

SW프로젝트 응용 계획서 (팀 프로젝트)						
선택주제	AI 사투리 번역 시스템			총 투입인원	5 명	
실제 프로젝트 주제 명	국문	대한민국 사투리 번역 AI: 말모이				
	영문	MALMOI: A Korean Dialect Translation AI				
수행기간	2025. 9. 5 ~ 2025. 9. 19. (2주간 진행)					
참여자	소속	참여자성명	학년	학번	휴대전화	이메일
	AI소프트웨어학과	박민준(팀장)	3	2021320028	010-5574-0027	4comma3@naver.com
	AI소프트웨어학과	이종구	3	2020315018	010-8802-1408	jonggu2020@naver.com
	AI소프트웨어학과	이준규	3	2021320024	010-9949-4634	leejungkyn2865@naver.com
	AI소프트웨어학과	김민재	3	2021320004	010-3512-3824	vinus12018@gmail.com
	AI소프트웨어학과	성찬영	3	2021320005	010-7192-2270	chanyounng6@gmail.com
프로젝트의 목표						
<p><b>최종 목표:</b> 인공지능(AI) 기술을 통해 대한민국 팔도의 다양한 사투리를 표준 서울말로 변환하여 지역 간 언어 장벽을 허물고, 모든 사람이 언어의 차이 없이 원활하게 소통하는 사회적 기반을 마련합니다.</p> <p><b>1. Why: 개발 동기 및 목적</b></p> <p>저희는 다음과 같은 문제의식을 바탕으로 본 프로젝트를 추진합니다.</p> <ul style="list-style-type: none"><li><b>소통의 장벽 해소</b>지역적 특색이 강한 사투리는 타 지역 사람들과의 의사소통에 어려움을 유발하며, 특히 공적인 자리나 비즈니스 환경에서 오해의 소지가 되거나 정보 전달의 효율성을 저해할 수 있습니다. 저희는 이러한 언어적 장벽을 기술로 극복하고자 합니다.</li><li><b>미디어 및 콘텐츠 접근성 향상</b>영화, 드라마, 유튜브 등 다양한 미디어 속 사투리는 비사투리 사용자의 콘텐츠 몰입과 이해를 방해할 수 있습니다. 자동 변환 기술을 통해 누구나 자막 없이 콘텐츠를 온전히 즐길 수 있도록 접근성을 획기적으로 개선합니다.</li><li><b>디지털 정보 격차 해소</b>사투리를 주로 사용하는 고령층이나 특정 지역민이 표준어 중심의 음성 인식 비서, AI 상담원 등 디지털 서비스를 사용하는 데 겪는 어려움을 해소하여 디지털 포용성을 높입니다.</li></ul> <p><b>2. What: 단계별 개발 결과물</b></p> <p>프로젝트 목표 달성을 위해 다음과 같은 단계별 결과물을 도출합니다.</p> <p><b>1차 목표: 사투리 ↔ 표준어 텍스트 변환 모델</b></p> <ul style="list-style-type: none"><li><b>핵심 결과물:</b> 다양한 지역의 사투리 텍스트를 입력하면, 문맥을 정확히 이해하여 자연스러운 표준어 텍스트로 변환하는 AI 번역 엔진을 개발합니다.</li><li><b>세부 결과물:</b><ul style="list-style-type: none"><li>고품질의 사투리-표준어 병렬 말뭉치(Corpus) 데이터셋 구축</li><li>웹 기반 텍스트 변환 데모 서비스 개발</li></ul></li></ul> <p><b>2차 목표: 음성인식(STT) 기반 실시간 변환 시스템</b></p> <ul style="list-style-type: none"><li><b>핵심 결과물:</b> 사용자가 사투리로 말하면 이를 실시간으로 인식하여 표준어 텍스트로 변환해주는 시스템을 구축합니다.</li><li><b>세부 결과물:</b><ul style="list-style-type: none"><li>사투리 음성 데이터에 특화된 음성 인식(Speech-to-Text) 모델 개발</li><li>음성 인식 모델과 1차 목표의 텍스트 변환 모델을 통합한 파이프라인 구축</li></ul></li></ul>						

