

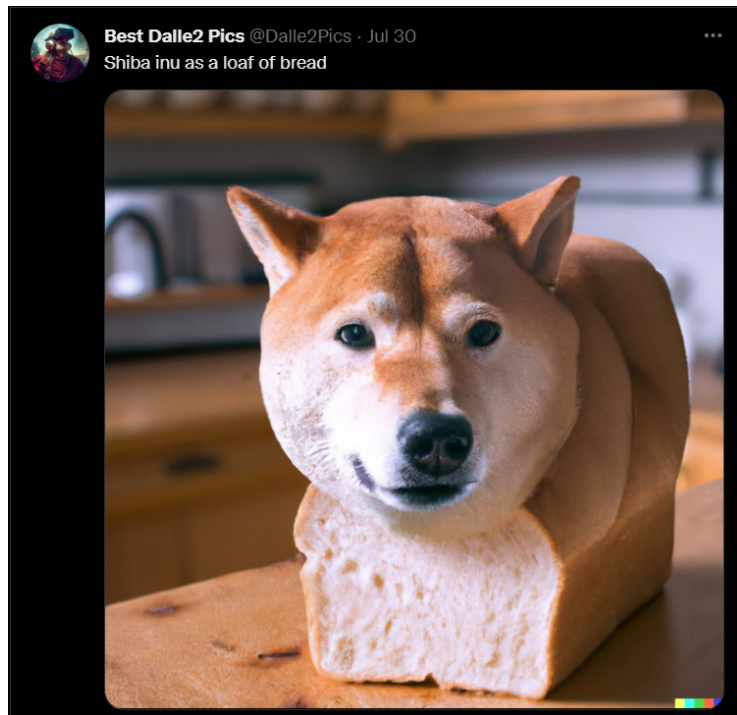
KoGPT2를 사용한 웹툰 줄거리 생성

트렌드리서치 팀

복숭아눅차 김세영 김지후

AI는 예술가가 될 수 있을까?

예술 창작 분야의 AI



■ AI는 예술가가 될 수 있을까?

예술 창작 분야의 AI

단어만 입력하면 AI가 대신 줄거리를 만들어 줄 수 없을까?

■ AI는 예술가가 될 수 있을까?

예술 창작 분야의 AI

웹툰 줄거리 데이터를 기반으로
단어만 입력하면 줄거리나 대사를 생성할 수 있을까?
시작 단어 또는 문장 입력 시
웹툰 줄거리를 자동으로 생성하는 모델

#1 데이터

- 크롤링

	id	title	author	day	genre	story
0	0	참교육	채용택 / 한가람	월	스토리, 액션	무너진 교권을 지키기 위해 교권보호국 소속 나화진의 참교육이 시작된다!<부...
1	1	신의 탑	SIU	월	스토리, 판타지	자신의 모든 것이었던 소녀를 쫓아 탑에 들어온 소년<그리고 그런 소년을 시험하는 탑
2	2	뷰티풀 군바리	설이 / 윤성원	월	스토리, 드라마	'여자도 군대에 간다면?' 본격 여자도 군대 가는 만화!
3	3	윈드브레이커	조용석	월	스토리, 스포츠	혼자서 자전거를 즐겨타던 모범생 조자현.<원치 않게 자전거 크루의 일에 자꾸 휘말...
4	4	퀘스트지상주의	박태준 만화회사	월	스토리, 드라마	[외모지상주의], [싸움독학], [인생존망]과 세계관을 공유하는 작품!<공부, 싸...

