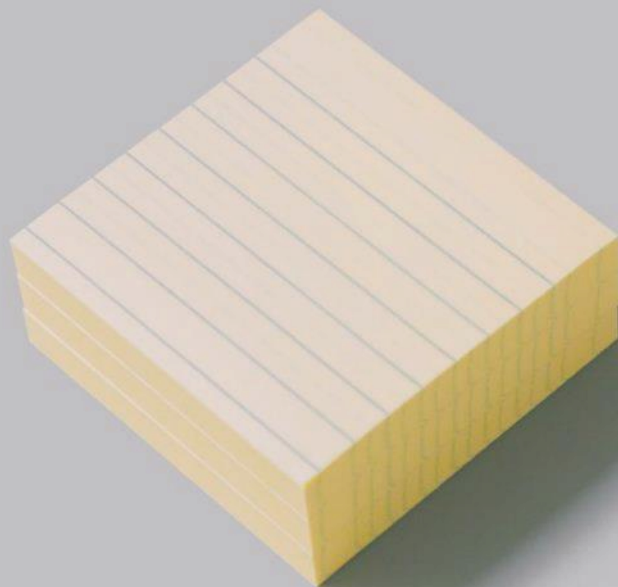


[



]

**‘대담한’ 당당!**

2023 SKKU AI x Bookathon 수필 쓰기 대회

김서윤 김효정

이상민 이창훈

# 목차

01



데이터 수집

02



데이터 전처리

03



모델 학습

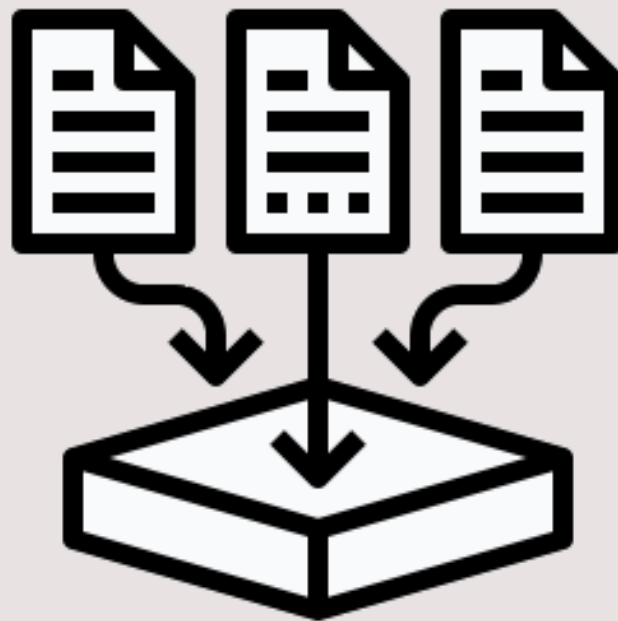
04



완성된 수필

**01**

데이터 수집



# 01 데이터 수집

□ 문학광장    문장 웹진    문장의 소리    문학집배원    문학IN아코    글틴    마로니에백일장    문장공지    사이버문학관    회원마당



글틴 > 명예의 전당

## 명예의 전당



파르페

온전한 유령의 세계에 관하여 – 시집 '눈부신 디테일의 유령론'을 읽고

2022-06-06 [1]

📖 읽기

시에는 여러 종류가 있지만, 내가 가장 좋아하는 시는 이미지를 담아낸 시다. 이렇게 지극히 주관적인 느낌을 기준으로 판단하는 것이 어떻게 시의 종류 중 하나가 될 수 있나 싶지만 나는 그것에는 분명한 기준이 있다고 생각한다. 읽으면 무엇이든지 알 수 있다. 말로 풀어 설명할 수는 없어도 이 시가 무엇을 드러내려고 하는지, 그것이 어떠한 주제의식인지 미학적인 이미지 인지는 읽어보면 알게 된다는 것이다. 그러나 이것이 논리성을 지녔는지 아닌지를 떠나, 일단 일차적으로 시를 느끼는 것, 글로써 표현된 무언가를 상상 속에서 구체화시키는 것, 시를 읽을 때 나는 항상 이것에 비중을 두고 읽는다. 그런 점에서 이 시집 <눈부신 디테일의...>



