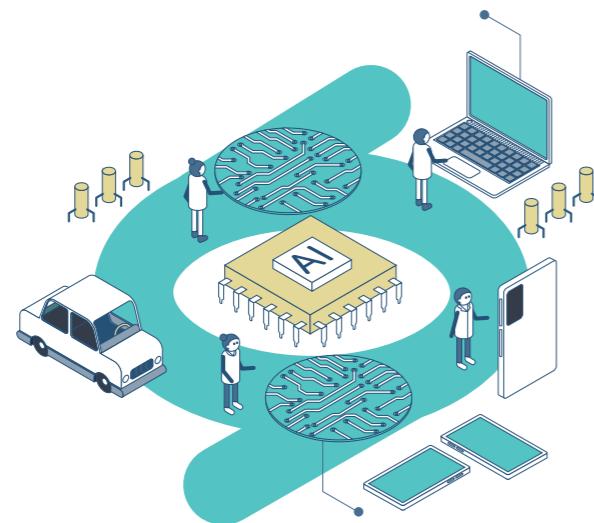


2023

데이터 청년 캠퍼스

— 자연어처리 기반 딥러닝기술 융합과정 —



과학기술정보통신부



K data 한국데이터산업진흥원

한국외국어대학교 AI 교육원

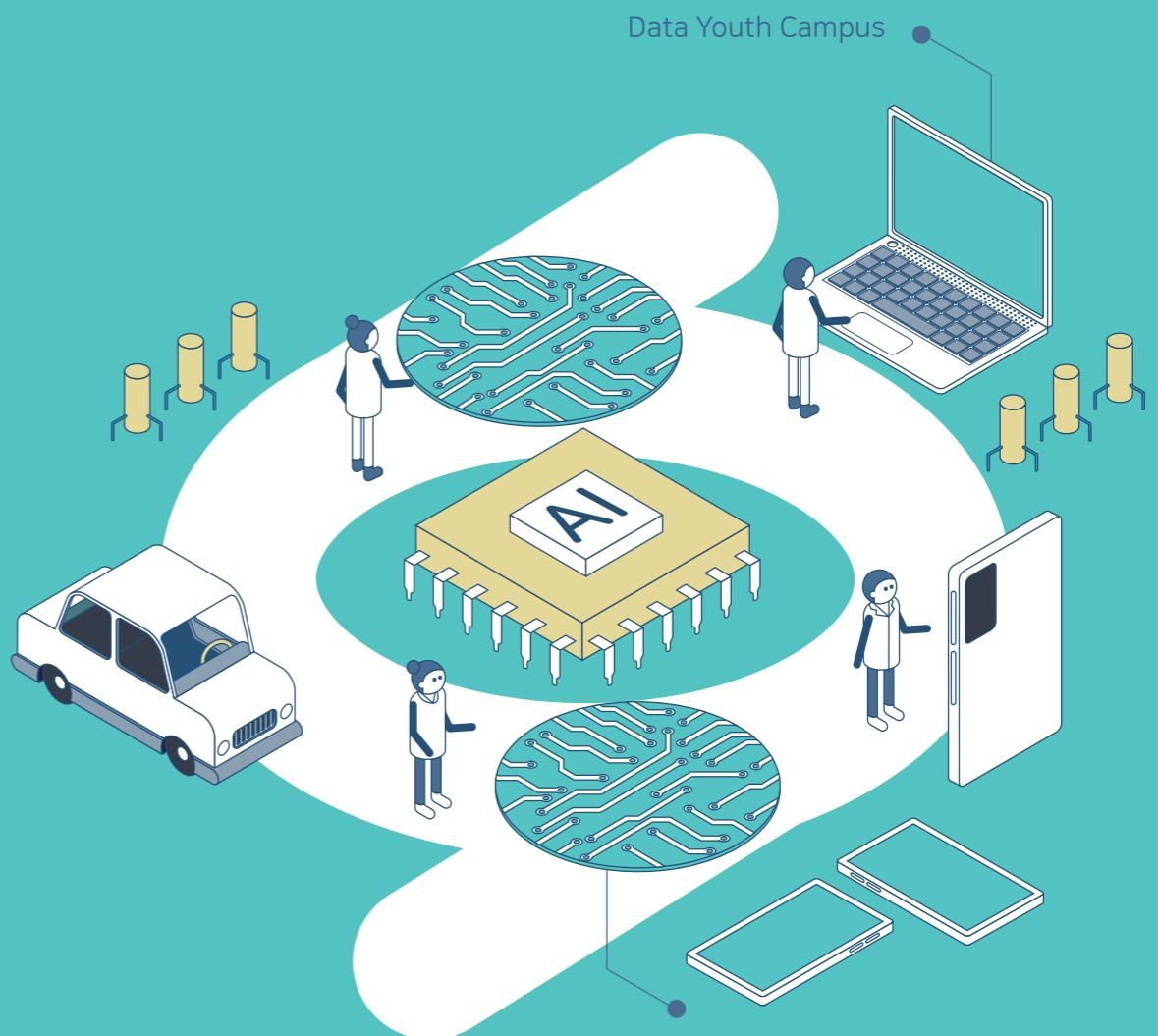
17035 경기도 용인시 처인구 모현읍 외대로 81

한국외국어대학교 글로벌캠퍼스 AI교육원(공학관 211-1호)

Engineering Building Gloal Campus , 81, Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin-si,

