

# 제 2회 AI x Bookathon

Team CONLP

인공지능융합전공 고준서 김민지 이상은

# 인공지능 조력자

AI as a Supporter

인간의 문맥 구성능력

+

인공지능의 생산력

# AI가 할 수 없는 것

우리가 할 수 있는거에 집중하자

입력 데이터의  
내용과 맞춤법이  
적절한지 파악

글의 내용이  
사실인지 아닌지  
파악

모든 글이  
적절한 문맥으로  
흘러가는지

# 우리가 할 수 없는 것

우리가 할 수 있는거에 집중하자

모델 아키텍처 변경

사전학습 모델 수정



Fine Tuning

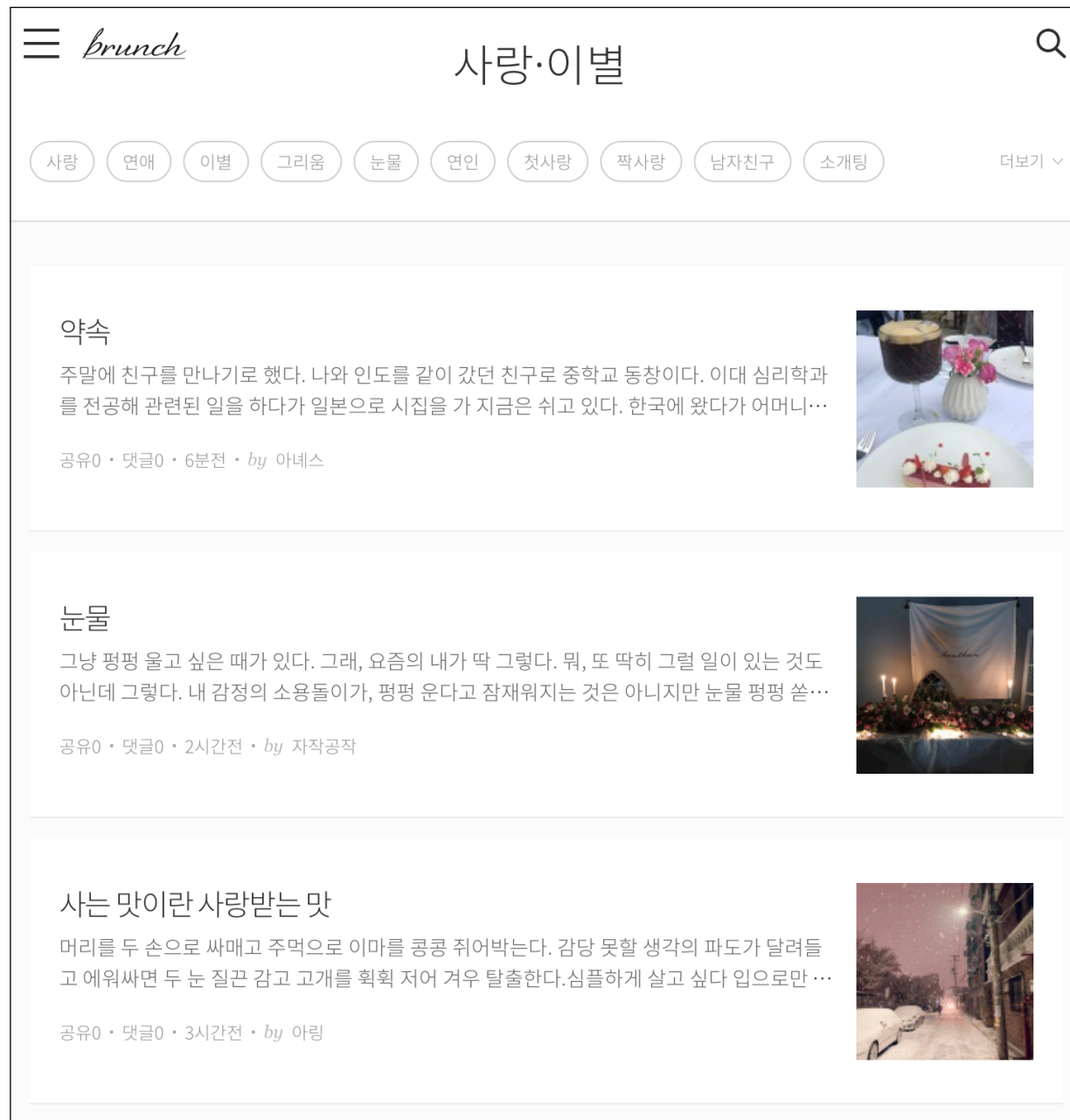
데이터 정제

전이학습을 활용한  
모델 훈련

훈련시 발생하는  
문제 대처

# Data Collection

## 데이터 크롤링



“여행을 테마로 한 글을 크롤링”

