

2024.12.27.(금)

# KOIMANG NOVEL HUB

팀명: 코리아IT의 망나니가 되었다

발표자: 최윤범





# CONTENTS

---

프로젝트 소개 **01**

---

아키텍처 및 기술 스택 **02**

---

웹서비스 개요 **03**

---

데이터 분석 **04**

---

딥러닝 모델링 및 시연 **05**

---

Q&A 세션 및 마무리 **06**

---



# Project Introductions

프로젝트 소개



1. 기획 의도 및 배경

2. 팀원 소개





# 기획 의도 및 배경

개발 기간: 11/18 - 12/26



## 웹소설 시장의 주요 과제

- 웹소설 시장은 낮은 진입 장벽에 비해 터무니없는 성공률과 창작 과정에서의 피로도라는 도전 과제가 있음.
- 소설가 지망생들에게 아이디어를 제공하고, 창작 활동을 지원하고자 하는 웹서비스를 계획.
- 소설가들에게 데이터 기반의 트렌드 분석 및 생성 AI 도구를 이용한 효율적인 창작 환경 제공.

이코노미스트

웹소설, 진입장벽은 낮지만 성공은 어렵다!

낮은 진입 장벽에 지망생 크게 늘어...성공하기는 쉽지 않아

실제로 많은 지망생이 '돈이 된다'는 소문에 이끌려 웹소설 작가 데뷔에 도전하고 있습니다. 하지만 대다수 지망생은 첫 번째 단계인 '매일 5500자 작성'에서 탈락합니다. 창작이라는 것은 그만큼 고통스러운 일입니다. 특히 매일매일 5500자를 작성한다는 것은 보통의 인내력이 아니고서는 쉽지 않은 작업입니다.

또 매일 5500자를 쓴다고 해서 모든 웹소설이 투데이 베스트에 오를 자격이 주어지는 것도 아닙니다. 기본적으로 '재미'가 있어야겠죠. 아울러 최근 유행하는 트렌드에 대한 공부도 요구됩니다.

# 팀원 소개



코리아IT의 망나니가 되었다 팀의 구성원을 소개합니다

<div><div>팀장</div><div>문정현</div><div>데이터 분석 AI 모델링</div></div>	<div><div>팀원</div><div>최윤범</div><div>AI 모델링 모델 Backend</div></div>	<div><div>팀원</div><div>화철수</div><div>Frontend Backend 배포</div></div>
--	--	--