Gyeonggi-do, Korea, (17035) Hankuk University of Foreign Studies

Tel. 031-8020-5123 | FAX. 031-330-4120 | E-mail. swcu@hufs.ac.kr



한국외국어대학교 AI 교육원

과학기술정보통신부

K data 한국데이터산업진흥원

2023

데이터 청년 캠퍼스

자연어처리 기반 딥러닝기술 융합과정



과학기술정보통신부



K data 한국데이터산업진흥원

한국외국어대학교 AI 교육원

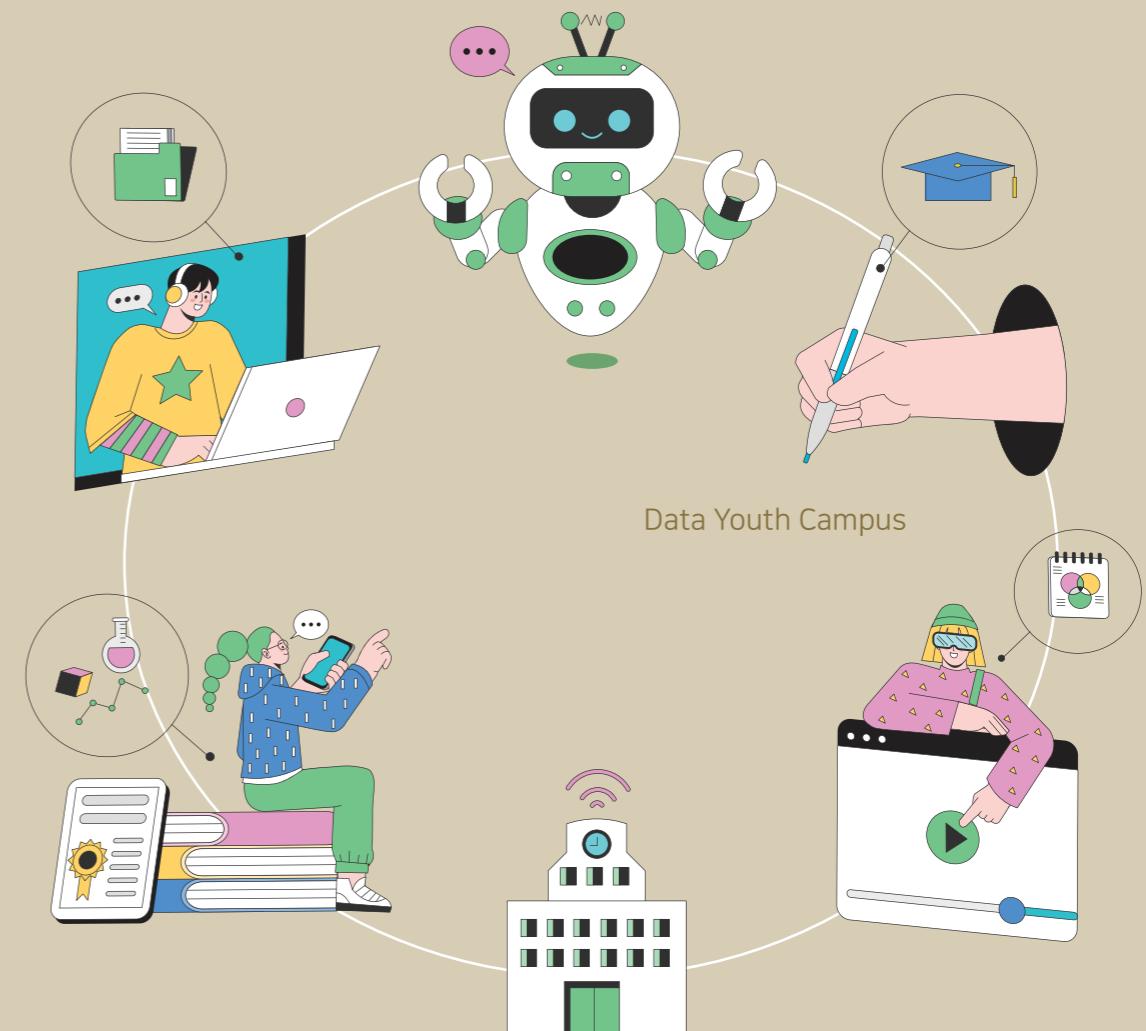
17035 경기도 용인시 처인구 모현읍 외대로 81

한국외국어대학교 글로벌캠퍼스 AI교육원(공학관 211-1호)

Engineering Building Gloal Campus , 81, Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin-si,

Gyeonggi-do, Korea, (17035) Hankuk University of Foreign Studies

Tel. 031-8020-5123 | FAX. 031-330-4120 | E-mail. swcu@hufs.ac.kr



Data Youth Campus

한국외국어대학교 AI 교육원



과학기술정보통신부



K data 한국데이터산업진흥원

01**데이터 청년 캠퍼스 소개**

데이터는 기존 산업의 혁신은 물론 신산업 창출 등 경제성장의 핵심 동력으로 미래 국가 경쟁력을 좌우하는 주요 자원으로 인식하고 해외 주요국은 데이터 기술 발전 및 시장 확대에 따라 데이터 활용의 기반이 되는 전문인력 양성을 위해 전략을 수립하고 투자를 확대 추진하고 있습니다.

또한 산업 수요 대비 국내 데이터 전문인력은 양적, 질적으로 부족하며 현장에서 필요로 하는 실무형 데이터 전문 인력의 미스매치 발생하고 있습니다. 이에, 과학기술정보통신부와 한국데이터산업진흥원은 산업에서 필요로 하는 데이터 고급인재 양성을 위해 데이터 관련 학부 전공자, 석·박사 등 청년 대상 융합 프로젝트 기반의 데이터 전문교육 및 일자리 연계를 지원하고 있습니다.

02**지원 대상**

- 데이터 관련 전공자(3~4학년, 졸업유예생 포함)
- 만 34세 이하의 미취업 청년('87.1.1 이후 출생자)
- 데이터와 컴퓨터공학 관련 복수전공, 연계전공, 융합전공 지원가능. 단, 면접후 선정
- 교육인원: 70명
- 우대사항: IT관련 프로젝트 수행 또는 IT관련 업체 경력자(1년 이상)
- 준비사항: 노트북 지참 권장(CPU i5, 메모리 8G 이상 권장)

03**교육 기간**

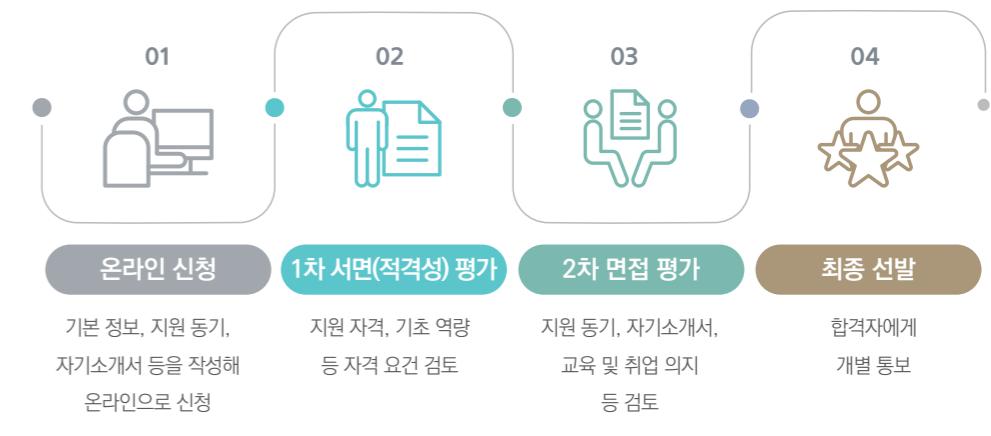
- 이론실습: 2022. 6. 27 ~ 7.22 / 10:00~18:00
(교육장: 한국외국어대학교 서울캠퍼스 사회과학관 206호, 사이버관 412호)
- 프로젝트 기간: 2022. 7. 25 ~ 8. 31 / 10:00~18:00
(팀선정 및 친교 프로그램, 기초 프로젝트 교육 등 수업에 따라 일정 변동 가능)
- 교육 모집 기간: 5/16 ~ 6/13

04**지원 내용**

- 언어공학 중심의 전문교과 과정 : 전문적인 언어학이론 수업은 영어 중심으로 진행되며 실용성을 바탕으로 컴퓨터언어학을 교육. 멀티모달 학습으로 교육 범위 확대
- 캐글(Kaggle) Competition을 이용한 팀 프로젝트 실습 : 본 사업에서 선정한 15개의 Competition 중 의무적으로 3개를 선택해 팀 프로젝트로 문제 해결
- 기업 멘토와 연계한 프로젝트 기반 실무 교육 : 학계/산업계 전문인력 Pool과 연계하여 단순 실습이 아닌 완성도 있는 실무형 프로젝트를 멘토와 진행하여 기획부터 수행까지 진행함으로써 취업 연계 가능

※ 교육과정 참여 혜택

	교육비 전액 지원		맞춤형 채용 정보 제공
	SQLD, ADsP 등 데이터 분야 전문 자격증 취득 지원		우수 프로젝트 팀은 경진대회를 통해 과학기술정보통신부장관상, 한국데이터산업진흥원 원장상에 선정될 수 있는 기회를 부여