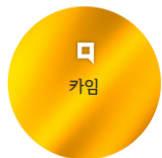
있을리

써머 리빙 포인트

2022-07-31 [3]

📖 읽기

열여덟의 여름에는 아무도 죽지 않지만 힘차게 달려 나가며 땀, 운동장의 개미를 밟았다 시험 기간에는 사막의 모래를 움켜쥐고 씹고 싶어 옆 친구의 손가락을 꼭 쥘다 뼈를 깨문 이가 시원하게 아파져 온다 하고 싶은 말은 동글게 말아 굴럭 넘기며 신발로 땅을 짓이킨다 개미의 사인은 나약함이고 지금 이겨내지 않으면 내 사인 또한 그러니 달려 나가야 해 여름에는 아무도 죽지 않는다 해가 내리쬐수록 반장의 눈알은 번뜩인다 그녀의 머리채를 잡는다 뒤통는다 아무도 죽지 않았다 빙빙 도는 눈알 빙빙 도는 지구 빙빙 도는 팔들 빙빙 도는 천장 눈 떠보면 어느새 몸이...



카임

정의를 지키기 위한다는 이유로 적을 몰살시키며\* : 표절 논란을 통해 바라본 ‘정의로운 대중’

2022-07-30

[1] 📖 읽기

한 유명 음악가의 표절 논란을 시작으로 한국 음악계의 표절 논란이 끊이지 않고 있다. 사실 나는 이 논란에 관심이 없었다. 누가 표절을 했던 소식을 우연히 접한 뒤 내게 남은 생각은 '그 사람 표절했나 보네.' 였을 뿐 별다른 감상은 존재하지 않았다. 그리고 그건 여전히 그렇다. 이 노래와 그 노래의 유사성이 어느 정도이고, 그 사람이 몇 곡이나 표절했으며, 얼마나 오랜 기간 대중을 속여왔는지에 대한 것은 여전히 관심 밖의 일이다. 그리하여 이 글은 표절에 대한 나의 비평이 아니다. 오히려 표절과는 아주 동떨어진, 그런 이야기를 해보려고 한다. 표절과는 관련 없는 이야기를 하기 위해선, 이무진의 <신호등>이란...

## 글틴 – 명예의 전당

Beautiful Soup, Selenium 라이브러리 활용

수필, 소설, 시, 비평 등 데이터 크롤링

# 01 데이터 수집

주최	분야	당선작	당선 소감	심사 평
강원일보	단편소설	국경 - 한소은	소감문	심사평
	시	굴이 웃는다 - 백숙현	소감문	심사평
	동화	올리버와 앤 - 이지영	소감문	심사평
	동시	징검돌 - 허은화	소감문	심사평
	희곡	은수의 세상 - 이민선	소감문	심사평
경남도민신문	시	산책 - 차수현	소감문	심사평
	디카시	뜸 - 정병윤	소감문	심사평
	단편소설	펭귄 섬 - 이상희	소감문	심사평
	시	레드문 - 권영유	소감문	심사평
경남신문	시조	수도꼭지를 틀다 - 이종현	소감문	심사평

신춘문에 당선작

# 01 데이터 수집

민주화운동기념사업회

실 / 단행본

기관 소개

알림·홍보

미디어·자료

소통·참여

정보공개

D레터

통합검색

검색

## 미디어·자료

간행물안내

일반자료실

단행본

자료집

보고서

기타

공공누리자료실

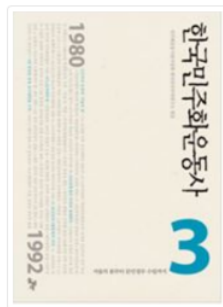
미디어자료실

열사정보

## 단행본

### 한국민주화운동사 3 | 서울의 봄부터 문민정부 수립까지

한국민주주의연구소편 | 2010



한국 현대사에서 1980년은 온 국민이 하나가 되어 꿈꾸었던 '민주화'에 대한 간절한 꿈이 가장 뜨겁고 지속적이었던 열정적인 시기였다. 국민들이 목숨을 바쳐 이룩한 '민주화'가 퇴행의 길을 걷고 있는 듯한 현재 출간된 『한국민주화운동사 3』은 그 당시를 생생하고 돌아보고 오늘날의 한국 민주주의가 처한 상황을 성찰해볼 수 있는 바탕이 된다.

