

프로젝트 기획서

2기 3조 추운자리
구선아 김진유 이재원 전상욱

주제	문제점	수작업으로 작성되는 회의록은 시간이 많이 소요되며 인적 오류가 발생할 수 있다.	
	문제점 해결 목표	1. STT 기술을 사용해 회의록을 생성하여 시간을 절약하고 인력 소모를 줄인다. 2. sLLM을 활용한 자동 요약 기능을 통해 중요한 내용을 빠르게 파악할 수 있도록 한다.	
시스템 구성	Database	Amazon RDS for MySQL	Tables -- User, Meeting, Participant, Minutes, Notification ...
	Model	AWS EC2 Instance	
	Web	Django + AWS EC2 Instance	
	비고	Model에 Spot Instance를 사용하는 것을 고려	
모델링 방안	방안	Huggingface의 transformers 패키지를 활용하여 기존 모델을 로드한 후, 모델을 회의록을 요약하는데 최적화하기 위해 요약 데이터를 fine-tuning한다. 제한된 시간, 제한된 리소스 내에서 fine-tuning을 하기 위해 LoRA 혹은 QLoRA 기법을 활용한다. 적절한 양식의 회의 요약문을 얻기 위해 prompt를 적용한다.	
	후보 모델	현재 ko-LLM 리더보드 기준 ai-human-lab/EEVE-Korean_Instruct-10.8B-expo (Llama base) 검증된 모델 nlpai-lab/KULLM3 (Llama base)	
데이터		음성인식 모델	회의록 요약 모델(sLLM)
		회의 음성 (248.33 GB) 주요 영역별 회의 음성인식 데이터 (563.73 GB) 출처 - AIHub * 데이터 크기가 매우 크기 때문에, 데이터의 일부만 랜덤추출하여 학습시킬 계획	요약문 및 레포트 생성 데이터 (472.59 MB) 출처 - AIHub
			국회 회의록 말뭉치 2021 (307MB) 국회 회의록 요약 말뭉치 2022 (71MB) 국회 회의록 요약 말뭉치 2023 (58MB) 출처 - 국립국어원 모두의 말뭉치
R&R	구선아	AI 모델링(sLLM) / AWS / 문서 및 일정 관리	
	김진유	Web(django) / 발표	
	이재원	Web(django) / AWS	
	전상욱	AI 모델링(STT) / 문서 및 일정 관리	