05**선발절차****06****교육カリ큘럼**

구분	교육분야	상세내용	강사	시간	시설·장비	교·보재
이론	언어공학개론	언어공학 분야에 대한 개요와 필요성, 언어학과 언어공학, 데이터 처리에 필요한 컴퓨터 정보 처리 이론	김진아 (장지원)	21	프로젝터	자체교재
이론/ 실습	머신러닝/ 딥러닝	머신러닝 알고리즘에 대한 기본 이론 뿐만 아니라 데이터 전처리, 차원 축소, 모델 평가 등에 대해 학습 딥러닝에 대한 기초적인 이론과 실제 학습. 표준 신경망 모델, 학습기법, 영상·언어·음성 인식 문제의 기초 연구	한민규 (박민규)	35	프로젝터, PC, 서버	머신러닝 교과서
이론/ 실습	자연어처리	자연어처리에 대한 통계 기법 등을 활용하여 형태소 분석, 구문 분석, 의미 분석, 담화 분석에 이르는 언어 처리 과정 및 모델을 학습	고원희 (백선희)	42	프로젝터, PC, 서버	개발교재 (자연어처리)
실습	캐글 실습	캐글에서 초보자 대상의 학습용 Competition을 이용하여 주제선정과 프로젝트 진행 능력을 학습	박민규 (김경섭)	31	프로젝터, PC, 서버	자체교재
실습	기초 프로젝트	팀선정, 주제선정, 요구사항 분석, 기초 설계	오세종 (이지효)	31	프로젝터, PC, 서버	
실습	심화 프로젝트	산업체 멘토와 연계한 산업체 관점의 산학연계 프로젝트 수행	두일철 외 4인	210	프로젝터, PC, 서버	
평가 보완	프로젝트 보완	프로젝트 발표식때 수정보완 지시에 따른 프로젝트 보완 실행	공동	0	PC, 서버	
합계				350		

PROJECT 01

생태계 교란조



AI 기반 피싱 탐지 서비스

산업체 및 멘토 교수: 트위그팜·김현아 매니저, 두일철·오세종·박민규 교수(한국외국어대학교 AI교육원) / 팀장: 홍수빈(고려대학교 금융보안학과)
팀원: 백지현(광운대학교 정보융합학과), 이규영(동국대학교 정보통신공학과), 장명재(세종대학교 컴퓨터공학과), 정희수(한국외국어대학교 통계학과)

개발동기 및 목적

개발동기 및 목적

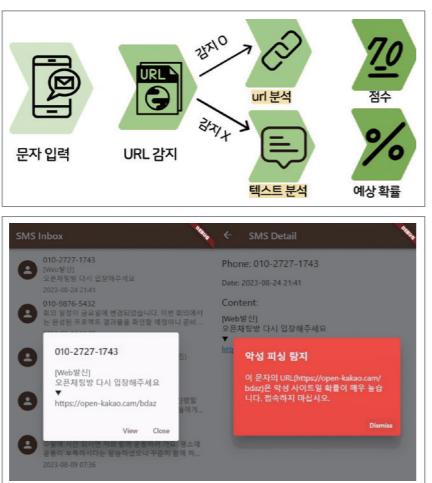
불특정 다수의 개인정보를 탈취해 금전을 요구하는 사기 수법인 피싱(Phishing)이 사회적으로 여전히 기승을 부리고 있다. 특히 금융감독원에 따르면 코로나 이후 피싱으로 인한 피해가 더욱 증가했다고 밝혔다. 최근 4년 사이 스미싱으로 인한 피해액이 21배나 증가했고 보이스피싱으로 인한 피해액 또한 날이 갈수록 증가하는 추세다. 하지만 기존에 존재하는 상용 앱인 후후(WhoWho)나 알약미 등 특정 상황에서 탐지할 수 없다는 문제를 확인했다. 따라서 이러한 문제를 해결하고자 인공지능 모델을 활용해 스팸, 스미싱 문자와 보이스피싱 통화를 탐지하고자 한다.

개발목표

텍스트피싱 탐지 기법은 다음과 같다. SMS 텍스트 내 접속 가능한 정상적인 URL이 존재하면 URL을 파싱한다. 존재하지 않는다면 SMS 텍스트를 스팸 분류 인공지능 모델(KoBERT)에 서빙한다. 파싱한 URL은 1차로 사전 구축된 서버에서 블랙리스트 DB를 대조하고, 2차로 웹 소스코드가 피싱 사이트임을 확인할 수 있는 여러 보안 평가항목에 따라 분석한다. LLM인 GPT-3.5 Turbo 모델에 미리 주입된 풀프로토콜에 따라 웹 소스코드 악성 여부 평가를 요구해 각 항목당 점수를 반환한 뒤 총점을 도출한다. SMS 텍스트에 실제로 접속할 수 있는 정상적인 URL이 존재하지 않는다면, 인공지능 모델(KoBERT) 및 Attention+LSTM 양상들을 통해 텍스트에서 스팸 문자의 확률값을 도출한다. 보이스피싱 부분은 음성 대화 문맥을 인공지능 모델(KoBERT)을 활용해 실시간으로 분석한다. 이어 미리 학습된 모델에서 보이스피싱 문맥과 얼마나 유사한지 확률값을 도출한다. 결과적으로 고도화된 인공지능 모델을 활용해 스팸 및 스미싱 문자와 보이스피싱을 탐지한다.

최종결과물

01 텍스트 피싱 탐지



02 보이스 피싱 탐지

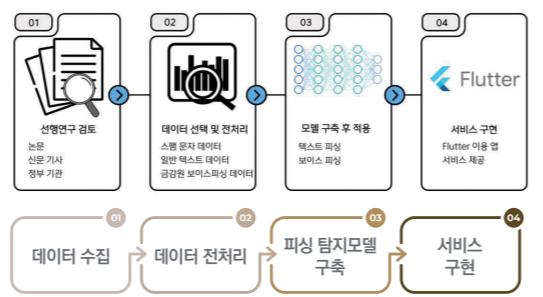


추진배경

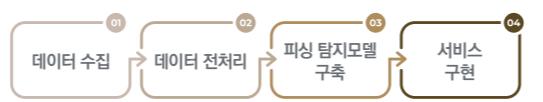
불특정 다수의 개인정보를 탈취해 금전을 요구하는 사기 수법인 피싱(Phishing)이 사회적으로 여전히 기승을 부리고 있다. 특히 금융감독원에 따르면 코로나 이후 피싱으로 인한 피해가 더욱 증가했다고 밝혔다. 최근 4년 사이 스미싱으로 인한 피해액이 21배나 증가했고 보이스피싱으로 인한 피해액 또한 날이 갈수록 증가하는 추세다. 하지만 기존에 존재하는 상용 앱인 후후(WhoWho)나 알약미 등 특정 상황에서 탐지할 수 없다는 문제를 확인했다. 따라서 이러한 문제를 해결하고자 인공지능 모델을 활용해 스팸, 스미싱 문자와 보이스피싱 통화를 탐지하고자 한다.

프로젝트 추진 내용

프로젝트 진행과정



프로젝트 구현과정



기대효과

경제적 측면

프로젝트의 핵심인 고도화된 피싱 탐지 기법은 기존 상용 애플리케이션에 비해 훨씬 뛰어난 탐지 능력을 가졌다. 이는 사용자들이 피싱으로 인한 경제적 손실을 크게 줄일 수 있게 해주며, 덕분에 개인의 금융 안정성이 강화된다. 더 나아가, 보이스피싱 및 스팸 문제도 줄어들면 그로 인한 분쟁과 보상, 그리고 피해 예방을 위한 시스템 개발 및 유지 관리 비용 등 금융기관의 부담도 크게 감소할 것이다. 이러한 절감 효과는 궁극적으로 관련 수사 및 재판에 드는 비용과 시간 역시 줄어들게 된다. 법원 및 수사기관의 부담 감소와 함께 사회 전체의 자원 활용 효율성도 증대될 것으로 예상된다.