# Architecture & Tech Stack

아키텍처 및 기술 스택

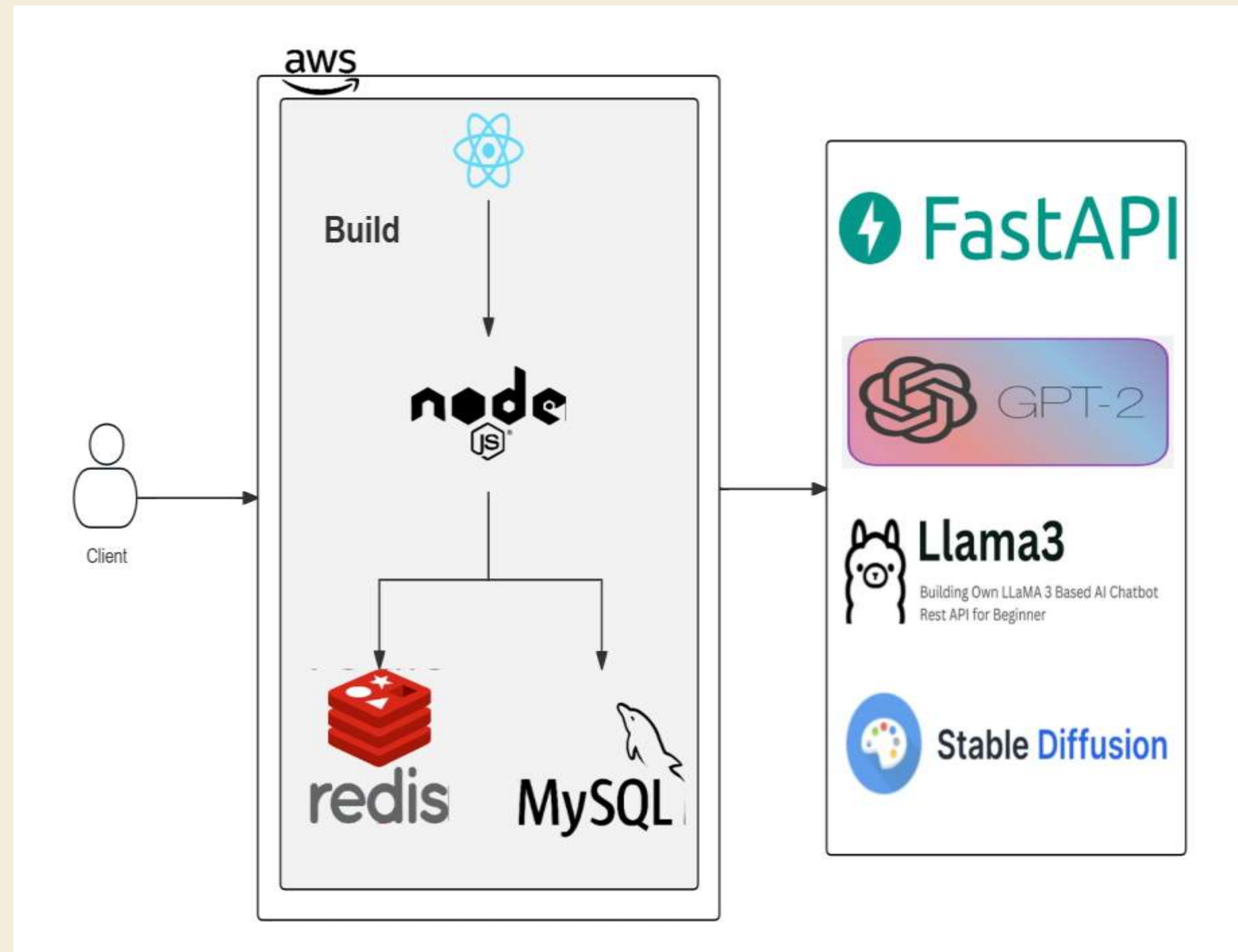


1. 아키텍처 다이어그램

2. 기술 스택 소개



# 아키텍처 다이어그램



# 주요 기술 스택



분류	기술 스택
크롤링	BeautifulSoup, Selenium
데이터 분석	Pandas, NumPy, SciPy, scikit-learn, Matplotlib, seaborn
딥러닝 및 모델링	PyTorch, TensorFlow, Hugging Face 라이브러리
서버 및 백엔드	Express, Passport, Sequelize, FastAPI
데이터베이스	MySQL
프론트 및 CSS	React, Tailwind
배포	AWS Lightsail, PM2