네이버의 연재/완결/도전 웹툰 줄거리 크롤링 (총 4,567개의 데이터)

#1 데이터

- 전처리

```
def alpha_num(text):
    return re.sub(r'^ㄱ-ㅎㅏ-ㅣ가-힣 A-Z a-z 0-9 ]', ' ', text)
```

story	cleanstory
무너진 교권을 지키기 위해 교권보호국 소속 나화진의 참교육이 시작된다!&부...	무너진 교권을 지키기 위해 교권보호국 소속 나화진의 참교육이 시작된다 부활...
자신의 모든 것이었던 소녀를 쫓아 탑에 들어온 소년&그리고 그런 소년을 시험하는 탑	자신의 모든 것이었던 소녀를 쫓아 탑에 들어온 소년 그리고 그런 소년을 시험하는 탑
'여자도 군대에 간다면?'본격 여자도 군대 가는 만화!	여자도 군대에 간다면 본격 여자도 군대 가는 만화
혼자서 자전거를 즐겨타던 모범생 조자현.&원치 않게 자전거 크루의 일에 자꾸 휘말...	혼자서 자전거를 즐겨타던 모범생 조자현 원치 않게 자전거 크루의 일에 자꾸 휘말리...
[외모지상주의], [싸움독학], [인생존망]과 세계관을 공유하는 작품!&공부, 싸...	외모지상주의 싸움독학 인생존망 과 세계관을 공유하는 작품 공부 싸움...

- 1) NULL 값과 중복값 삭제
- 2) 정규표현식을 사용해 문장부호 제거
- 3) 1개 이상의 공백을 하나의 공백으로 치환

#2 모델

- 선정 이유

GPT

(Generative Pre-trained Transformer)

Transformer 구조를 바탕으로 넓은 범위에 걸친 학습 가능
사전학습 모델의 구조 변화를 최소화하여 효과적으로 파인튜닝

#2 모델

- GPT-2

GPT-2는 논문 Attention Is All You Need에서 제시한
트랜스포머 구조에서 인코더 블록을 제거하고 디코더 블록만 사용한 모델

Masked Self-Attention을 사용하여 문장 생성에 최적화

#2 모델

- GPT-2



- 1) 각 단어의 순서들을 고려한 임베딩 행렬 입력
- 2) 각 디코더의 Self Attention 과정을 거친 후 신경망 레이어 통과
- 3) 출력값을 임베딩 벡터와 곱함. 해당 결과값은 각 단어가 다음 단어로 등장할 확률값으로 출력됨
- 4) 이 중 가장 높은 확률값을 가지는 단어를 출력하며, 이 값은 다음 입력값이 됨

#2 모델

- Pre-Training

KoGPT2

GPT-2의 부족한 한국어 성능을 극복하기 위해
40GB 이상의 텍스트로 학습된 한국어 디코더 언어모델

#2 모델

- Fine-Tuning

Fine-Tuning

: 모델의 파라미터를 미세하게 조정하는 과정

사전 학습된 KoGPT2의 Task에 맞는 웹툰 줄거리 데이터 추가 학습

#2 모델

- Fine-Tuning 하기 전 결과

[입력 키워드] 첫사랑

첫사랑을 위해 헌신하는 모든 분들에게 감사드린다"고 말했다. 한편 이날 행사에는 김희정 여성가족부 장관, 박원순 서울시장, 이낙연 국무총리, 홍준표 경남도지사, 이재명 성남시장. 김경수 경남지사 등 각계 인사 500여명이 참석했다.</d> 서울중앙지방법원은 지난달 30일 열린 선고공판에서 "피고인은 피해자 유족들로부터 합의금 명목으로 1억원을 받아 가로챘다"며 징역 1년6월에 집행유예 2년, 추징금을 명령했다.

#3 프로젝트 결과

- 아쉬운 결과

[입력 키워드] 가까운



생성 줄거리 :

가까운 녀석들을 만나고 싶다고요 그럼 우리도 같이 해볼래욧 <unk> ㅋ 그치만 좀 도와줘용
진짜 귀여운 고양이들 아니면 저기 있는 거야 난 고양이가 아니거든 왜 자꾸 사람을 마주치
는 거지 궁금해 요거 이게

- ‘<unk>’로 표현해야 하는 OOV(Out of Vocabulary) 문제가 발생
- 자연스럽지 않은 흐름

#3 프로젝트 결과

- 아쉬운 결과

[입력 키워드] 복잡한 관계



생성 줄거리 :