사회적 측면

AI 기반 피싱 탐지 서비스는 보안 기술과 제도 발전에 중요한 역할을 할 것이다. 기존 상용 앱 대비 실질적인 피싱 탐지 메커니즘이 구축되면서 새로운 형태의 사기 및 보안 위협에 대응하는 능력도 강화된다. 또한 정보 유출 위험이 줄어들고 개인정보보호는 더욱 강화되며, 소비자들은 다양한 서비스나 제품을 걱정 없이 이용할 수 있게 될 것이다. 특히 디지털 소외계층인 노인 등에게는 정보 접근성을 크게 높여주며 디지털 사각지대를 해소하는데 중요한 도구가 될 것이다. 또한, 스팸과 피싱을 잘 탐지해 사고를 예방하거나 줄이는 것은 다른 유형의 범죄나 사기를 예방하는 데도 긍정적인 영향을 미칠 것이다. 사이버 범죄의 경우 공격 기법이 빠르게 진화하고 있기 때문에, 이러한 AI 기반 탐지 기법은 다양한 형태의 사기나 보안 위협에 대응하는 능력을 갖추는데 중요한 발판이 될 것이다. 결국, 이 프로젝트는 개인의 정보보호와 사회 안전성 강화, 그리고 경제적 효율성 증대라는 세 가지 주요 목표를 동시에 달성하는 방안을 제시한다. 이는 전체 사회적 분위기를 긍정적으로 바꾸며, 디지털 리터러시와 정보보호 의식 향상을 통해 우리 사회가 건강한 정보사회로 한 걸음 더 나아가는데 크게 기여할 것으로 예상된다.

산학협력

팀원활동

이규영: Quality Assurance, 발표자료 제작 및 최종 발표
백지현: KoBERT 및 Attention+LSTM 모델링, 논문 리서치
장명재: KoBERT 및 (Bi)LSTM 모델링, 백엔드 및 플러터 개발
정희수: 데이터 전처리, Whisper 모델 파인튜닝, 논문 리서치
홍수빈: LLM Application 및 백엔드 개발, 프로젝트 총괄

멘토활동

조원별 RnR 및 구체적인 서비스 기능 재정의
프로젝트 간 활용할 AI 모델 정의 및 학습 데이터셋 피드백
실시간 보이스피싱 탐지기술 고도화 관련 주가 리서치 결과 공유
심화 프로젝트 최종 발표자료 및 리허설 피드백



PROJECT 02

다잇소



BridZe : 베트남 다문화가정 아동을 위한 언어와 정서 길잡이 서비스

산업체 및 멘토 교수: (주)하니엠앤비·김석원 대표, 두일철·오세종·박민규 교수(한국외국어대학교 AI교육원) / 팀장: 양체연(한국외국어대학교 언어인지과학과), 팀원: 권영후(한국외국어대학교 국제통상학과), 박현빈(한국외국어대학교 영어통번역학과), 정성욱(한국외국어대학교 컴퓨터공학부), 한건희(고려대학교 컴퓨터학과)

개발동기 및 목적

개발동기 및 목적

베트남 다문화가정의 아동들은 언어 장벽의 문제로 인하여 부모와의 유대관계 형성의 어려움, 자기 존중감 저하 등 다양한 정서 문제를 겪고 있다. 뿐만 아니라 학업 성적 부진, 학급 내 소외 현상과 같이 좋지 못한 환경에 처해있다. 더욱이 심각한 문제는 정보 소외 대상자이자 디지털 소외 계층인 다문화가정의 부모님들은 한국 부모님들과 달리 아동의 문제를 정확히 묘사하는 것에 어려움을 겪고 있으며 해결책을 찾지 못하는 경우가 많다는 것이다. 아동기의 언어와 정서 문제는 성인기까지 이어지기 때문에 문제와 문제 해결방법을 몰라 아동기를 헛되이 보내는 일이 없도록 '아동 관찰, 정보 수집, 방대한 정보 중 정보 선별'의 과정을 대신하는 서비스를 개발하고자 한다.

최종결과물

01 언어 평가

아동의 나이별 언어 습득 수준에 맞는 지정 발화를 읽어 베트남인의 한국어 발화에 특화된 whisper 모델을 통해 발음과 읽기 능력을 상, 중, 하로 평가받게 된다.

언어 평가의 임계치는 AI HUB에 있는 한국 아동의 발화 데이터 점수를 활용해 통계 기법으로 설정. 베트남인 부모님 역시 동일한 평가 방식으로 언어 수준을 평가하여 하, 중의 평가를 받은 경우에는 비대면 교육 프로그램을 추천받게 된다.



02 정서 평가

유대관계 테스트에 답하여 상, 중상, 중하, 하로 평가된다. 중간중간 토리찌 캐릭터를 통해 아동의 속마음 영상을 촬영하여 SEvggnet 모델로 부모님에 대한 질문에 답하는 아동의 주된 표정을 평가하고 아동의 감정에 따른 표정 모음집과 영상을 별도로 제공한다.



03 평가 결과 시각화

아동의 언어 평가와 정서 평가 모두 포함된 결과를 보여주는 차트. 아동의 언어와 정서 점수를 비교해能看到. 차트에는 '언어 평가'와 '정서 평가'의 점수와 평가 결과가 표시된다.



04 기관 연결

기관 방문자의 리뷰, 우수 기관 선정 여부, 휴관 여부와 특화된 교육 프로그램 등을 확인하여 직접 선별한 기관들로 아동의 평가 수준에 맞는 기관을 추천받게 된다. 원하는 지역을 클릭하면 해당 지역의 기관을 추천받을 수 있고 편의성을 위하여 비대면 프로그램까지 추천이 가능하다.



개발목표

정보 소외 대상자이자 디지털 소외 계층인 베트남 다문화가정의 아동들이 언어 습득의 불완전성을 보완하고 부모와의 유대관계 형성에 비롯한 정서적인 문제를 해결할 수 있도록 아동 맞춤형 기관을 안내해줌으로써 이들이 제대로 된 아동기와 성년기를 보낼 수 있도록 한다. 아동의 문제를 인지하고 묘사하는 것에 어려움을 겪는 다문화가정 부모님들을 위하여 시가 아동의 문제를 제대로 된 정서 문제로 인지 혹은 두 가지 모두 문제인지를 평가해주어 문제 인식의 어려움을 해소한다. 뿐만 아니라 아동의 평가 결과에 맞는 기관을 추천해주는 방대한 정보 속 정보를 선별해야하는 어려움을 해소한다. 추후, 베트남어를 제외한 다른 언어권을 추가하여 다른 국가로 자원이 되는 언어와 정서 길잡이 서비스가 활성화 가능성을 가지고 있다. 마지막으로 현재는 선별된 기관을 추천하고 있지만 전문기관과 협업하여 플랫폼 내에서 심층 진단과 상담 및 치료가 가능하도록 개발을 이어나간다면 편리성 증대가 기대된다.