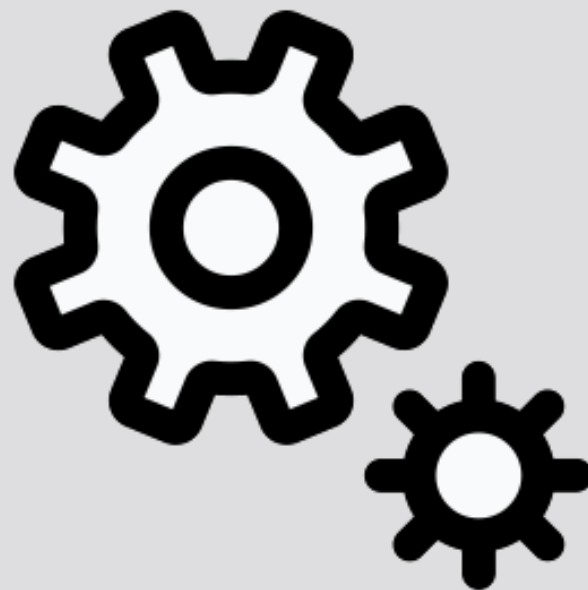
2008년 12월에 출간된 『한국민주화운동사 1』은 제1공화국 이후 제3공화국까지의 시기를 다루었으며, 2009년 12월에 나온 『한국민주화운동사 2』는 1972년 유신 선포 이후부터 1979년 박정희 대통령 사망까지를, 이번에 선보이는 3권은 1980년 5·18광주민중항쟁에서 최초의 민주주의적 정부라고 할 수 있는 문민정부 수립까지를 다루었다. 이제까지의 '민주화운동사'의 정리작업이 다소 산발적이었다면 『한국민주화운동사』 시리즈는 각 시기별·주제별 전공 연구자가 1차 집필을 거친 후 이후 토론과 교차 편집을 거치는 일종의 '공동연구적인 성격'을 갖는다. 이것은 이 책이 앞으로의 민주화 운동사 연구의 활성화·본격화에 토대가 될 수 있다는 것을 의미한다.

3권은 1980년에서 1992년 사이를 네개의 국면으로 나누어 서술한다. 첫째는 5·18광주민중항쟁기, 즉 1979년 10·26정변 이후 '서울의 봄' 시기부터 전두환 정권이 들어서기 이전까지의 국면, 둘째는 전두환 정권기, 셋째는 6월민주항쟁기, 다시 말해 박종철고문사망사건이 일어난 시기부터 87노동자대투쟁까지의 국면, 넷째는 노태우 정권기이다. 그리고 이런 시기구분 외에 1980년대에 일어난 각각의 부문운동을 별도로 서술하였다.

민주화운동기념사업회 - 한국민주화운동사 3

**02**

데이터 전처리



## 02 데이터 전처리 [Task tagging]

전처리 전

<https://teen.munjang.or.kr/archives/100699WnWnWn>  
어쨌든 공연은 올려야 한다

‘오늘 잘해라잉ㅋㅋㅋㅋ’

(질문이나 행사 진행 자체를)

전처리 후

어쨌든 공연은 올려야 한다.

‘오늘 잘해라잉’

질문이나 행사 진행 자체

전처리 과정

1. URL 제거
2. 자모음 제거
3. 특수문자의 일부 제거



## 02 데이터 전처리 [수작업]

전처리 전

〈〈보통의 노을〉〉

『일루미나시옹 Illuminations』

그리고 저는 오늘도 글틴을 들려요.