# Web Services Overview

웹서비스 개요



- 
1. 웹사이트 구현에 사용된  
주요 기술 요약
  2. 서비스 기능별 UI 설명
- 

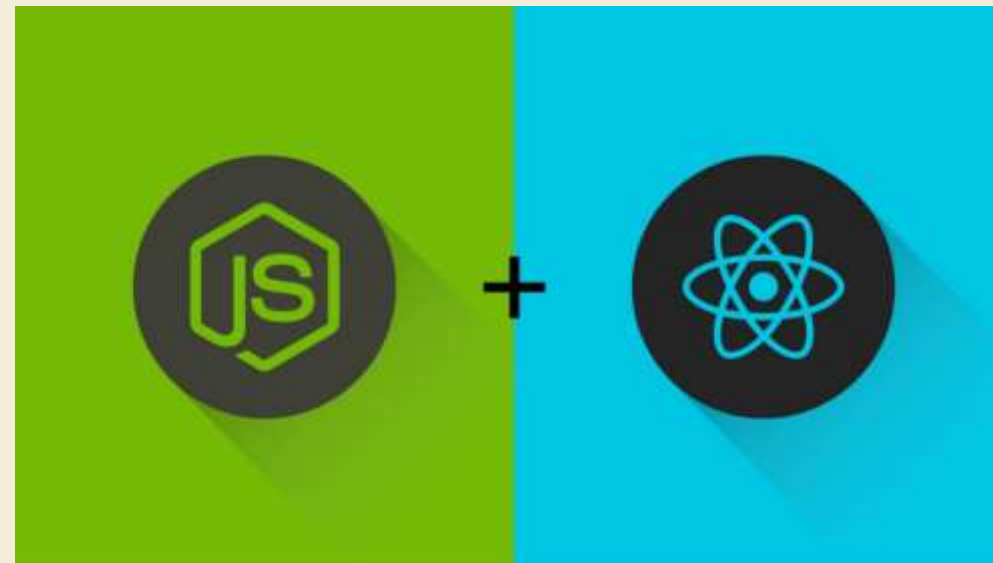


# 웹기술 정리



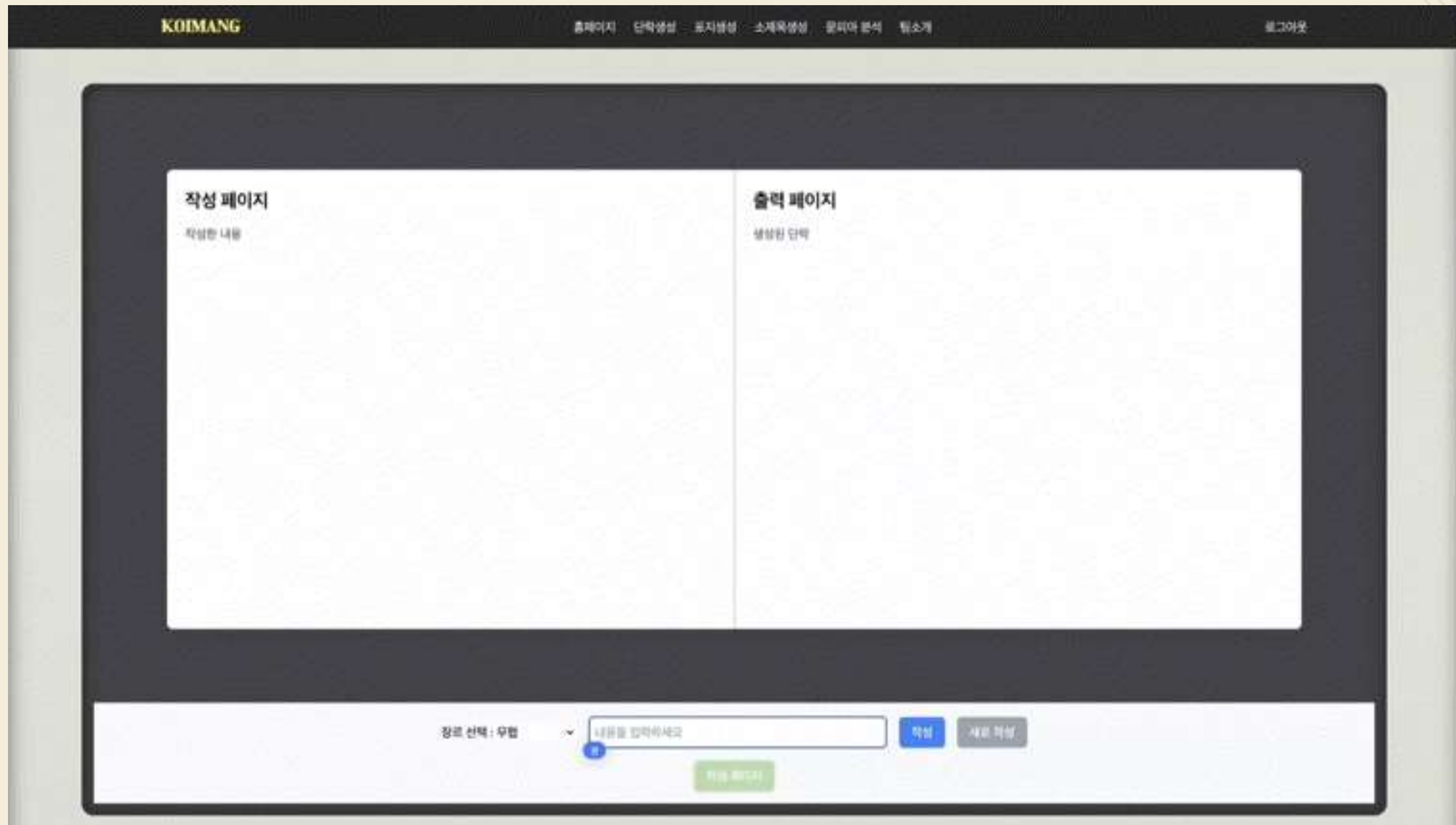
- Node.js 의 비동기 처리를 이용한 대규모 데이터 요청 처리 효율성 증대

