

DU-도전학기 결과보고서

과제명	Ai 마케팅 프로젝트 : 광고카피 자동 생성하기		
참여자	성명	소속	학번
	김 [REDACTED]	경영학과(Ai소프트웨어)	[REDACTED]
	변 [REDACTED]	무역학과(Ai소프트웨어)	[REDACTED]
		이하 여백	
지도교수 의견	<p>본 팀은 자연어처리 응용 기술을 토대로, 블로그의 다양한 주제의 텍스트 콘텐츠를 자동 생성하여 블로그 개발과 운영에 효율적으로 적용 가능한 기술을 연구하였다. 학습한 AI모델 (알파카로라의 경량 버전, 페이스북 openSW)을 로컬에서 구동가능하여 블로그 사용자의 접근성이 용이하도록 하였다. 학습데이터는 2만5천 라인의 텍스트를 활용하였다. 향후 성능 향상을 위해서 일반 사용자가 키워드에 어느정도 부합되는 주제 글이 생성된다는 피드백을 적용하여 보다 활용성을 높일 수 있다.</p> <p>연구와 구현 결과는 자연어처리 및 웹 기술의 학습 성과를 충분히 달성하였다고 사료된다. 또한 도전학기 교육목표에 부합되는 활동으로 학점 부여 및 경쟁력 확보에 기여하였음을 확인하였다.</p> <p>앞으로 이 학생들의 훌륭한 성장이 기대된다.</p> <p style="text-align: right;">(소속) AI학부 (성명) [REDACTED] 인)</p>		

1. 도전 과제 내용

프로젝트: 블로그의 다양한 주제의 텍스트 콘텐츠 자동생성
(변경사항 내용은 중간보고서에 첨부했음을 알림)

목표: 블로그 사용자들이 다양한 주제의 텍스트 콘텐츠를 손쉽게 생성할 수 있도록 지원하고, 사용자들의 지식과 관심사의 폭을 넓히는 것이 목표입니다.

(1) 데이터 수집, 가공

1. AI 모델에 적용할 블로그 내용 데이터를 정의한다.
2. 인터넷 상에서 다양한 주제의 블로그 포스트를 수집한다 (예: 여행, 요리, 영화 리뷰 등).
3. 주제별로 데이터를 분류하여 정리하고, 각 주제에 대한 키워드 및 주요 패턴(표현, 어투 등)을 추출한다.
4. 학습 데이터와 라벨링 작업을 수행한다.

- 수집된 블로그 포스트를 군집화(Clustering)하여, 유사한 주제와 스타일의 글들끼리 묶는다.
- 각 군집의 대표적인 예제들을 선별하여 라벨링 작업을 수행한다. (예: 문장 구조, 주제 태그, 스타일 등을 라벨로 정의)
- 라벨링된 데이터를 기반으로 AI 모델을 학습시킨다.

(2)모델 개발 및 테스트

- alpaca-LoRa 를 기반으로한 자연어 처리 모델을 개발하며, 학습된 데이터를 적용한다.
- 모델이 생성한 블로그 포스트의 주제, 문장 구조, 스타일 등이 사용자의 요구와 일치하는지 테스트한다.

2. 도전 과제 수행 결과 및 성과

- 자연어 처리 기술 습득: 텍스트 데이터 전처리, 토큰화, 임베딩 등의 기술을 이해하고 적용하였습니다. 이를 통해 AI 모델에서 필요한 입력 데이터를 처리할 수 있는 역량을 키웠습니다.
- 웹 크롤링 및 데이터셋 전처리: 웹 크롤링을 통해 필요한 데이터를 수집하고, 데이터 전처리 과정을 거쳐 효율적인 모델 학습을 위한 데이터셋을 구축하였습니다.
- 웹 개발 역량 강화: HTML, CSS, JavaScript 등의 프론트엔드 기술을 익혀 사용자 친화적인 인터페이스를 구현하였으며, 이를 통해 AI 모델과 연결되는 웹 플랫폼을 개발하였습니다.
- AI 모델 학습: 주어진 데이터셋을 활용해 모델을 학습시키고, 끈질기게 시도한 결과로, 답변이 과거보다 개선되었으며, 모델 수정 및 학습에 총 12시간이라는 막대한 시간이 걸렸습니다. 하지만 결과적으로 생성 AI 모델 제작에 실패했지만, 그 과정 속에서 많은 것을 배울 수 있었습니다.각 분야에서의 경험과 지식 쌓기를 통해 과제 수행이 가능했으며, 앞으로의 대학생활에도 큰 도움이 될 것입니다.