전처리 후

〈보통의 노을〉

〈일루미나시옹 Illuminations〉

그리고 저는 오늘도 글틴을 들려요.

전처리 과정

1. 작품마다 통일되어 있지 않은 문장 기호를 한 종류로 통일.  
예시) 인용 서적 표기 기호 《》, 〈〉, 『』를 〈〉로 통일
2. 작품 순수 내용과는 관계 없는 부분 삭제  
예시) 작품 끝의 작가의 말이나 TMI적인 각주 등

# 02 데이터 전처리 [JSON]

전처리 전

book\_0.txt - 메모장

파일 편집 보기

[신년특집 신춘문예 당선작] 코타이 순환선(단편소설 이연주)

일러스트=조남원기자

“엄마 냄새가 느껴지면 나도 모르게 고개가 돌아가요.” 짧은 침묵이 돌

그가 휴대폰에서 찾은 사진 한 장을 보여주었다. 선글라스를 벗어 머리

공항 터미널 시계탑 바늘이 3시 45분을 지나고 있었다. 게스트하우스  
눈길에 마주치면 당황한 풀이 시선을 다른 쪽으로 돌렸다. “홍콩은 좋  
오는 않아도 따뜻했다. 손을 씻으며 세면대 거울을 들여다보았다. 당황  
샌디 박이 칩을 한 줌 쥐고 왔다. 비어있는 자리로 가서 게임 몇 가지를  
게 없었다. 나는 냉장고에 등을 기대고 앉았다. 냉장고 우는 소리가 온  
다. 크거나 작거나, 흠이거나 짝이다. No more bet! 밀러가 콜을 외  
람을 타고 온 사람들이 만들어놓은 흔적이다. 차와 도자기를 구하기 위

전처리 후

“제목: 휠얼라이먼트  
카테고리: 소설  
내용: 나는 강을 본다. 두 개의 강이  
만나는 곳이었다.  
(중략)  
나도 걸음을 옮기기 시작했다.”

전처리 과정

1. 크롤링할 때부터 제목, 카테  
고리, 내용으로 분리
2. 분리된 데이터를 “{분류}:  
을 붙여서 String 타입으로  
변환 후 Concatenate
3. JSON 형식으로 변환