- 서비스 확장성을 높이고, 라이브 협업과 추가 데이터 분석 기능 구현의 기반

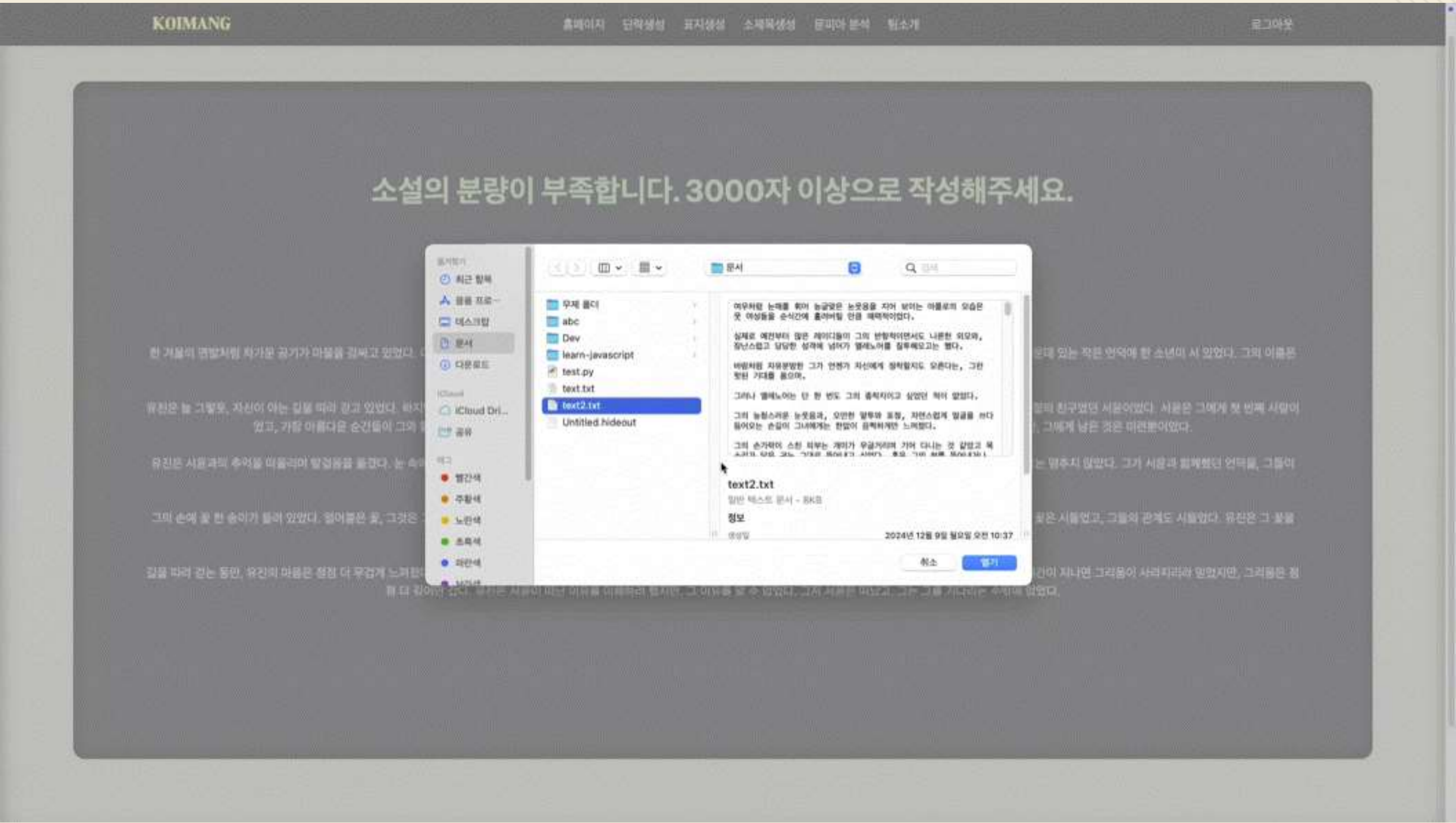


- 확장 가능하고 가용성이 높은 웹 호스팅 솔루션을 제공하여 기업이 웹 사이트를 쉽게 호스팅하고 관리할 수 있도록 지원

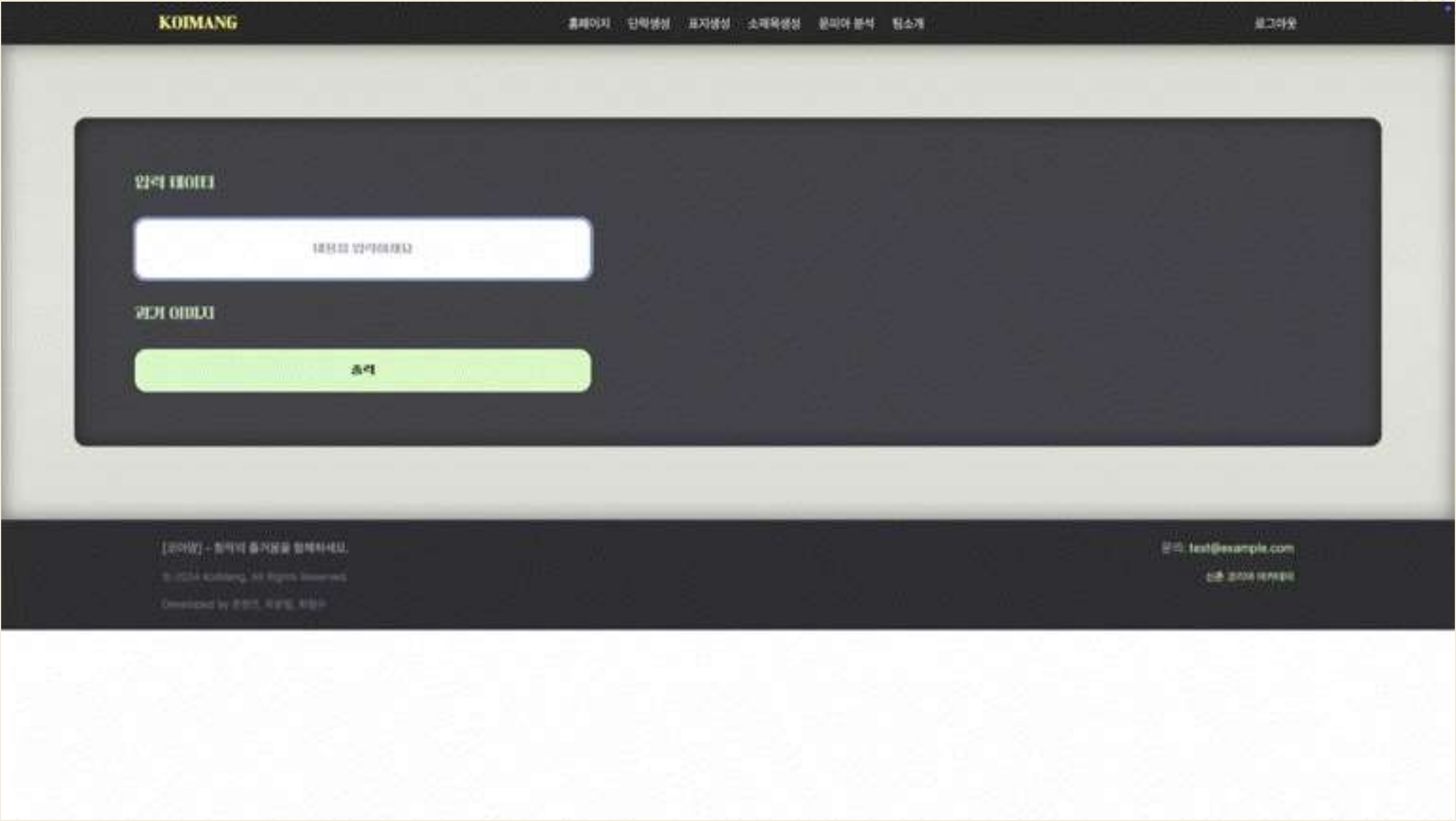
# 소설 생성 모델 UI



# 소제목 생성 모델 UI



# 표지 생성 