|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **과제 보고서**  **(2024.07.26)** | | 결  재 | 작 성 자 | 선 임 | | 부 장 | | 위 원 |
|  |  | |  | |  |
| 작 성 자 | 서원형 | | | | 작 성 일 | | 24.07.26 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 과제명 | 전자약 LLM 모델 구현 | 진행도(%) | 100 |
| 결론 | - 과제로 주어진 논문의 LLM 모델(3종)을 구현함  [남은 문제점]  모델을 고도화하기 위해서는 데이터가 추가적으로 필요함. | | |
| 진행 과정 | **[문제점]**    <구현해야 하는 논문 분석 모델>  논문에 작성된 모델을 구현하기 위해서 데이터 선정, 개발환경 구축, 데이터 전처리, 모델 학습, 모델 평가, 모델 최적화 순서로 계획을 잡고 LLM 모델 구현을 실행하였다.  **[해결 방법]**  논문에서 사용된 데이터는 4년전 특정 대회에서 쓰여진 데이터로 쉽게 구할 수 없으며, 치매 포럼의 회원만이 다운로드가 가능했다. 따라서, Audio를 생성형 모델에게 치매환자처럼 만들어 달라고 할 수 없었으므로 가능한 text를 사용하여 아래 그림의 모델을 구현하였다.      <text로 구현 가능한 부분은 구현을 완료하였다.>  개발환경은 google colab을 사용하였고, 유료 GPU를 사용할 수 없었기 때문에, 세션이 비정상적으로 종료되어 학습이 정상적으로 불가능한 경우도 있었지만, 모델의 성능보다는 모델 구현에 중심을 맞춰 파라미터를 조절하여(에폭시를 3으로 두는 등) 모델을 생성하였다.  **[결론]**   |  |  | | --- | --- | | **모델** | **결과** | | **텍스트 데이터 활용 BERT 모델** | Separate Model - Precision: 100.0000%, Recall: 57.1429%, F1-Score: 72.7273%, Specificity: 100.0000%, Accuracy: 78.8732% | | **GPT 의견 활용 BERT 모델** | Separate Model - Precision: 100.0000%, Recall: 57.1429%, F1-Score: 72.7273%, Specificity: 100.0000%, Accuracy: 78.8732%. | | **텍스트 + GPT 의견 활용 트랜스포머 인스턴스 공유 BERT 모델** | Shared Model - Precision: 100.0000%, Recall: 97.1429%, F1-Score: 98.5507%, Specificity: 100.0000%, Accuracy: 98.5915% |     <공유 모델> <분리 모델 가중치 0.5>  기존 논문의 정확도와 혼동행렬 지표를 뛰어넘은 것도 존재할만큼 유의미한 모델이 추출되었다. 또한, 데이터를 바꾸면 새로운 모델이 나올 수 있게 템플릿의 형태로 활용이 가능하다.  **[남은 문제점]**  데이터가 정상적으로 모두 제공되지 않은점, GPT가 생성한 데이터이기 때문에 실제 알츠하이머 환자의 데이터와는 괴리감이 있을 수 있다는 점을 고려하여, 데이터를 확보한 후에 모델의 고도화가 가능하다고 생각한다. 해당 논문 데이터는 지도교수의 소속, 신분등을 치매 포럼에 제출 후 이용목적을 검증받는다. 이후 회원으로 인정이 되면 이용이 가능하다. | | |