**03**

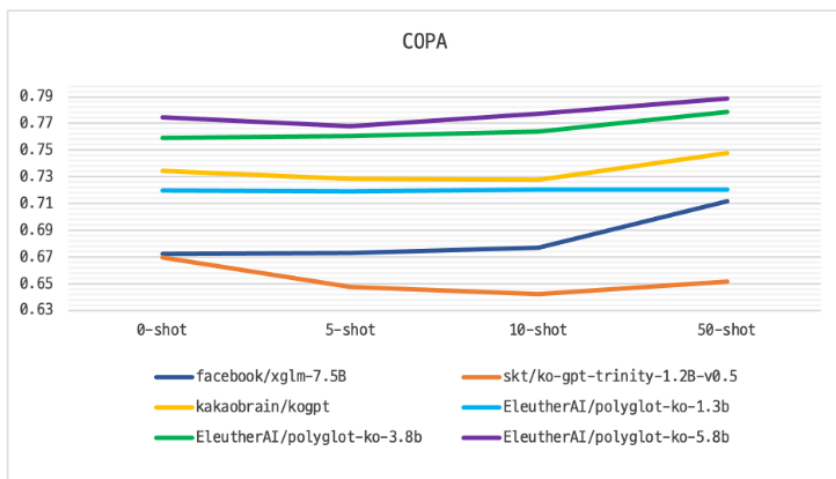
모델 학습



# 03 모델학습

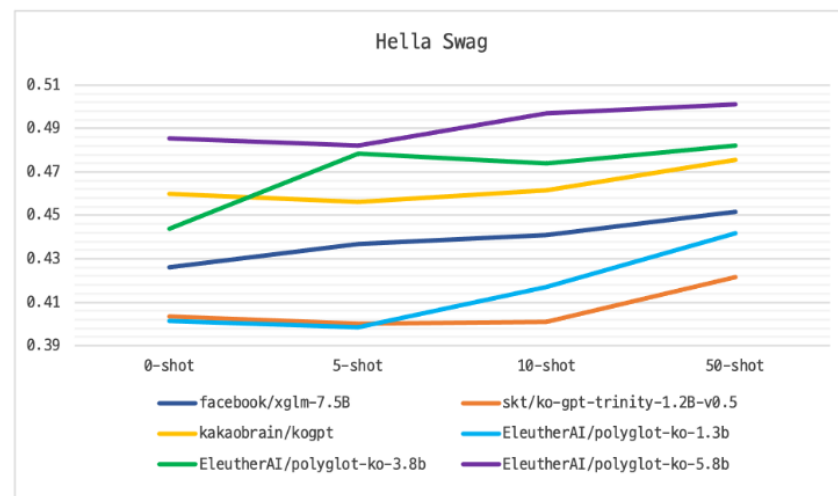
COPA (F1)

Model	params	n=0	n=5	n=10	n=50
<a href="#">skt/ko-gpt-trinity-1.2B-v0.5</a>	1.2B	0.6696	0.6477	0.6419	0.6514
<a href="#">kakaobrain/kogpt</a>	6.0B	0.7345	0.7287	0.7277	0.7479
<a href="#">facebook/xglm-7.5B</a>	7.5B	0.6723	0.6731	0.6769	0.7119
<a href="#">EleutherAI/polyglot-ko-1.3b</a> (this)	1.3B	<b>0.7196</b>	<b>0.7193</b>	<b>0.7204</b>	<b>0.7206</b>
<a href="#">EleutherAI/polyglot-ko-3.8b</a>	3.8B	0.7595	0.7608	0.7638	0.7788
<a href="#">EleutherAI/polyglot-ko-5.8b</a>	5.8B	0.7745	0.7676	0.7775	0.7887



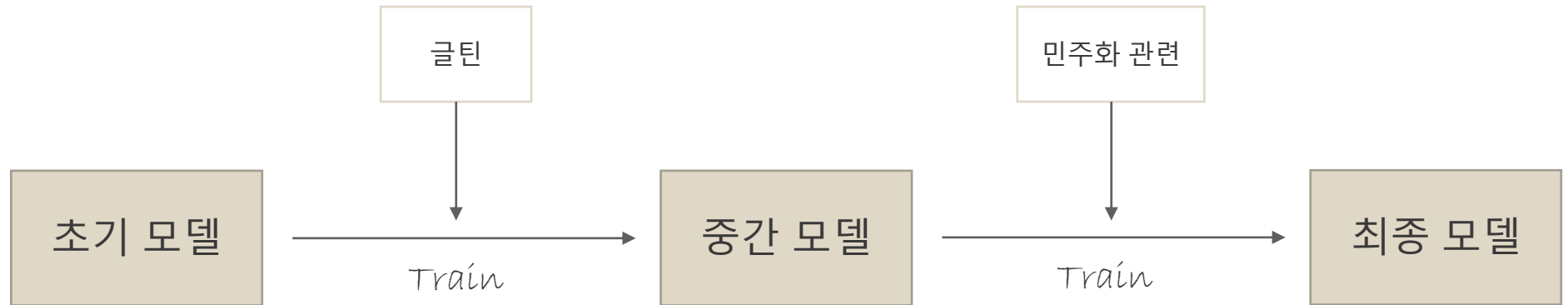
HellaSwag (F1)

