, 영화 제작 등에 활용되는 AI 음성이 더 사실적으로 묘사될 수 있을 것으로 기대됩니다. 더미 데이터 테스트를 통해 생성된 저희만의 데이터 셋은 높은 잠재적 가치를 지닐 것입니다.

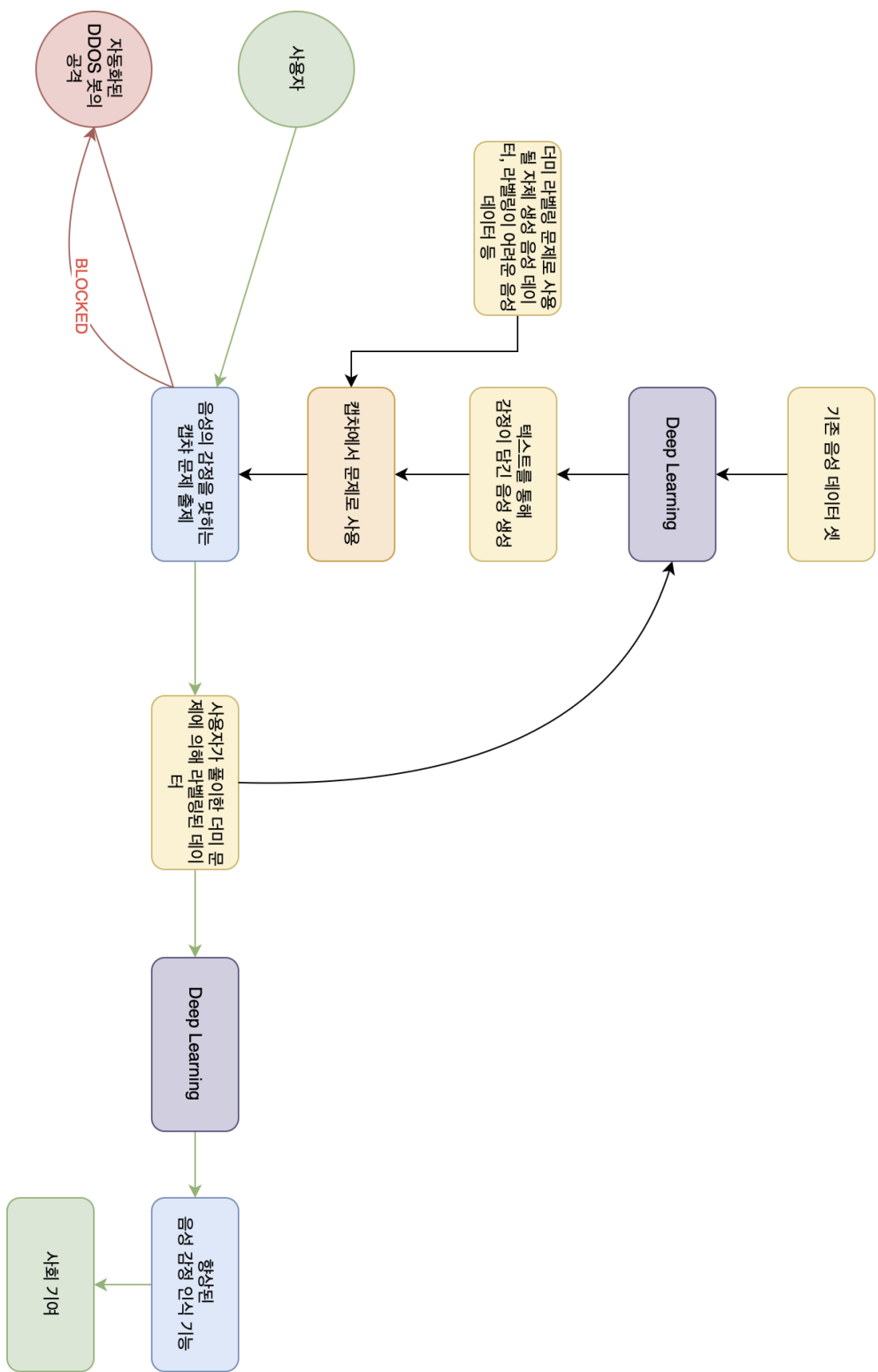


그림 1. [전체 구상도]

작품  
구상도(사진)  
(아이디어를  
쉽게 설명할 수  
있는 구상도 등)





  
  
  
[Fogot password?](#)  

그림 2. 로그인 화면

2 / 3

Play audio and select proper emotion.

anger	fearful	disgust	sad
neutral	calm	happy	surprised

그림 3. E-CAPTCHA 화면