추진배경

지속되는 저출산으로 인하여 1인당 부양해야하는 고령 인구수는 매년 늘어나고 있다. 이에 따라 전체 아동 비율 중 다문화가정 아동이 차지하는 비율 역시 매년 늘어나고 있다. 인구수가 줄어들고 있음을 이유로 부모에게 출산을 강요할 수는 없다. 다문화가정 아동들을 제대로 된 환경 속에서 성장시키며 대한민국 경제에 기여할 수 있는 일원으로 성장시키는 것이 저출산의 현실적인 대안이 될 것으로 보인다. 올바른 성인기를 위해서는 아동기의 언어 습득과 정서 형성이 가장 중요然而다문화가정 아동이 한국 사회에 녹아들 수 있는 첫걸음이 되고자 프로젝트를 진행하게 되었다.

프로젝트 추진 내용

프로젝트 진행과정



기대효과

경제적 측면

베트남 다문화가정 아동들이 제대로 된 환경에서 아동기를 보낼 수 있다면 한국 사회에 맞는 사회화가 가능하여 한국 경제의 일원으로 경제 활동에 참여할 수 있게 된다. 이들이 고령 인구를 부양할 수 있는 인구가 되어준다면 청년층 1인당의 부담은 줄어들 것으로 보이며 한국 경제를 성장시키는 것에 도움이 될 수 있다. 사회적 약자를 위해 정보 소외 격차 등의 근본적인 문제를 해결하여 현재 다문화가정을 위한 다양한 지원과 자원의 활용도를 높이고 자원에 대한 접근성을 높여 좋은 자원이 낭비되는 것을 막을 수 있다.

사회적 측면

세계화로 인하여 다문화가정의 가구 비율은 매년 증가하고 있다. 하지만 여전히 다문화가정의 아동들은 한국 사회 속에서 제대로 통화되지 못하고 있다. 여러 언론사의 인터뷰를 참고해보면 언어 습득의 불완전성으로 인하여 한국 아동들과 함께 하는 조별 과제가 어렵고 학교 폭력도 심각하다고 보고 있다. 베트남 다문화가정 아동이 언어 장벽 문제와 부모와의 유대관계 문제만 해결하더라도 훨씬 나은 학교 생활과 일상 생활을 할 수 있을 것으로 기대된다. 사회적 약자가 한국 사회에 제대로 통화될 수 있는 첫 시작점이 되어줄 수 있다.

산학협력

팀원활동



공동: 주제 선정 회의
업체: PM, Whisper 모델 구축, 웹/앱(UI), 기관 데이터 수집, 발표자료 제작 및 발표, 보고서 작성
권영후: 음성 데이터셋 전처리, SEvggnet 모델 구축, 기관 데이터 수집
박현빈: 웹/앱(UI, 프론트), 기관 추천 알고리즘 작성
정성우: Whisper 모델 구축, 이미지 데이터셋 전처리, SEvggnet 모델 구축 및 성능 강화
한건희: 웹/앱(프론트, 백)
홍석준: 웹/앱(UI, 프론트)

멘토활동

한국인의 경우 극우 사용이나 외적인 이유로 인하여 서양인과 달리 표정으로 분류가 되지 않는 감정이 있음을 제시
평가 과정에 있어 불필요한 내용이 들어가지 않고 핵심만 들어갈 수 있도록 서비스 과부하를 중재 전문가의 시각에서 보았을 때 해당 서비스의 명칭 결정에 있어 방향성 제시

PROJECT 05

오조은데

오조은데(5조)



언감생心: 언어 감수성 분석 모델 구축 및 ChatGPT의 언어 감수성 교정

산업체 및 멘토 교수: 트위그팜·고원희 매니저, 두일철·오세종·박민규 교수(한국외국어대학교 AI교육원) / 팀장: 허은(한성대학교 컴퓨터공학)
팀원: 김보경·박지선(한국외국어대학교 언어인지과학), 박찬호(서강대학교 중국문화), 양시현(상명대학교 휴먼지능정보공학), 이예건(한국외국어대학교 컴퓨터공학)

개발동기 및 목적

개발동기 및 목적

- 언어 감수성이란들은 사람의 상황과 감정을 고려하여 언어를 사용하는 능력을 말한다. 우리 사회는 의도치 않게 차별적 의미, 혹은 편견과 고정관념이 내포되어 있는 단어 표현을 종종 사용한다. 코로나19의 첫 등장 후, 감염 경로를 알 수 없는 확진자를 깜깜이 환자로 지칭한 중앙재난안전본부는 시각 장애인을 고려하지 못한 표현이라고 질타를 받은 후 사과문을 발표하고 감염경로 불명으로 수정한 바 있다. 이러한 차별적 표현은 일상 대화에서도 쉽게 찾아볼 수 있다. ‘어미 모’라는 특정 성별을 지칭하는 한자를 사용한 유모차, 팔이 반인 티셔츠를 의미하는 반팔티가 그 예시이다. 언어 감수성을 고려한다면 유모차 대신 유아차, 반팔티 대신 반소매를 사용해야 한다. 실제로 언어 감수성을 고려한 사회적 분위기가 증가하고 있으며 해당 키워드에 대한 언급량 또한 높아지고 있는 추세이다.
- 따라서 기존 언어 감수성 선행 연구에 대해 공학적 접근을 시도해보자라는 의견이 모여 프로젝트를 시작하게 되었다. 여러 논문을 기반으로 하여 언어 감수성 사전을 최초로 구축하고 문장에서 언어 감수성을 고려하지 못한 표현을 검출하고 대안어를 제시하는 모델을 통해 사용자가 언어 감수성을 고려한 표현을 사용하도록 제안하고자 한다.

최종결과물

01 언어 감수성 사전 구축

from	to
가정부	가사관리사
결승가족	가족
결승가정	가정
김 여사	운전 미숙자
김 여사	운전 미숙자
고아원	아동복지시설
경력단절여성	고용중단여성
광부	광산 노동자
과부	故 000씨 배우자
공부	학습
기생오라비	
기생오라비	
개그맨	코미디언
개그우먼	코미디언

02 언어 감수성 분석 모델 구축

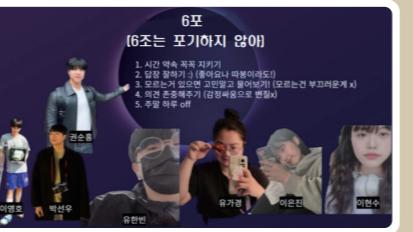
- 입력 문장의 감성 분석을 통해 부정적 여부를 판단하고, 동시에 언어 감수성이 낮을 여부를 판단. 해당 문장의 언어 감수성이 낮다면 언어 감수성 사전의 카테고리를 분류하고, 대안어를 검색하여 출력

언어 감수성 분석 결과
1. 문장은 부정적이지 않은 문장입니다.
언어 감수성 모델 실행 시간: 0.0173초
2. 언어 감수성 분석 결과, 이 문장은 언어 감수성이 낮습니다
언어 감수성 모델 실행 시간: 0.0173초
3. 언어 감수성 카테고리 모델 결과, 이 문장은 “장애 및 장애인”, “사회적 신분” 카테고리에 해당됩니다.
언어 감수성 카테고리 모델 실행 시간: 0.0173초
4. 추천 로직 코드 실행
단어 “장애” “은/는” 언어 감수성이 낮은 단어로 “장애” “은/는”를 표시하는 것이 좋아보입니다.
단어 “장애인” “은/는” 언어 감수성이 낮은 단어로 “장애인” “은/는”를 표시하는 것이 좋아보입니다.
추천 로직 코드 실행 시간: 0.0354초