Model	params	n=0	n=5	n=10	n=50
<a href="#">skt/ko-gpt-trinity-1.2B-v0.5</a>	1.2B	0.4036	0.4	0.4011	0.4214
<a href="#">kakaobrain/kogpt</a>	6.0B	0.4599	0.456	0.4616	0.4754
<a href="#">facebook/xglm-7.5B</a>	7.5B	0.4261	0.437	0.4409	0.4517
<a href="#">EleutherAI/polyglot-ko-1.3b</a> (this)	1.3B	<b>0.4013</b>	<b>0.3984</b>	<b>0.417</b>	<b>0.4416</b>
<a href="#">EleutherAI/polyglot-ko-3.8b</a>	3.8B	0.4438	0.4786	0.4737	0.4822
<a href="#">EleutherAI/polyglot-ko-5.8b</a>	5.8B	0.4853	0.482	0.4968	0.5012

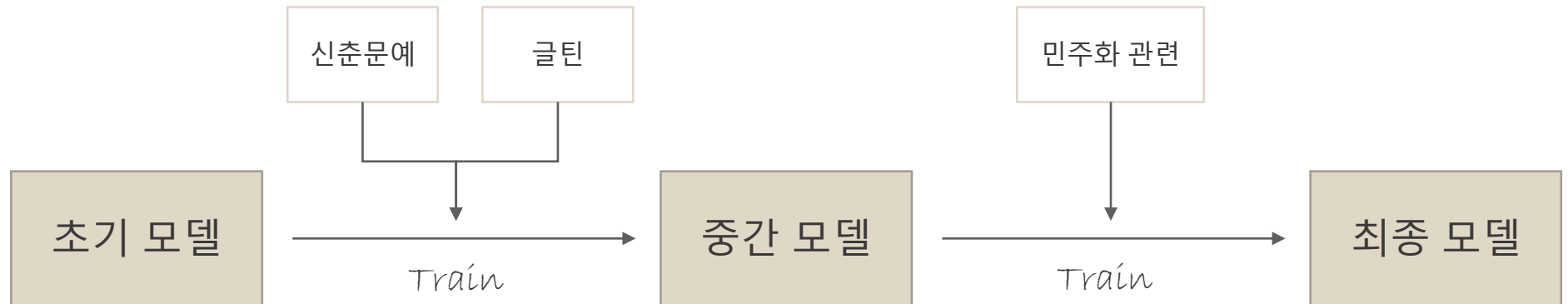


# 03 모델학습

Version 1 :

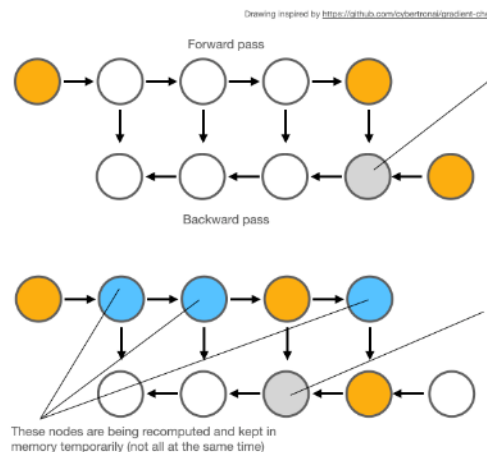


Version 2 :



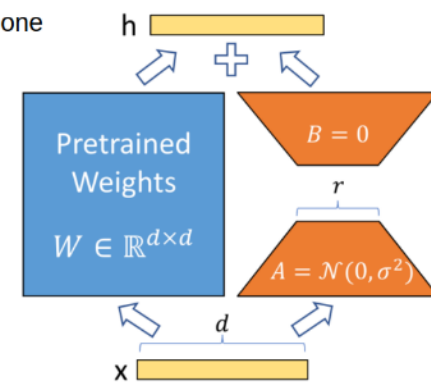
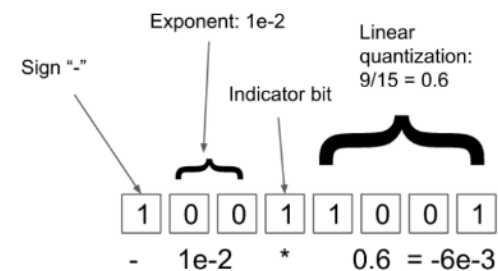
# 03 모델학습