# Data Collection

데이터 크롤링

관동별곡

혼자 떠나는 여행

오직 한사람을 위한 여행

일동장유가

여행의 이유

서유기

# Data Collection

## 양 VS 품질

난 누구고, 어떤 내가 되고 싶은 거니? "윤진, 넌 너를 누구라고 생각해?"라고 스스로 물어보았다. 묻고 보니 질문 자체가 적절한 거 같진 않다. 왜냐면, 시간이나 장소가 바뀌면, 나 스스로 낯선 내 모습을 발견할 때도 있고, 십 년 전에 있던 나와 지금의 나도 다를 거니까. 아마 2~3년 후의 내 모습 또한 지금의 나와는 같지 않을 거다.

### 좋은 품질의 텍스트

^^문화와 접목된 새로운 콘텐츠, 더불어 1인 미디어에 관심 있으신 분들의 많은 참여 바랍니다. 보다 자세한 사항은 아래 링크를 참고하시면 됩니다.  
<https://cafe.naver.com/mpiousernetwork/17830>문화로(路) 프로젝트 [1인크리에이터전문가 2급 민간자격증 교육]문화로(路) 프로젝트 문화로...

### 나쁜 품질의 텍스트

띄어쓰기 오류, 광고성글, URL, 한자...

# Data Collection

양 vs 품질

날 것의 텍스트



학습에 용이한 텍스트 형태

전처리가 할 수 없는 것: 내용에 관한 것