03 ChatGPT API 적용한 웹 구현



PROJECT 06 6포(6조는 포기하지 않아)



산업체 및 멘토 교수: 글나무·문선희 대표이사, 두일철·오세종·박민규 교수(한국외국어대학교 AI교육원) / 팀장: 유한빈(한국외국어대학교 정보통신공학과) / 팀원: 권순홍·이영호(한국외국어대학교 정보통신공학과), 박선우·이은진(한국외국어대학교 컴퓨터전자시스템공학부), 유가경(한국외국어대학교 국제경영학과 경영정보전공), 이현수(한국외국어대학교 공간환경학부)

개발동기 및 목적

개발동기 및 목적

- 교육과정 및 프로젝트 과정을 진행하던 중 발표자리가 부쩍 늘어난 것을 느꼈다. 하지만, 사회초년생은 우리에게 발표는 익숙하지 않기에 두려워하는 경우가 많다는 것을 알았다.
- 주변 지인뿐 아니라 사회 전체의 발표에 대한 인식이 부정적이라는 것을 깨닫고 이를 해결해보고자 프로젝트를 기획하게 되었다.

개발목표

- 발표에는 많은 요소가 전달력을 좌우한다. 이 때문에, 발표에 대한 평가와 피드백을 언어적, 비언어적 요소로 나누어 제공함으로써 발표자의 연습을 효과적으로 돋구고 하였다. 발표자가 쉽게 연습할 수 있는 환경이 되어주기 위해, 각 요소별 평가기준과 평가내용 및 피드백을 보기 쉽게 가시화하여 제공할 것이다. 우리는 사용자가 자신의 히스토리와 피드백을 통해 보다 발표능력을 발전시킬 수 있도록 하는 길잡이가 되고자 한다.

추진배경

- 현대 사회는 다양성을 존중하고 여러 가지 감수성 문제에 민감하게 반응하고 있다. 이 중에서도 ‘언어 감수성’은 인권, 젠더, 그리고 여러 사회적 이슈를 정교하게 표현할 수 있는 중요한 도구로서의 역할을 하고 있다. 이런 배경 아래, 언어 감수성을 고려하는 사회의 변화와 흐름을 더욱 정확하게 반영하여 현대 언어를 분석하고 검증하는 언어감수성 프로젝트를 진행하게 되었다. 또한, 오늘날 여러 분야에서 중요한 역할을 하는 디지털 인공지능 서비스 ChatGPT의 언어 사용은 그 영향력에 비해 사회적 흐름을 따라가지 못하는 현상을 보였다. 인공지능 서비스는 편향된 언어를 사용할 경우 사용자나 특정 집단에게 불쾌감을 줄 수 있다. 서비스의 답변을 사용자가 무비판적으로 사용할 경우를 대비해, 언어감수성 모델을 통해 ChatGPT의 언어감수성을 보완하고자 하였다. 이를 통해 언어 모델이 사회적 다양성을 존중하고, 긍정적인 사회적 변화를 촉진할 수 있는 도구로 활용될 수 있도록 하고자 한다.

추진배경

- 발표의 중요성은 이전 시대에서부터 계속 증대되고 있는 상황이다. 하지만, 사람들은 발표를 연습할 수 있는 환경이 부재하기에 발표를 쉽게 연습해 볼 수 없다. 이러한 상황은 사람들에게 발표에 대한 부담감과 공포로 이어지고 있다. 따라서, 우리는 이러한 문제의 해결사가 되어줄 플랫폼을 개발하고자 하였다.

최종결과물

01 평가 결과 및 피드백 페이지

- 각 요소(발화 속도, 목소리 크기, 밤음 정확도, 말더듬, 자세) 별로 평가한 점수와 맞춤 피드백 제공



02 히스토리 및 팁

- DB를 통해 저장된 유저의 평가 추이 확인 및 발표관련 팁 제공



프로젝트 추진 내용

프로젝트 진행과정



프로젝트 구현과정

음성처리 구현과정



영상처리 구현과정



기대효과

경제적 측면

- 현재 국내에서 발표를 연습할 수 있는 서비스는 부재하다. 이 때문에, 우리 플랫폼이 배포를 시작한다면 시장 진입 가능하다. 플랫폼 특성상 타겟이 명료하지는 않으나 학생, 사회초년생 등 발표자를 접하는 기회가 많아지는 사람들을和服务을 연습하고 싶어하는 사람 모두가 유저가 될 수 있다는 점에서 범용성이 적용된다.

- 시장 독점성과 타겟의 범용성. 일회성 서비스가 아니라라는 점에서 수익창출이 꾸준히 가능하다.

사회적 측면

- 발표를 두려워하거나 경험에 부족한 모든 사람들에게 도움이 될 수 있다. 향후 스크립트 생성 기능과 디자인 기능, 다양한 발표 환경 적용 기능을 제공함으로써 사용자층을 확대하여 범용성을 극대화 할 예정이다. 국내뿐만 아니라 해외에서도 발표에 대한 인식이 부정적이지만 중요성을 인지하고 있는 점에서 우리 플랫폼이 제공하는 서비스는 사회에 큰 기여를 할 수 있을 것이다.

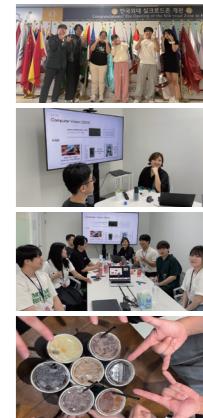
산학협력

팀원활동

- 공동: 발표 자료 제작
- 유한빈: 음성 처리 및 분석 모델 구축, 보고서 작성
- 권순홍: 음성 처리 및 분석 모델 구축
- 박선우: 영상 처리 모델 구축, 발표
- 유가경: UI 제작 및 웹 개발(프론트엔드)
- 이영호: 영상 처리 모델 구축, 웹 개발(프론트엔드)
- 이은진: 웹 개발(백엔드)
- 이현수: 웹 개발(백엔드)

멘토활동

- 데이터 구축 방법 및 음성 분석 모델 방향성 제공
- PT 전체적인 방향성 제시 및 조언
- 가시성 좋은 PT 자료와 효과적인 PT 방법 조언



PROJECT 07

데복이



Alice: 112신고 음성기반 주소완성 및 코드 분류 서비스

산업체 및 멘토 교수: (주)Post AI·차지윤 선임연구원, 두일철·오세종·박민규 교수(한국외국어대학교 AI교육원) / 팀장: 이동현(강남대학교 소프트웨어전공)
팀원: 김이서(한국외국어대학교 언어인지과학과), 박경민·윤대열·홍은진(한국외국어대학교 ELLT 학과), 함상균(한국외국어대학교 전자물리학과)

개발동기 및 목적

개발동기 및 목적

- 프로젝트 주제를 정하면서 112 경찰 시스템에서 존재하고 있는 다수의 문제점에 대해 인식하게 되었다. 현재 경찰 시스템은 신고 전화가 들어오면 해당 지역의 112 종합 상황실로 접수가 되며, 종합 상황실에서 주소를 파악하고 사건의 긴급성에 따라 코드를 0번부터 4번까지 분류해 출동시간이 확정된다.
- 이 과정에서 정확한 주소를 일일이 검색 시스템에 입력을 해야 하고, 사람이 직접 사건의 코드를 판단하여 분류해야 해 경찰관 개인에게 지어지는 책임과 부담이 과도할 것이라는 판단을 하게 되었다. 이러한 문제 인식 과정을 거친 후 조사를 해본 결과 실제 기준 경찰이 사용하는 112 시스템의 폴эм은 업데이트가 늦고 로드맵 기능도 잘 출현되지 않는데다가 멤버십상도 찾아 내야하거나 다음 지역에서 재검색이 이루어지는 경우가 많았으며 검색어가 외래어인 경우에도 절차 하나만 달라도 검색이 되지 않았다. 또한 동일한 사건에 대해 신고 전화를 받는 접수자가 다를 경우 각각 다른 코드가 부여되어 발생하게 된 문제도 많았다.
- 따라서 사람의 직관과 AI의 객관적인 판단과 자동화를 더해줌으로써 좀 더 신속한 접수 및 대응으로 현 경찰 시스템에 존재하고 있는 문제점을 개선하고 경찰관님의 업무를 도와주는 시스템을 만들고자 했다.