큰 모델들은 GPU의 램에 모두 들어가지 못할 확률이 높아서 덩치를 줄일 기법을 여러가지 활용함

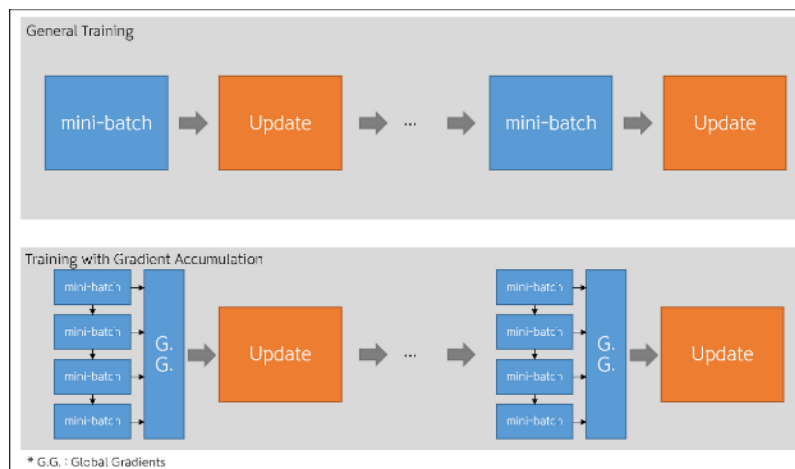


Gradient Checkpointing

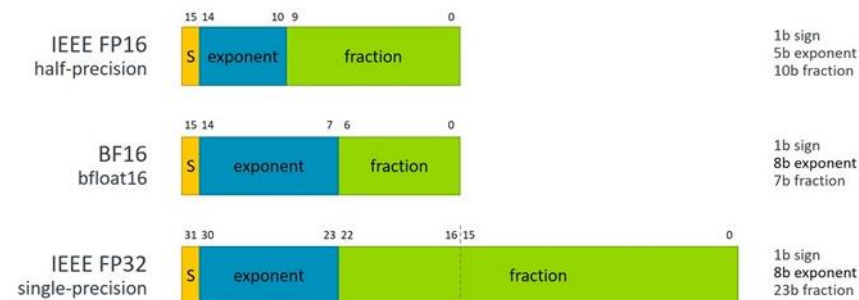
Right: low-rank adapters (LoRA) with 8-bit backbone  
Below: dynamic quantization by Dettmers et al.



8-bit optimizer



Gradient Accumulation



Mixed/Half precision

**04**

완성된 수필



## 04 완성된 수필

담대한  
膽大  
daring

담대한 포부

담대한 도전

담대한 꿈

담대한 대학생의 도전

민주화 운동



## 04 완성된 수필

담대한 단어

담대한 꿈의 실현

담대했던 나의 20대

담대하게만 보였던 아버지

담대한 목표

## 04 완성된 수필 - 프롬프트 작성

**“제목:** 민주화 운동 그

**카테고리:** 민주화라는 후

**필자:** 학생 운동 당시를 담대한 꿈을 꾸고 도전했던 대학생 시절을 회상하는 내용 담은 수필

**내용:** 아버지의 나이가 되어서 옛날의 엄격하던 아버지에 대해 다시 생각해본다 아버지에서 돌아가시기 전까지 내게

보여주신 삶이란 것이 과연 무엇인가?”

모델

## 04 완성된 수필

민주화라는 담대한 꿈을 꾸고 도전했던 대학생 시절을 회상한다.

엄격한 군인 아버지 밑에서 열심히 공부해 좋은 대학에 진학함.  
정치에는 관심이 없이

군부 독재로 인한 아버지의 실직과 친구의 죽음으로 민주화에 대한 결  
심을 함

어머니의 반대와 친구들이 다치고 힘들어함

결국 포기하지 않고 **6월 10일** 민주화 운동을 하고 **6월 29일** 민주화 선  
언을 봄. 지금의 청년들에게 담대한 도전에 대한 이야기를 함

# 감사합니다