# Data Collection

높은 품질의 데이터를 위해

직접 내용을 검토하고 수정하는 작업을 거침

광고성 글  
URL  
해시태그  
약간의 맞춤법 수정

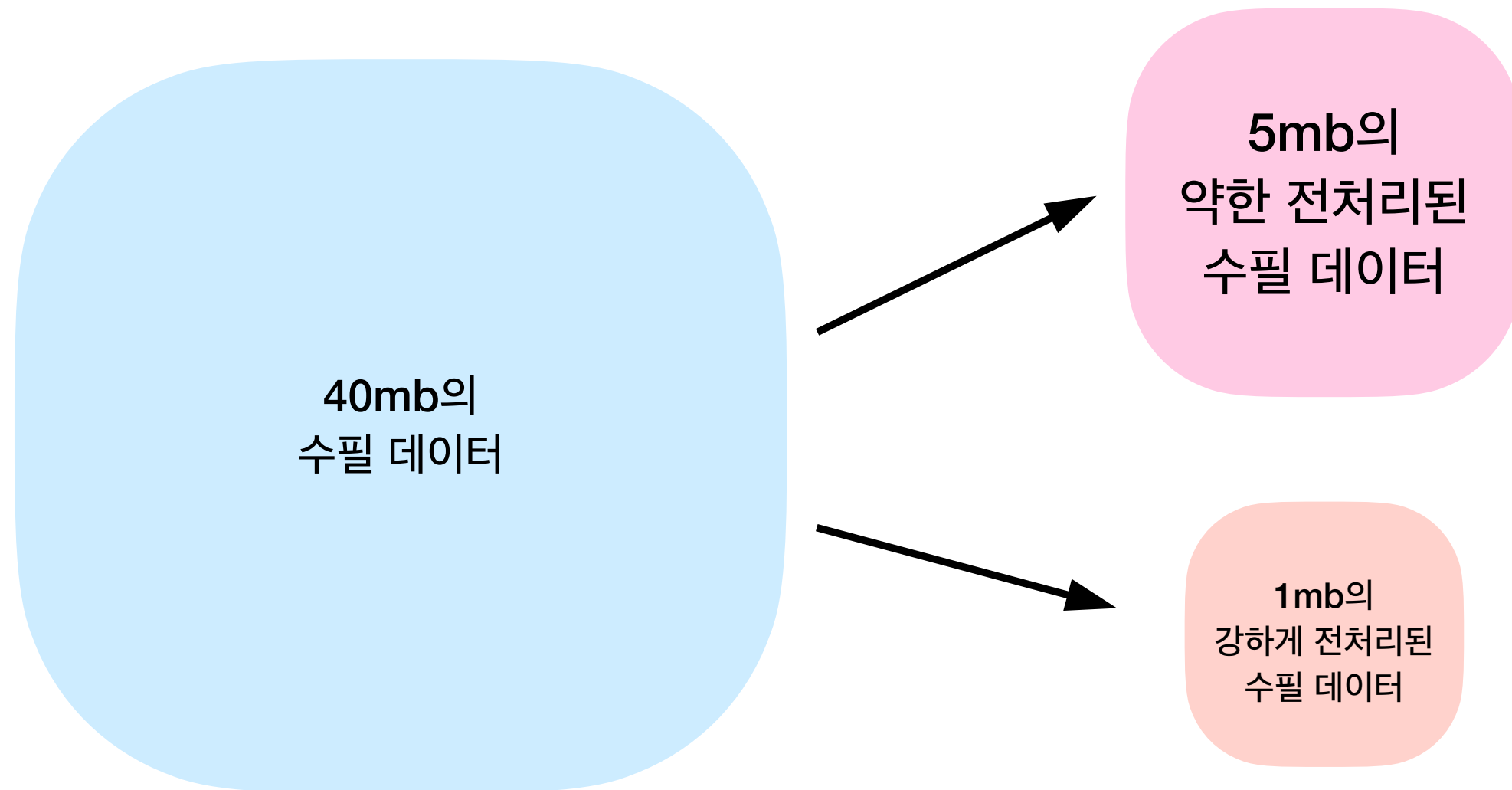
빠른 훈련 시작을 위한  
약한 전처리

광고성 글  
URL  
해시태그  
내용이 별로인 글 제거  
대화체 제거  
내용 안맞는 부분 다듬기  
문맥 다듬기  
강한 맞춤법 검사

모델의 정교화를 위한  
강한 전처리

# Data Collection

사용한 데이터



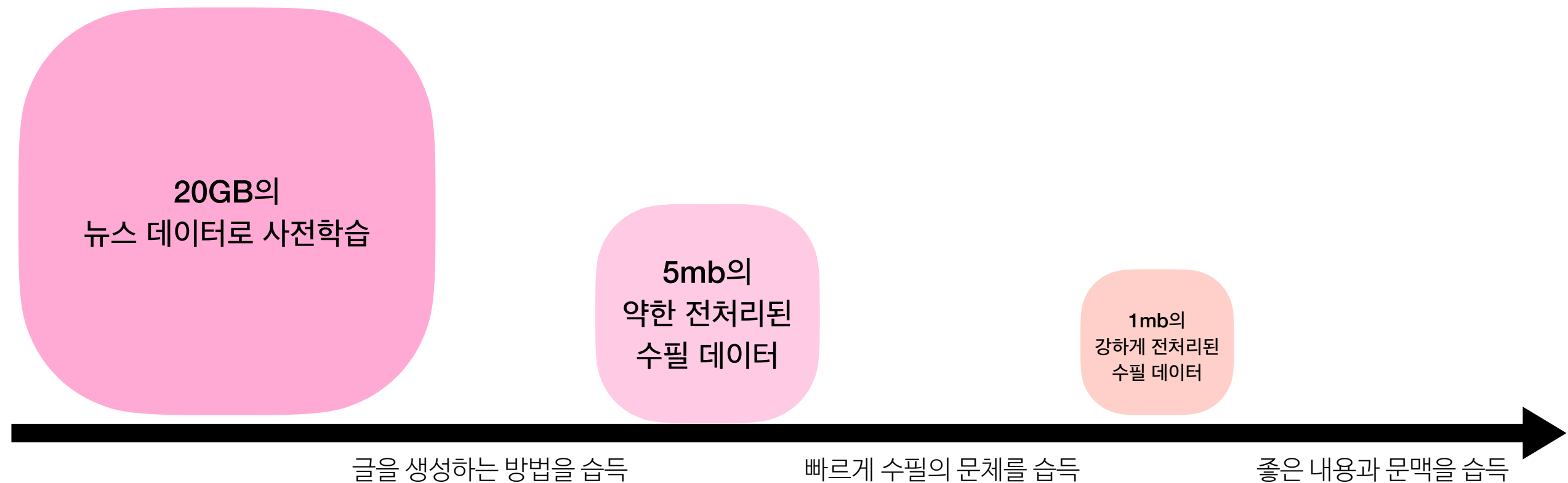
# Model GPT-2

모델 원리



# Model Training

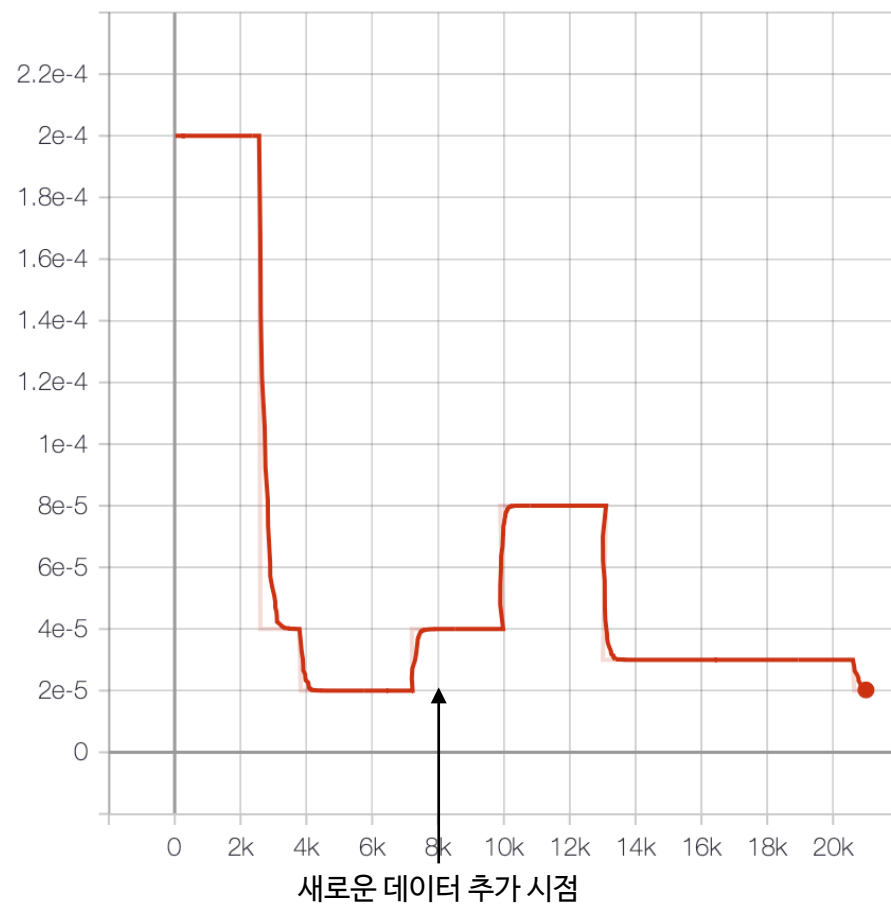
## 모델 훈련 전략



# Model Training

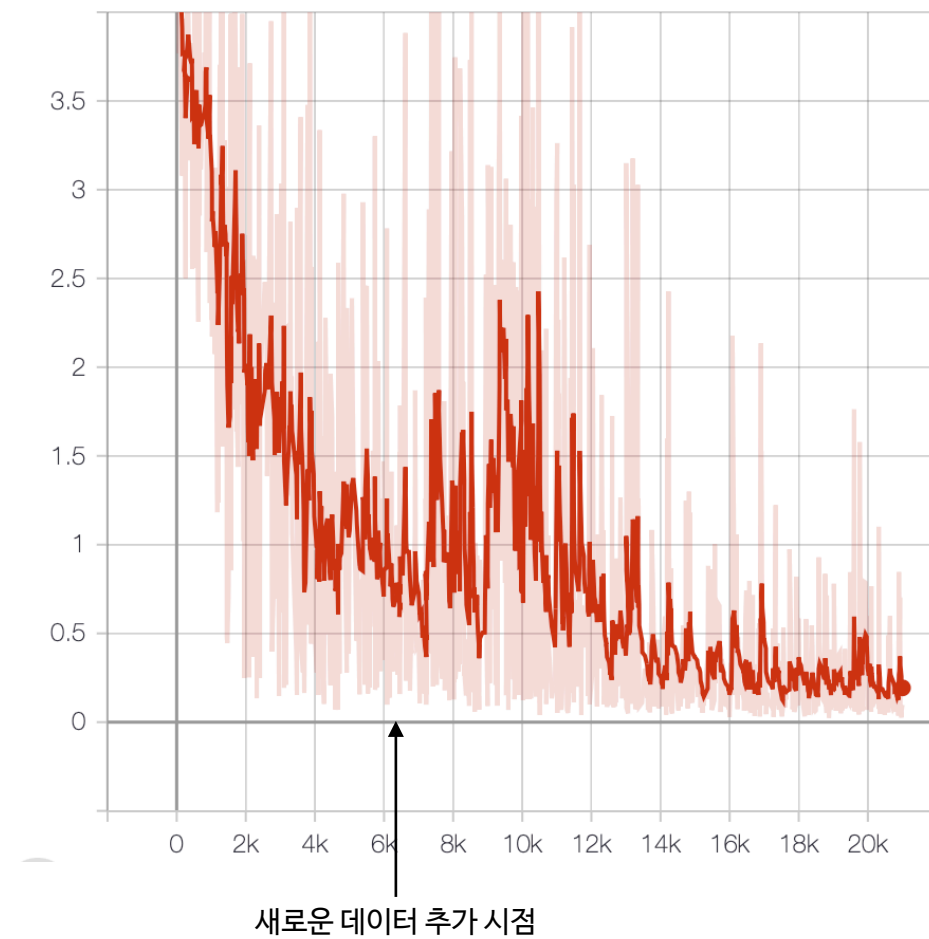
## 모델 훈련 전략

learning\_rate



Learning Rate Decay

loss



23,000 Steps Training

# Model Training

학습 중간 단계에서 생성한 원본 텍스트

3000 Steps