개발목표

01 STT 02 주소 자동 특정 및 검색 03 코드 자동 분류

- 신고자의 말이 부정확해 알아듣기 어려운 경우가 많았는데, 인공지능을 활용해 신고자의 말이 다소 부정확하더라도 자동으로 입력만 된다면 접수시간을 상당히 단축시켜 출동을 앞당김으로써 신속 대응에 도움이 되리라고 본다는 경찰관님의 의견을 바탕으로 신고 전화가 접수되었을 때, 신고 내용을 정확하게 텍스트로 전사하는 것을 목표로 한다.**
- TT를 통해 텍스트로 전사된 신고자의 발언은 문제점은 파악하고자 관계자에게 신고 내용에 대한 코드를 자동으로 분류해주고, 사건 카테고리를 특정해준다. 이를 통해 위치의 안심상권수를 보여주는 것을 목표로 한다.**
- 이를 통해 사람의 직관적인 판단에 딸려었던 사건의 긴급성 코드 분류를 객관적으로 향으로써 경찰관님의 좀 더 합리적으로 판단을 내릴 수 있게 될것이라 기대한다.

최종결과물

01 웹 기본 화면



02 음성이 들어왔을 때



03 조사 작성 자동화



추진배경

- 문제 인식 과정을 거친 후 실제 경찰관님 대상 인터뷰를 통해 시스템의 필요성 및 효용성을 파악하기 위해 파출소 및 종합 상황실에서 근무하시는 경찰관님들과의 인터뷰를 진행했다. 인터뷰 결과를 바탕으로 Alice 시스템을 구축되며 현재 경찰 시스템에 존재하고 있는 '주소 특정' 및 '코드 분류'에 대한 어려움을 AI를 통해 완화할 수 있게 되고, 실제 업무에 많은 도움이 될 것이라는 판단을 하게 되어 프로젝트를 추진하게 되었다.

프로젝트 추진 내용

프로젝트 진행과정



PROJECT 09

구강9조



뭐멍냥? : 반려동물 식단 관리 서비스

산업체 및 멘토 교수: 윌로그·공현규 팀장, 두일철·오세종·박민규 교수(한국외국어대학교 AI교육원) / 팀장: 백제민(한국외국어대학교 정보통신공학과)
팀원: 꽈서연(한국외국어대학교 전자물리학과), 조유원(한국외국어대학교 컴퓨터공학과), 서범근(명지대학교 경영정보학과), 김소정(세종대학교 데이터사이언스학과)

개발동기 및 목적

개발동기 및 목적

반려동물 식단관리 앱의 문제점

- 반려동물의 개별적인 특성과 영양학적 필요성을 고려하지 않고, 일반적이고 표준화된 식단을 제공하는 경우가 많습니다. 예를 들어, 반려동물의 나이, 성별, 품종, 체중, 활동량, 건강상태 등에 따라 적절한 칼로리와 영양소의 비율이 달라지는데, 이를 반영하지 않고 같은 식단을 권장하는 앱이 있을 때 문제입니다.
- 반려동물의 사료 성분을 분석하거나 비교하는 기능이 부족하거나 없는 경우가 많습니다. 사료는 반려동물의 건강과 직결되는 요소인데, 대부분의 앱은 사료의 브랜드나 가격만을 기준으로 추천하고, 사료의 성분이나 원산지, 안전성 등에 대한 정보를 제공하지 않음
- 반려동물의 식단을 관리하기 위한 다양한 기능이 부재하거나 미흡한 경우가 많습니다. 예를 들어, 반려동물의 식사량이나 시간을 기록하고 알림을 주는 기능, 반려동물의 체중이나 체지방률을 측정하고 관리하는 기능, 반려동물에게 적합한 간식이나 수제사료 레시피를 제공하는 기능 등이 있음
- 국내 반려동물 관련 시장이 3조원에 근접하고 반려동물 인구가 1000만에 이를 정도로 급격하게 성장하면서 펫푸드 시장에 대한 관심도 증가하고 있다. 그럼에도 불구하고 자작하고 싶은 것은 반려동물 관련 산업체뿐만 아니라 반려동물 가족들이 펫푸드를 잘못 이해하고 있다는 점이다. 기죽은은 사료이지만 반려동물용은 음식 또는 푸드라고 부르며 조금은 인간과 더 친밀한 명칭을 부여했다.
- 이는 같은 동물임에도 영성이 다르고 우리 곁에 두는 목적도 다르다는 것을 명확하게 하고 있다고 생각한다. 이러한 이유로 최근 사람 음식처럼 질 높은 펫푸드가 시장을 넓혀가고 있다. 다만 그 누구도 펫푸드는 논란에 있어 동물과 사람과의 차이에 대한 고민은 부족하다는 생각이 만연함

개발목표

- 반려동물을 키우는 세대수가 지속적으로 증가하고, 펫 사료나 반려동물 서비스가 코로나 19 이후 모바일 분야에서 수요가 급증하고 있다. 반려동물을 가족처럼 가족 같은 반려동물에게 좋은 음식을 제공해 주고 싶은 것은 너무나 당연하지만, 펫푸드는 논란에 있어 동물과 사람과의 차이에 대한 고민이 다소 부족하다. 기존과는 다른 관점으로 접근해서 반려동물이 필요한 영양소를 분석해서 해당 반려동물이 먹으면 되는것과 그렇지 못한 것을 구분해줘서, 반려동물의 수명증기의 효과를 기대할 수 있다.

추진배경

- 반려동물을 키우는 인구가 점점 늘면서 펫푸드 산업과 펫먹방 수요 증가하고 있다. 반려동물에게 더 좋은 음식을 먹이고 싶어하는 사람들이 증가하면서, 펫푸드 시장은 지속적으로 규모가 커지는 중이다. 기존에 제공되고 있는 서비스에서 부족하고 아쉬웠던 부분을 채우고, 다른 방향성도 고려해보려고 다음과 같은 프로젝트를 진행하게 되었다.