3. 자기 평가

김 ■■■

기존 목표를 완전히 달성하지 못했음에도 불구하고, 이번 1학기 동안 자연어 처리, 웹 제작, 모델 학습과 같은 다양한 분야에서 소중한 경험과 지식을 쌓을 수 있었습니다. 이러한 경험은 개인적인 성장과 전문성 증진에 큰 도움이 되었습니다. 특히, 자연어 처리와 모델 학습 분야에서 얻은 지식을 토대로 일반적인 글 생성까지는 성공했다는 점에서 성취감을 느낍니다. 그러나 아쉬움도 남습니다. 목표로 했던 블로그 글 생성에 이르지 못한 점은 개선해야 할 부분이라 생각합니다. 끊임없이 노력하고 지속적으로 발전해 나가면서 더 나은 결과물을 만들어내려는 의지가 필요합니다. 종합적으로, 이러한 경험을 바탕으로 앞으로의 과제와 도전에서 더욱 발전하여 목표한 블로그 글 생성 AI를 완성할 수 있게끔 준비하겠습니다. 차츰 개선과 도전을 통해 더 나은 성과를 거두며 자신감을 기르고 역량을 강화할 것입니다.

변 ■■■

AI 모델 개발을 위한 1학기 동안의 활동 수기는 다음과 같습니다. 1학기 초에 프로젝트를 시작하면서, 자동으로 글을 생성해주는 AI를 개발하는 목표가 세워졌습니다. 우리 팀은 열정적으로 자연어 처리, 웹 제작, 모델 학습 등의 다양한 분야에 대한 지식을 쌓았으며 그 과정에서 원활한 협업과 함께 서로의 전문성을 배워 나갔습니다. 자연어 처리에서는 텍스트 데이터 전처리와 토큰화, 임베딩 기술을 익히며 문맥과 상황에 알맞은 문장을 생성하는 모델 개발에 한걸음씩 다가갔습니다. 웹 제작 과정에서는 사용자 친화적인 인터페이스를 구현하기 위해 HTML, CSS, JavaScript와 같은 프론트엔드 기술을 익혔습니다. 이를 통해 사용자들이 AI를 손쉽게 사용할 수 있는 웹 플랫폼을 만들어내었습니다. 모델 학습에서는 많은 시행착오에도 불구하고, 일반적인 글을 출력하는데까지 성공하였습니다. 끈기와 노력으로 목표한 블로그 글 생성까지는 어렵게도 이르지 못했지만, 얻은 경험과 지식은 앞으로 발전해서 더 나은 결과물을 만들어낼 수 있다는 자신감을 심어 주었습니다. 이러한 과정을 통해 학습한 기술과 노하우를 앞으로의 프로젝트에도 활용하여, 언젠가는 목표로 했던 블로그 글 생성 AI를 완성하는 날이 오기를 기대합니다. 이러한 경험을 계기로 팀원 모두가 개인적으로 성장할 수 있었으며, 앞으로도 지속적인 발전을 이루어 나가기 바랍니다.

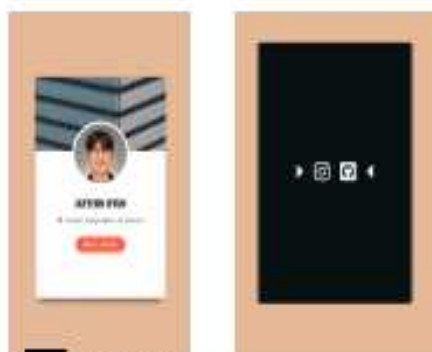
4. 최종 결과물

- 개인 결과물에 대한 상세한 내용은 웹보고서에 모두 적어놓았습니다.



■ 개인 결과물

<https://hibboprofile.netlify.app/>



■ 개인 결과물

<https://byun-profile.netlify.app/>

팀 결과물

- 팀 결과물에 대한 상세한 내용은 웹보고서에 모두 적어놓았습니다.

해당 프로그램은 사용자가 입력한 내용을 주제로 하여 관련된 글을 자동으로 생성해주는 AI 기반 텍스트 생성 서비스입니다. 사용자는 주제나 키워드를 입력하면, 프로그램은 이를 바탕으로 연관된 내용을 포함한 글을 작성하여 제공합니다. 하지만 아직 성능이 매우 부족한 편이나 이러한 기능의 성능이 올라간다면 효율적인 콘텐츠 생산과 크리에이티브한 아이디어 발굴에 도움을 줄 수 있습니다.

프로그램은 자연어 처리 기술, 웹 개발 역량, 데이터셋 구축 및 전처리, AI 모델 학습 등 여러 과정을 통해 구현되었습니다. 이를 통해 사용자 친화적인 인터페이스와 효율적인 모델 학습의 기반을 제공할 수 있게 되었습니다. 앞으로 이 프로그램을 발전시켜 더 다양한 주제와 분야에서 글의 자동 생성을 지원할 수 있도록 노력하겠습니다.

참고자료 : <https://github.com/tloen/alpaca-lora>

