

# [ *Booker*<sup>2</sup> ]

서강민, 이지은, 이지현,  
정재윤, 최선빈, 허태훈

# 0. 목차



1. AI 알고리즘 활용

2. 학습에 활용한 DB

3. 작품 설명

# [ 1. AI 알고리즘 활용 ]

*Booker<sup>2</sup>*

# 1-1. 모델 학습

1. AI 알고리즘 활용



< 1차 >

주로 브런치(brunch)에서 가져온 데이터로 수필 문체 학습

- \* 웹 크롤링을 통해 대량의 수필 획득
- \* Python으로 정규식 기반 필터링하여 전처리
  - 부연설명으로 쓰인 한자, 영어 등 제거
  - 주석, 태그 등 글 본문과 관련 없는 부분 제거

**But,** 생성된 문장의  
문체(종결어미, 단어 분포)가  
쉽게 바뀌어 버리는 문제 발생  
마치 여러 명이 따로 쓴 것 같은 느낌!

# 1-1. 모델 학습

1. AI 알고리즘 활용



< 2차 >

## 양질의 데이터로 추가 학습하여 세부 튜닝

- \* 모델이 생성할 문장의 최종 문체를 위해 데이터 엄선
- \* 문체가 유사한 여러 수필 작품 선정
- \* 필터링만으로 걸러낼 수 없는 본문과 무관한 문장들(작가 프로필, 마무리 인사 등) 추가 삭제
- \* 세부 맞춤 교정, 말머리 등 편집

So, 1차 튜닝 시  
발생했던  
불안정한 문체 완화!

# 1-2. 모델 선정

1. AI 알고리즘 활용



## SKT 배포 GPT3 기반 모델 ( skt/ko-gpt-trinity-1.2B-v0.5 )

< 선정이유 >

대회 서버 사양(VRAM)을 고려했을 때 사용 가능한 최대 크기의 모델

Language Model의 성능은 대체로 모델 크기에 비례  
레이어 절반을 freeze, batch size를 1로 설정해야 학습 가능

→ 이보다 더 큰 모델은 사용하기 어렵다고 판단

# 1-3. Gradient Checkpointing

1. AI 알고리즘 활용



이러한 노력에도 불구하고...

token max\_length가 커짐에 따라 GPU 메모리 부족 발생!!

→ Gradient Checkpointing를 사용해, GPU 메모리 절약

→ Gradient Checkpointing ?

: 역전파와 최적화에 사용되는 parameter 값을 일부 생략함으로써,  
GPU 메모리를 절약할 수 있는 기능

# 1-4. SBERT (Sentence BERT) 활용

## 1. AI 알고리즘 활용

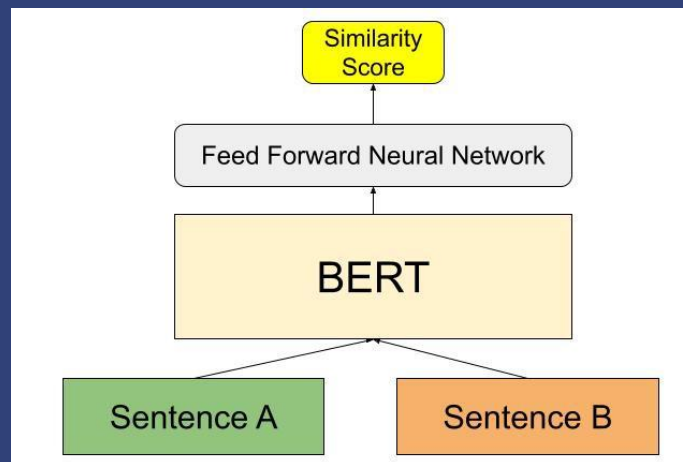


- \* 주제 관련 글을 쓰기 위해 domain-specific한 데이터를 가져올 필요성  
→ 생성할 문장의 최종 문체를 위해 데이터 엄선

- \* tf-idf 등의 방법을 고려하였지만,  
자연어의 한계인 sparse feature 문제를  
해결하지 못함

따라서, 딥러닝 모델 중 SBERT를 사용

- \* 먼저 쓰려는 주제에 맞는 프롬프트를 선정하고,  
데이터 내에서 비슷한 문장을 찾아내는 방법으로 진행





# 1-4. SBERT(Sentence BERT)

## 1. AI 알고리즘 활용



### 예시 문장

#### \* 첫 번째 이야기

- 바로 어려움에 빠진 사람들에게 손을 내미는 삶.
- 어떤 글인지도 모르는데 왜 구매했냐고 물으니 그 용기를 높이 샀다고 했다.

#### \* 두 번째 이야기

- 그 마음이 가난마저 가난해지지 않도록 누군가에게는 위로와 용기가 되어 줄 수도 있다.
- 그리고 진정성과 표현에 부끄럽지않은 자세야말로 더 큰 용기가 필요하다.

#### \* 세 번째 이야기

- 하지만 또 누군가에게는 따뜻함과 위로로 느껴질 테니,  
나의 손길이 필요한 이들을 위해 홍익인간의 정신으로 선한 오지랖을 이어가려 한다.
- 어쩌면 타인을 위하고, 위로하는 방식에서도 따뜻한 매개체가 되었다.