모델 UI





# Data Analysis

데이터 분석



1. 분석 과정

2. 분석 결과

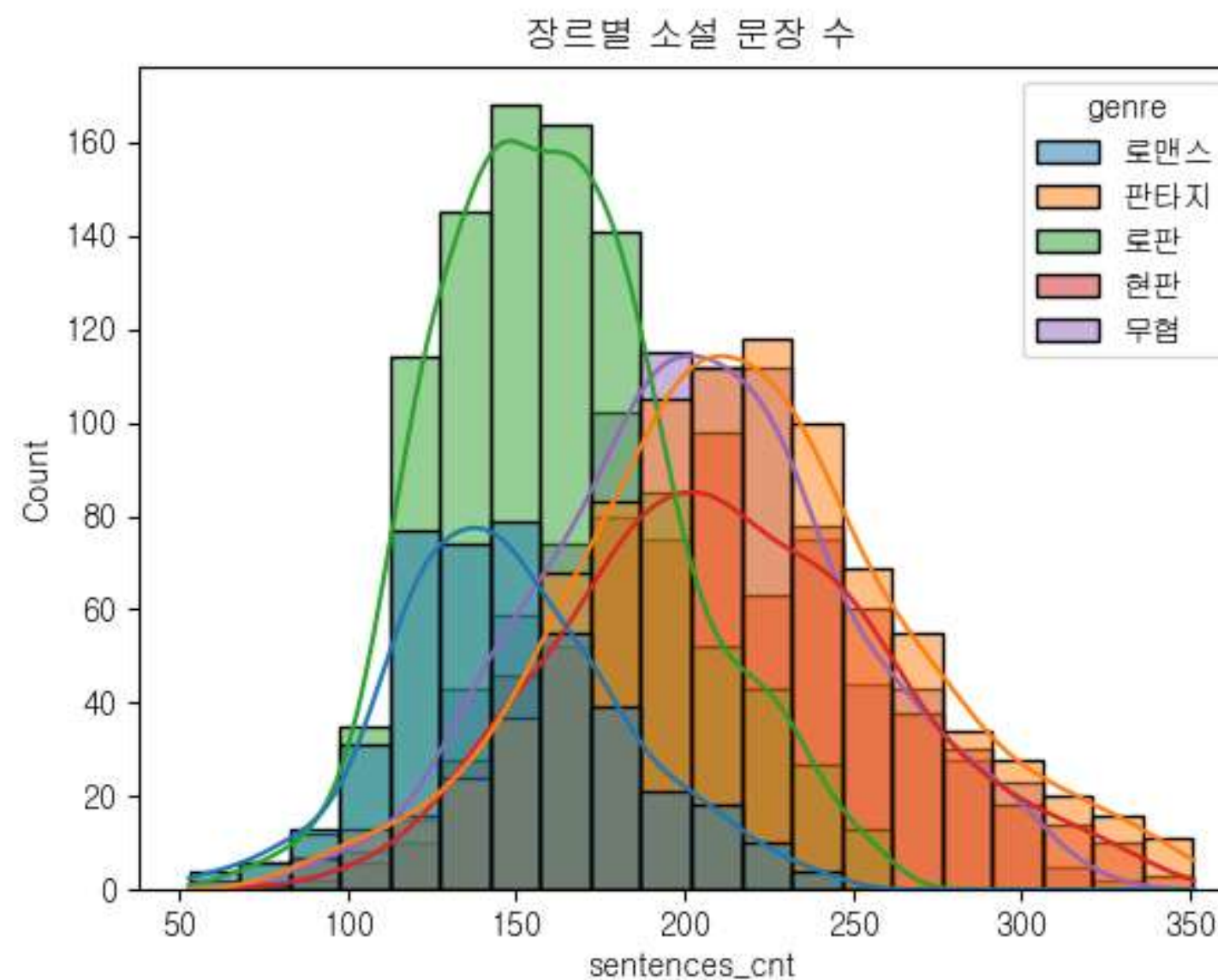


# 장르별 소설 분량 분석



## 비교분석 서비스 제공

- 텍스트 파일을 업로드하면 파일의 문장 수를 분석하고 장르별 인기 소설 문장 수와 비교
- 로맨스와 로판은 평균 150문장 내외
- 무협, 판타지, 현판은 평균 200문장 내외