**한국으로 돌아왔다.** 쿠바로 갈 예정이었지만 인천 공항에서의 실수로 인해 비행기가잡은 공항에 문제가 생겼다. 비행기는 쿠바를 출발해 한국으로 돌아갔다. 인천에서 쿠바를 경유하는 대신 대한항공을 이용했는데, 실제 비행시간은 1시간 50분 정도였다. 쿠바를 입국하는 비행기에 문제가 생긴 건 이번이 처음이었다.

7200 Steps

**한국으로 돌아왔다.** 쿠바 여행 중 만난 한국인 여행자에게 매력적인 첫인상을 남긴 곳은 라벨로이다. 스페인의 정취를 닮은 붉은 지붕이 인상적인 라벨로에 도착하니 인테리어는 한결 부드러워졌고 분위기도 매우 순조로웠다. 역시 스페인이 딱 그랬다. 스페인은 세계여행의 성지다.

12000 Steps

**한국으로 돌아왔다.** 쿠바 여행 중 만난 한국인 여행자에게 쓰러졌던 일화는 유명하다. 공산정권 하의 쿠바 공산주의와 자본주의의 영향을 받았던 쿠바는 물놀이 강국이었다. 그곳의 수질은 캐나다 밴쿠버의 1. 2 비와 어린이의 공간으로 소문이 자자한데 이마저도 물놀이 용품을 파는 분수대로 다시 돌아왔다.

16000 Steps

**한국으로 돌아왔다.** 쿠바로 갈 때까지만 해도 좋았는데. '비행기값이 50만원 정도 나왔구나.' 라는 생각이 들었다. 앞으로는 무조건 싼 티켓으로 와야겠다. 100만원 넘는 걸 샀는데 더 비싸졌구나. 아찔했다. 내 생일은 친구의 지인과 친구의 지인의 SNS를 통해 빠르게 퍼졌다.

# Generated Text

최종 모델 (21,000steps)이 생성한 텍스트 예시