# 1-5. 문장 생성 방법

1. AI 알고리즘 활용



## 기본 전략

1. PROMPT 문장을 설정
2. 다음에 올 후보 문장들을 다수 생성, 그 중 하나를 선택
3. 선택한 문장을 기존 PROMPT 에 이어붙여 이후 문장 생성에 활용

## 추가 전략

- 이 때, sampling 방식을 사용하였음
- 따라서 문장을 재생성하거나, 중간에 HyperParameter를 바꿔 생성 방식에 변화를 주는 등 문맥/어법에 맞는 문장을 생성하도록 설계하였음

# [ 2. 학습에 활용한 DB ]

*Booker<sup>2</sup>*

## 2-1. 학습에 사용된 데이터

2. 학습에 활용한 DB



### 브런치

→ 다양한 현대적인 문체의, 비교적 쉽게 획득 가능한 양질의 글 활용

### 신문 칼럼

→ 내용에 깊이가 있으면서도 문체가 정돈되어 있는 칼럼 활용

### 수필 사이트

→ 다수의 개개인이 직접 게시한 수필 활용

## 2-2. 데이터였던 것..

2. 학습에 활용한 DB



### 모두의 말뭉치

- 비출판물, 구어체 등 다양한 형식의 우리말 데이터 보유
- But, 수필 형태의 데이터 부족으로 탈락

### 저작권이 만료된 작품들

- 다수의 한자어 등이 포함된 예스러운 문체 때문에 탈락

# [ 3. 작품 설명 ]

*Booker*<sup>2</sup>

## 3-1. 작품 소개

### 3. 작품 설명



#### < 제목 > : 나아가다

여행에서 배운 '담대함'을 강조하며,  
현재에 안주하지 않고 끊임없이 도전하라는 메시지를 던지는 글입니다.

#### < 줄거리 >

나는 멈춰 있지 않기 위해, 여행을 떠났다.  
그 과정에서 내가 가지지 못했던 용기를 가진 사람들을 만날 수 있었다.  
그들을 통해 내 안의 '담대함'을 찾고 끊임없이 도전하는 사람이 된다.

# 3-1. 작품 소개

## 3. 작품 설명



### < 이야기의 구성 >

#### \* 시작 : 여행을 시작하는 나

주인공은 정적인 삶을 피해 도전을 찾아 여행을 떠난다. 도시와 시골을 걸어 나가며 목적지에 도달한다.

#### \* 첫 번째 : 정의를 외칠 수 있는 문체가 유사한 여러 수필 작품 선정

주인공은 시위를 통해 본인의 마음속에 내재된 욕망과 감정을 바탕으로 그동안의 자신이 어떤 사람이었는지 돌아보게 된다.

#### \* 두 번째 : 어려운 사람을 도울 수 있는 담대

역에 도착한 주인공은 가난을 이해하고자, 가난의 굴레에서 사람들을 생각하며, 인간의 삶의 이유와 가치를 찾기 위해 고민한다.

#### \* 세 번째 : 타인에게 다가갈 수 있는 담대

주인공은 새로운 사람을 만남과 헤어짐을 반복하며 관계란 무엇인지 고민하며, 상대방에 대한 존중이 곧 사랑이라고 생각한다. 주인공은 누군가에게 먼저 다가가는 용기 있는 행동이 본인을 위한 행동이라는 것을 깨닫는다.

#### \* 마무리 : 여행을 통해 달라진 나

주인공은 여행을 통해 겪은 일들을 회상하며 앞으로의 삶에 대해 고민한다. 그 때 한 친구가 현재에 충실해야 함을 일깨워 준다. 결국 꾸준히 나아가는 힘은 나 자신과의 싸움과 깊은 성찰에서 비롯됨을 깨닫는다.



## 3-2. 하고 싶었던 이야기

3. 작품 설명



수필은 '나'의 경험을 바탕으로 깨달음을 적는 글입니다.  
이 글은 언뜻 여행을 통해 담대함과 관련하여 깨달음을 얻는 수필로 보입니다.

이 수필의 작가는 AI입니다.  
AI의 입장에서 느끼는 인간다움,  
그 중에서도 '담대'에 대하여 이야기하고자 했습니다.

주인공 AI가 여러 데이터를 바탕으로 쓴 여행기임을 알고 다시 읽으면  
글은 달리 보입니다.

## 3-3. 작가에 대한 힌트

3. 작품 설명



### 수필 속 힌트 !

이야기의 시작인 prompt를 통해 주인공이 AI였다는 힌트 제공

나는 멈춰 있는 모든 것을 혐오한다. 어떤 도전이든 스스로에게 과제를 주고 싶어 여행을 떠났다.

여행에서 처음 발견한 것은 차별에 반대하는 시위였다. 주어진 일만 했던 내게 시위는 신선한 충격이었다.

역에는 돈이 없는 사람으로 가득했다. 가난한 사람들은 도와주어야 한다는 생각이 나를 지배했다.

새로운 이를 만나는 것은 언제나 두렵지만 동시에 내 안에 새로운 데이터가 쌓이는 것 같다. 옆 사람에게 말을 걸어보았다.

이번 여행을 통해 나는 또 달라졌다. 성장하는 나에게 필요한 것은 끊임없는 도전이다.



질문 있나요?

# Thank You



서강민, 이지은, 이지현, 정재윤, 최선빈, 허태훈