### 3차 목표: 실시간 음성 통역 솔루션

- **핵심 결과물:** 사투리 음성 입력을 받아 표준어 음성(TTS)으로 실시간 출력하는 완전한 형태의 통역 솔루션을 완성합니다.
- **세부 결과물:**
  - 2차 목표 시스템에 텍스트-음성 변환(Text-to-Speech) 기술 결합
  - 모바일 앱 또는 웨어러블 디바이스용 실시간 통역 프로토타입 개발

### 3. Impact: 기대 효과 및 창출 가치

본 프로젝트는 다음과 같은 사회적, 경제적, 기술적 가치를 창출할 것입니다.

- **사회적 가치**
  - **지역 간 소통 증진:** 출신 지역에 관계없이 명확하고 원활한 의사소통을 가능하게 하여 사회 통합에 기여합니다.
  - **교육 자료로의 활용:** 학생들이 각 지역 사투리와 표준어를 비교 학습하는 교육용 도구로 활용될 수 있습니다.

#### 문화적/경제적 가치

- **K-콘텐츠 글로벌 경쟁력 강화:** 사투리가 사용된 한국 콘텐츠를 표준어로 자동 변환하여 해외 팬들의 이해를 돕고, 글로벌 확산을 가속화합니다.
- **새로운 서비스 시장 창출:** AI 컨택센터(AICC), 비대면 회의, 공공 민원 서비스, 관광 앱 등 다양한 산업에 기술을 접목하여 새로운 부가가치를 창출합니다.

#### 기술적 가치

- **한국어 자연어 처리(NLP) 기술 선도:** 국내 방언 AI 연구를 활성화하고, 고난도 언어 처리 기술력을 확보하여 국내 AI 기술 생태계에 기여합니다.
- **독자적인 데이터 자산 확보:** 프로젝트를 통해 구축된 대규모 사투리 데이터셋은 그 자체로 귀중한 기술적 자산이 됩니다.

#### 개발내용

IDE: VScode

언어: Python, JavaScript

프레임워크: PyTorch, Hugging Face Transformers, React

API서버구축: FastAPI, Node js

DB: Firebase, Mysql

사용할 외부 서비스: GIT, Wandb.ai, 외부 api

모델 선정 : DNN, 랜덤 포레스트(Xgboost 등)

사용할 데이터 :

경상도 지역 사투리 모음 (<https://cronquist0.tistory.com/8751531>)

제주어 사전 (<https://www.jeu.go.kr/culture/dialect/dictionary.htm>)

전국 사투리 모음 사이트,

전라도 사투리 모음 사이트

(<https://m.cafe.daum.net/obh2004/2le1/106?listURI=%2Fobh2004%2F2le1%3FboardType%3D>)

#### 추진 전략

- \* 개발에 필요한 방법 및 추진전략을 요약
- \* 참여개발자간의 역할분담 및 추진 일정

주요내용	추진 일정 (주)			
	1주차	2주차	3주차	4주차
데이터 수집 및 전처리	o			
요구사항 수집 및 정의서 작성	o			
핵심 필수 기능 개발	o	o		
인공지능 모델 학습 및 성능 검증		o	o	o
프로토타입 테스트 및 수정			o	o
결과 보고서 작성 및 발표 문서 작업				o

#### 팀만의 전략

\* 어떤 방식으로 협업을 진행할 것인지에 대해서 기술하시오.

GIT을 이용하여 버전관리를 한다.

Wandb.ai를 이용하여 모델버전 관리 및 학습을 한다.

notion을 이용하여 스케줄 관리를 한다.

google doc를 이용하여 다같이 문서 작성 및 관리를 한다.

카카오톡, discord를 이용하여 회의를 진행한다.

#### 애로사항

데이터셋의 형태가 제각각이어서 데이터 전처리 진행할 때 오류가 발생할 것 같음.