최종결과물

01 YOLOv8를 활용한 데이터셋 구축

```
model.train(data='content/yolov3/data.yaml', epochs=100, batch=32, imgsz=640)

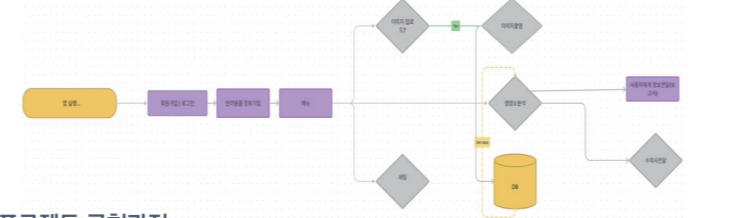
Epoch GPU_mem box_loss cls_loss df_l_loss Instances Size
100/100 4.65G 0.3678 0.2677 1.043 30 640:
Class Images Instances Box(P R mAP50
all 1315 1319 0.982 0.879 0.936
```

02 GPT API를 활용한 최종 영양성분 보고서 결과

사료 (100g)	
칼로리: 550kcal	
조단백: 31g (31%)	
조지아: 27g (27%)	
칼슘: 0.96g (0.96%)	
망: 0.04g (0.04%)	
영양 성분 분석	
트로피의 영양 지침과 탄수화물 함량이 높아 영양 불균형을 막자는 목표로 작용할 수 있습니다.	
조류들은 가나으로서 적극적인 영양 분석을 할 수 있습니다.	
사료의 조단백과 조지아에게 영양 분석을 진행해 보겠습니다. 하지만 칼슘과 인의 함량이 기준치를 넘어서 있으므로 주의가 필요합니다.	
영양 분석과 관련하여 수제사료의 상당부분이 칼슘과 영양 균형을 적절하게 관리하시길 바랍니다.	

프로젝트 추진 내용

프로젝트 진행과정



프로젝트 구현과정

- AI Hub에서 자연 음식 데이터셋 수집 후 라벨링 작업
- 데이터 전처리 라벨링 과정에서 선명도가 떨어지는 데이터는 모두 제거 후 모델링 성능 비교
- YOLOv8을 최종적인 모델로 선택하고, 이미지 데이터셋을 다시 학습 시키기
- GPT API를 사용해서 주간 및 월간 단위로 먹었던 영양소 성분 및 정보를 바탕으로 기록된 보고서를 최종적으로 도출

기대효과

- 개인화된 관리
 - 시간 절약
 - 건강한 반려동물
- 서비스를 통해 각각의 반려동물을 즉시적인 식단 정보 제공으로 정확한 영양소와 칼로리를 반려에 대한 맞춤형 식단 정보를 제시해 반려인은 원하는 정보를 제공함으로써, 그들의 공함으로써, 반려동물의 건강 찾기 위해 시간을 낭비할 필요가 없습니다.

- 신뢰성
 - 편리성
- AI 기반의 분석을 통해, 사용자는 반려동물의 식단에 대한 신뢰할 수 있는 정보를 얻을 수 있어, 반려인의 일상 속에서 편리하게 사용할 수 있습니다.

산학협력

팀원활동

김소정: 데이터 크롤링 및 데이터셋 구축
곽소연: UI, 발표 자료 제작
서범근: YOLOv8 성능 비교 및 모델 구축
백제민: 발표 자료 제작 및 발표
조유원: 데이터베이스 구축, 플러터 연결

멘토활동

발표 자료 분석
반려동물 동태 및 현황 분석 자료 제공
식단관리 서비스가 비즈니스화 될 수 있는 다른 방향 제시
기준에 반려동물 서비스의 문제점 분석자료 제공



PROJECT 10

NLPicasso

NLPicasso.

피카피카-

필한: 나현석

팀원: 김다빈

김미영

오세종

이승훈

한승윤

김현우

양현우

이현우

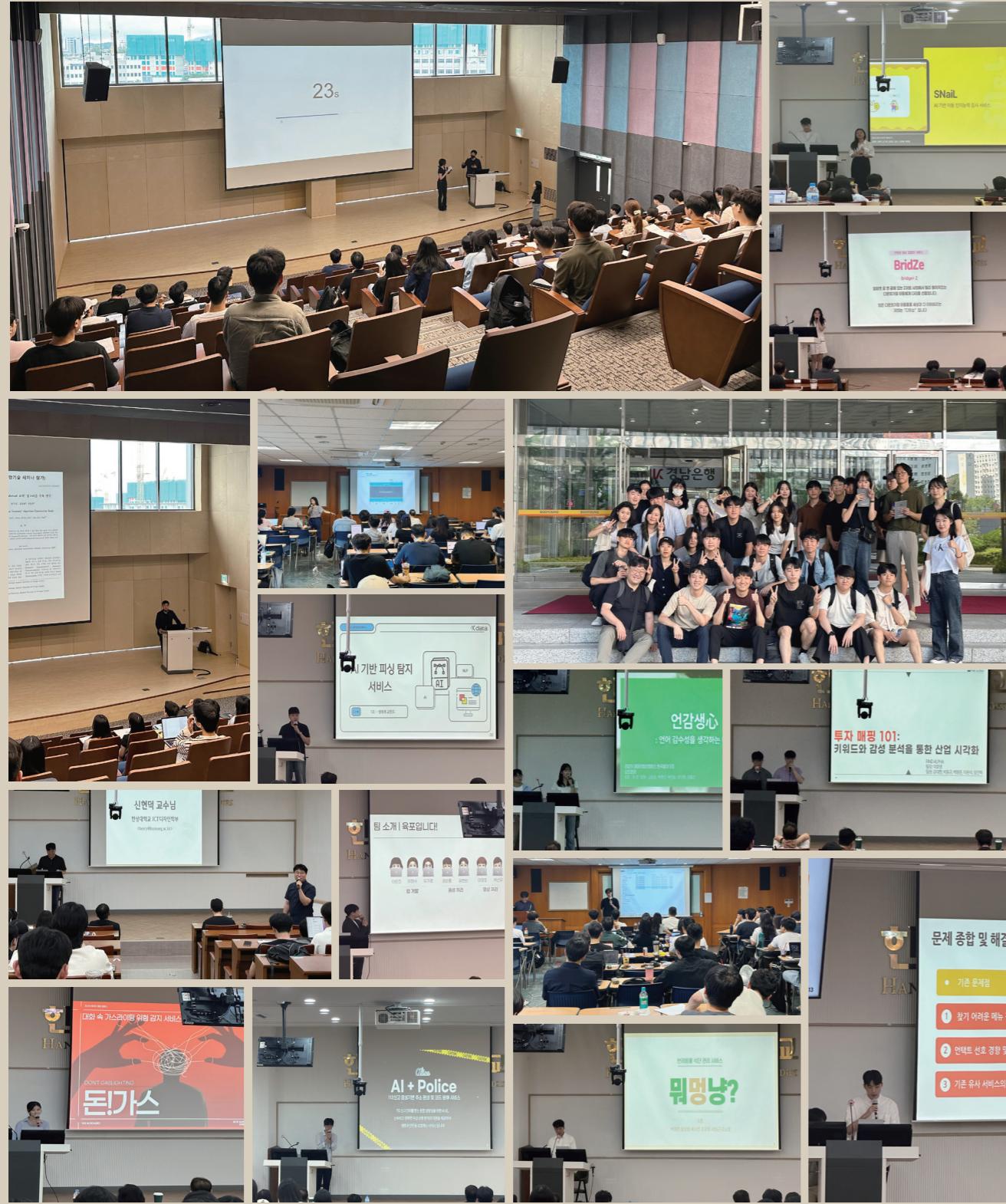
한승우

김현우

이현우

한승우

수업 및 발표 사진



팀별 회의 및 기타 활동 사진