“ 삶은 우리가 인생의 아름다운 균형을 찾는 방법이 아니라, 그저 자신이 정한 삶의 양식과 취향을 그대로 드러내는 하나의 예술작품으로 남게 되는 것이다.”

# Generated Text

텍스트 수정 과정에서 두가지 원칙

## 인공지능 조력자

AI as a Supporter

1. 모델이 생성한 문장 이외에 새로운 문장을 절대 넣지 않는다.
2. 새로운 문장이 필요할 경우, 모델이 생성하도록 한다.
3. 필요 없는 부분을 삭제하는 작업을 거친다.
4. 문맥에 맞지 않는 단어만 단어 단위로 수정한다.



# Generated Text

제목 - 저마다의 길: 여행끝에서 찾은 해답

주제 - 길

기

친구가 죽음의 길로 향해 갔고, 오랜시간 의지했던 친구이기에 슬픔에 빠지고 여러 고민을 함

승

세계 여러곳을 아무 생각없이 여행하며 즐거움으로 잊으려 하지만, 잘 되지 않아 여러 고뇌에 빠짐

전

순례길을 한발자국 생각하며 걸으며 친구가 걸은 길을 걸을 수 없고, 각자의 목적지를 향해 길을 걸어가야 한다는 것을 깨달음

결

깨달음을 바탕으로 한국에 돌아와 자신의 길을 걸어가기로 결심함

감사합니다.