# 장르별 소설 분량 분석

















# 소설 트렌드 업데이트

kakaopage 추천 웹툰 **웹소설** 책 바로가기

제목, 작가를 입력하세요. 🔍 📖 📁 👤

지금 핫한 실시간 랭킹 오늘신작 남성인기 **판타지** 현판 무협 드라마 로판 로맨스 단행본 요일 완전판

Top 300 24.12.19 11:00 기준 ⓘ

 1 ▲ 4 망검의 중심에서 듀얼을 외치다	 2 - 천재 궁수의 스트리밍	 3 ▲ 1 방치된 자동 사냥 게임의 흑막님	 4 ▼ 1 변경백 서자는 황제였다	 5 ▲ 1 마법학교 마법사로 살아가는 법	 6 ▼ 5 환생한 암살자는 검술 천재
 7 - 네크로맨서 학교의 소환천재	 8 ▲ 4 역대급 올마스터의 회귀	 9 ▲ 2 -99레벨 대마법사	 10 ▼ 1 신화로 초월하는 암살자	 11 ▼ 3 블랙기업조선	 12 ▲ 1 나 혼자 EX급 소환수

# 소셜 트렌드 업데이트



사람  
생각  
얼굴  
남자  
자신  
고개  
마음  
소리  
여자  
목소리  
자리  
표정  
시선  
오늘  
모습  
정도  
순간

고렵	NNG
대가	NNG
뒷담화	NNG
마족	NNG
만렵	NNG
백마장군	NNG
비무	NNG
스킬	NNG
플러팅	NNG
어쌔신	NNG
영지전	NNP
오우거	NNG
용병단	NNG
저렵	NNG