복잡한 관계 속에서 펼쳐지는 그들의 이야기 어느 날 갑자기 나타난 정체불명의 괴물들 때문에 모든 것이 엉망이 되어버린 세상 속 평범한 고등학생들의 성장 드라마 대한민국 명작 만화 리메이크 2 네이버 웹툰 작가들이 들려주는 2015 대학만화 최강자전 수상 작품입니다
한국

- “2015 최강자전 수상 작품” 과 같이 줄거리와는 관련 없는 내용이 생성
- “어느 날 갑자기 나타난 정체불명의 괴물들” 이라는 내용이 다른 입력 키워드에 대해서도 자주 등장

#3 프로젝트 결과

- 성공적인 결과

[입력 키워드] 첫사랑

💬 첫사랑을 하며 살고 있는 민준은 어느 날 갑자기 나타난 수상한 아이와 함께 평범하게 학교 생활을 시작한다 그런데 이 아이가 바로 전학생 강찬성 이라는 걸 알고 당황하는데 과연 그의 목적은 무엇일까 궁금증을 자아낸다

#3 프로젝트 결과

- 성공적인 결과

[입력 키워드] 꽃처럼




생성 줄거리 :

꽃처럼 맑은 눈빛으로 자신을 뽐내는 남자 백아진 과연 그녀는 자신의 마음을 전할 수 있을까요 사랑스러운 외모와 달리 냉철하지만 소심한 성격 때문에 항상 고생하는 민석이의 좌충 우돌 성장 스토리

#3 프로젝트 결과

- 성공적인 결과

[입력 키워드] 10년 후

 생성 줄거리 :

10년 후 그는 자신의 과거를 숨기고 자신을 변호할 수 있는 사람을 찾고 있었다 하지만 그 사람은 바로 과거 자신이 저지른 살인 사건에 연루된 사람이었고 결국 그의 손을 들어줬다는 것 이 모든 것은 그가 과거에 저질렀던 범행을 감추기 위한 것이었고 또 다른 범죄도 숨겨야 한다 이제

#3 프로젝트 결과

- 성공적인 결과

[입력 키워드] 10년 전




생성 줄거리 :

10년 전 사고로 인해 기억을 잃은 주인공 한솔은 우연히 과거의 상처를 치유하고 다시 살아 나기 위해 과거로 돌아가고 싶다는 마음을 먹는다고 한다 그런데 이 때 과연 자신의 과거를 극복하고 원래의 삶을 찾을 수 있을까

#3 프로젝트 결과

- 성공적인 결과

[입력 키워드] 은빛 금

 생성 줄거리 :

은빛 금붕어 그리고 인간들의 이야기 어느 날 갑자기 나타난 정체불명의 괴물들 때문에 모든 것이 엉망이 되어버린 세상 속 평범한 고등학생이었던 한별은 우연히 자신의 몸에 갇혀 있는 것을 알게 된다 하지만 그 순간부터 자신이 원하는 것은 무엇이든 이룰 수 있다는 사실을 깨닫게 되는데

프로젝트 의의 및 발전 가능성

AI 창작 분야를 웹툰으로 확장하여 **단어만 입력해도 웹툰 줄거리를 생성**해주는 모델 구현

웹툰 분야 외에도 동화, 드라마, 영화 줄거리 등 확장 가능성

Seq2GAN

대표적인 비지도 학습 모델인
GAN을 문장 생성에 활용

장르 별
생성

웹툰 줄거리별 장르 라벨링을 추가하여
장르별 웹툰 줄거리 생성

하이퍼파라미터
튜닝 및 데이터
정제

하이퍼파라미터 튜닝 및
불필요한 문장을 제거한다면
더욱 매끄러운 흐름의 줄거리 생성 가능

감사합니다

트렌드리서치 팀

복숭아눅차 김세영 김지후

Q & A