불용어 처리를 위한 STOPWORDS

사용자 정의 사전

## 불용어 처리

- 분석에서 의미가 없는 단어나 유용하지 않은 단어를 제외하는 과정

## 사용자 정의 사전

- 형태소 분석기가 사전에 정의된 단어를 인식할 수 있도록 단어와 품사를 작성함

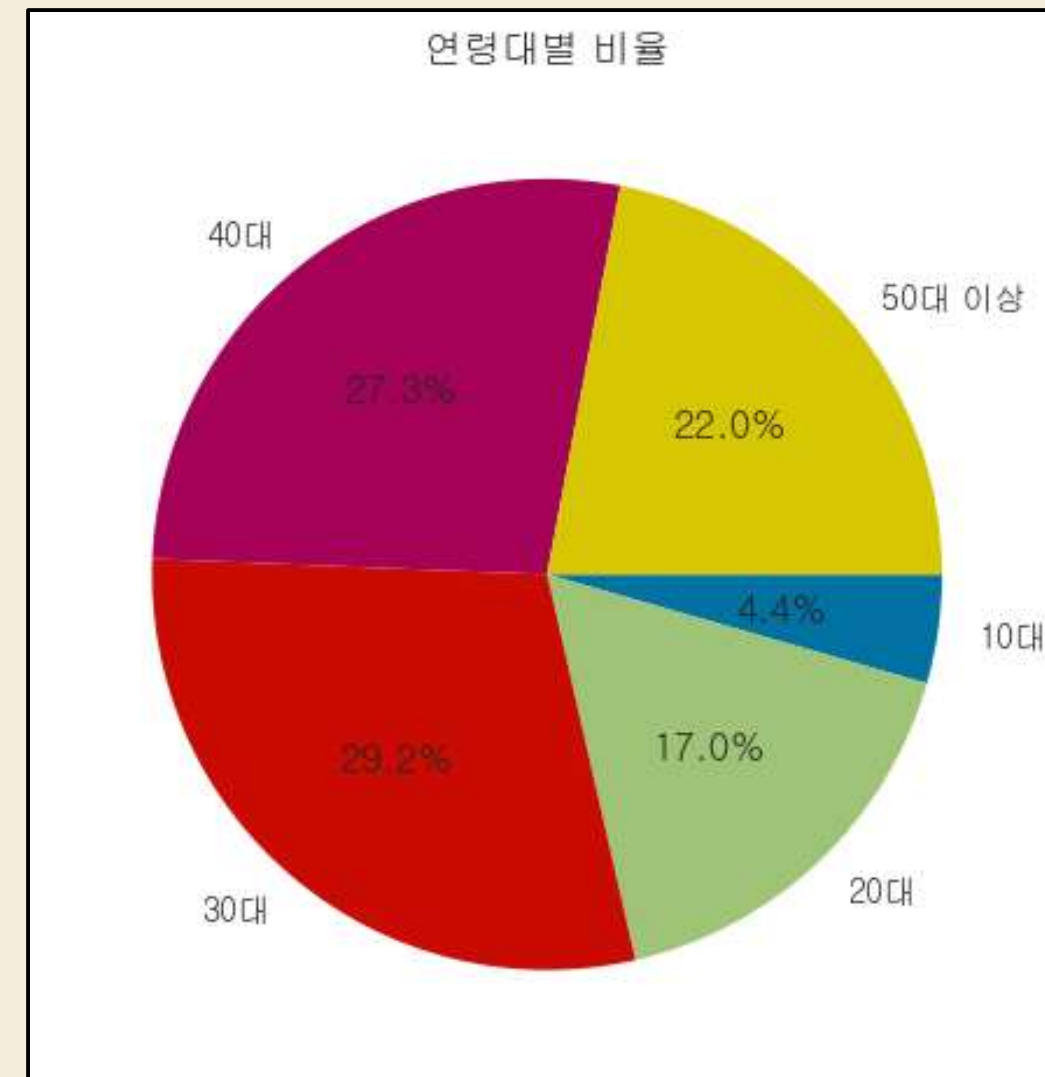
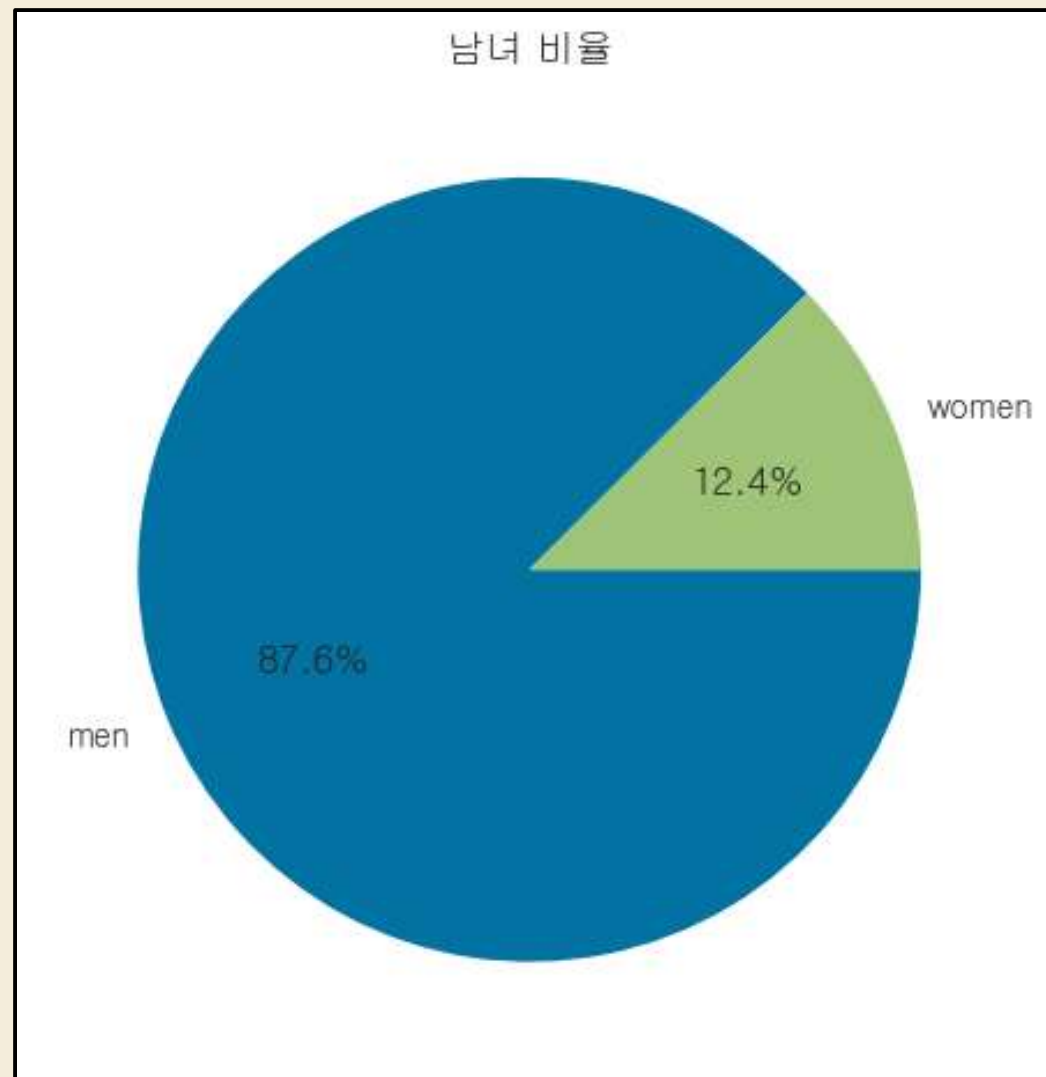


# 군집 분석

## 사용 데이터셋: 문피아 데이터

	title	genre	etc	tag	cycle	start	end	count	view	recommend	subscribe	string	men	women	10 대	20 대	30 대	40 대	50 대 이 상	유료
0	무공이 너무 쉽다	heroism	['새글']	['#현대판타지', '#퓨전', '#무협', '#무인', '#헌터', '#능력자']	['sun', 'mon', 'tue', 'wed', 'thu', 'fri', 'sat']	2024.11.10 14:09	2024.12.05 23:00	26 회	730,734	16,899	13,257	119,277	536,048	34,600	3	6	23	33	33	False
1	소요수선전(逍遙修仙傳)	heroism	['새글']	['#판타지', '#퓨전', '#무협', '#선협', '#환생', '#성장']	['mon', 'tue', 'wed', 'thu', 'fri']	2024.10.04 07:54	2024.12.05 21:51	45 회	800,150	24,034	10,308	297,637	590,516	48,126	3	9	22	28	35	False

# 군집 분석

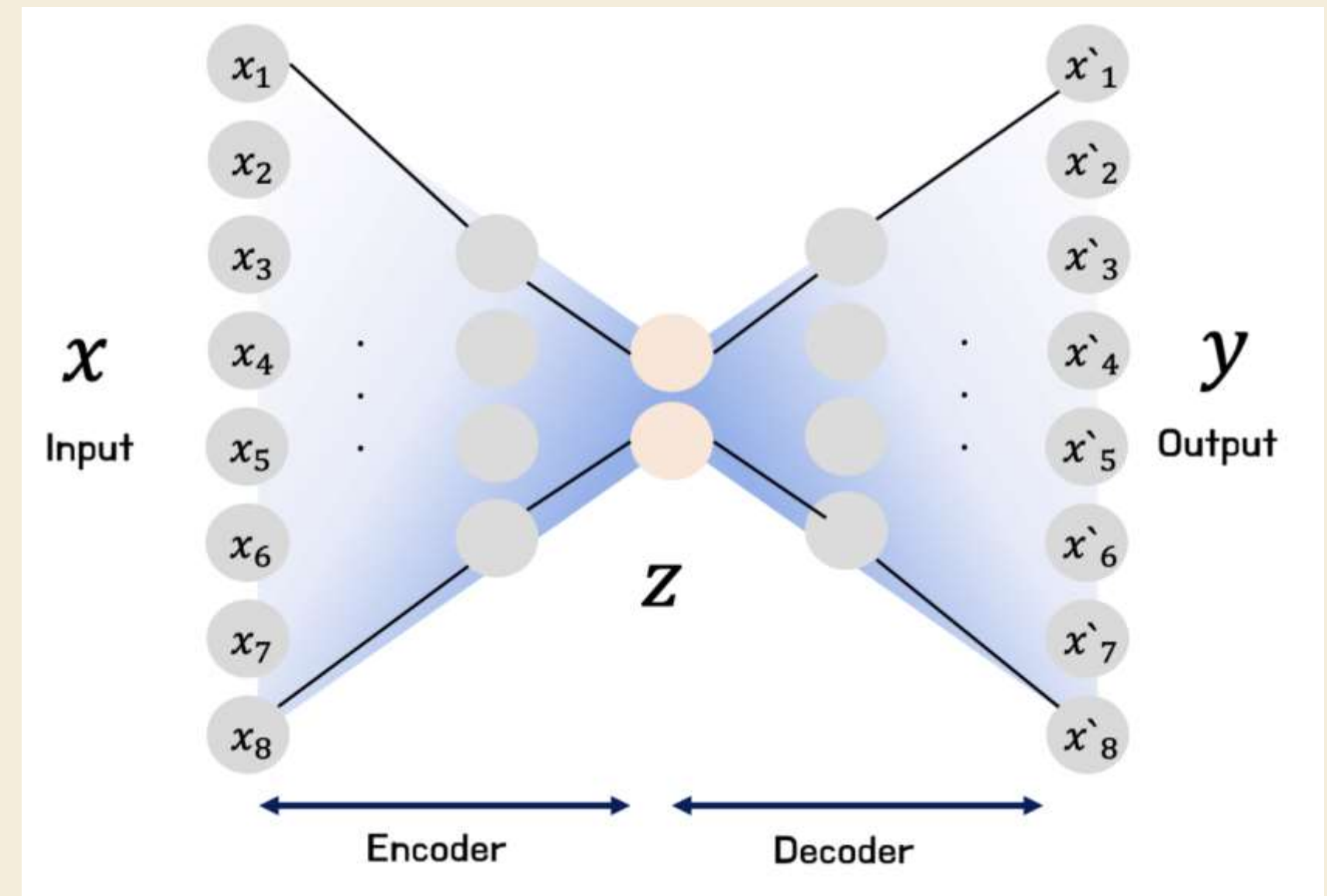


주 독자층: 남성 및 20대 이상 성인

# 군집 분석



- 딥러닝 오토인코더 기반의 군집분석을 진행
- 입력이 들어왔을 때 해당 입력 데이터를 최대한 압축시킨 후, 데이터의 특징을 추출하여 다시 본래의 입력 형태로 복원시키는 신경망



# 군집 분석



Layer (type)	Output Shape	Param #
input_layer_6 (InputLayer)	(None, 273)	0
dense_36 (Dense)	(None, 128)	35,072
dense_37 (Dense)	(None, 64)	8,256
dense_38 (Dense)	(None, 32)	2,080
dense_39 (Dense)	(None, 64)	2,112
dense_40 (Dense)	(None, 128)	8,320
dense_41 (Dense)	(None, 273)	35,217

Total params: 91,057 (355.69 KB)

Trainable params: 91,057 (355.69 KB)

Non-trainable params: 0 (0.00 B)

```
# 잠재 공간 모델
encoder = keras.Model(inputs = inputs, outputs = latent_Z)

# 잠재 특징 추출
latent_features = encoder.predict(input)
```

- 임베딩 레이어의 latent feature를 추출하여 군집분석에 사용

# 군집 분석

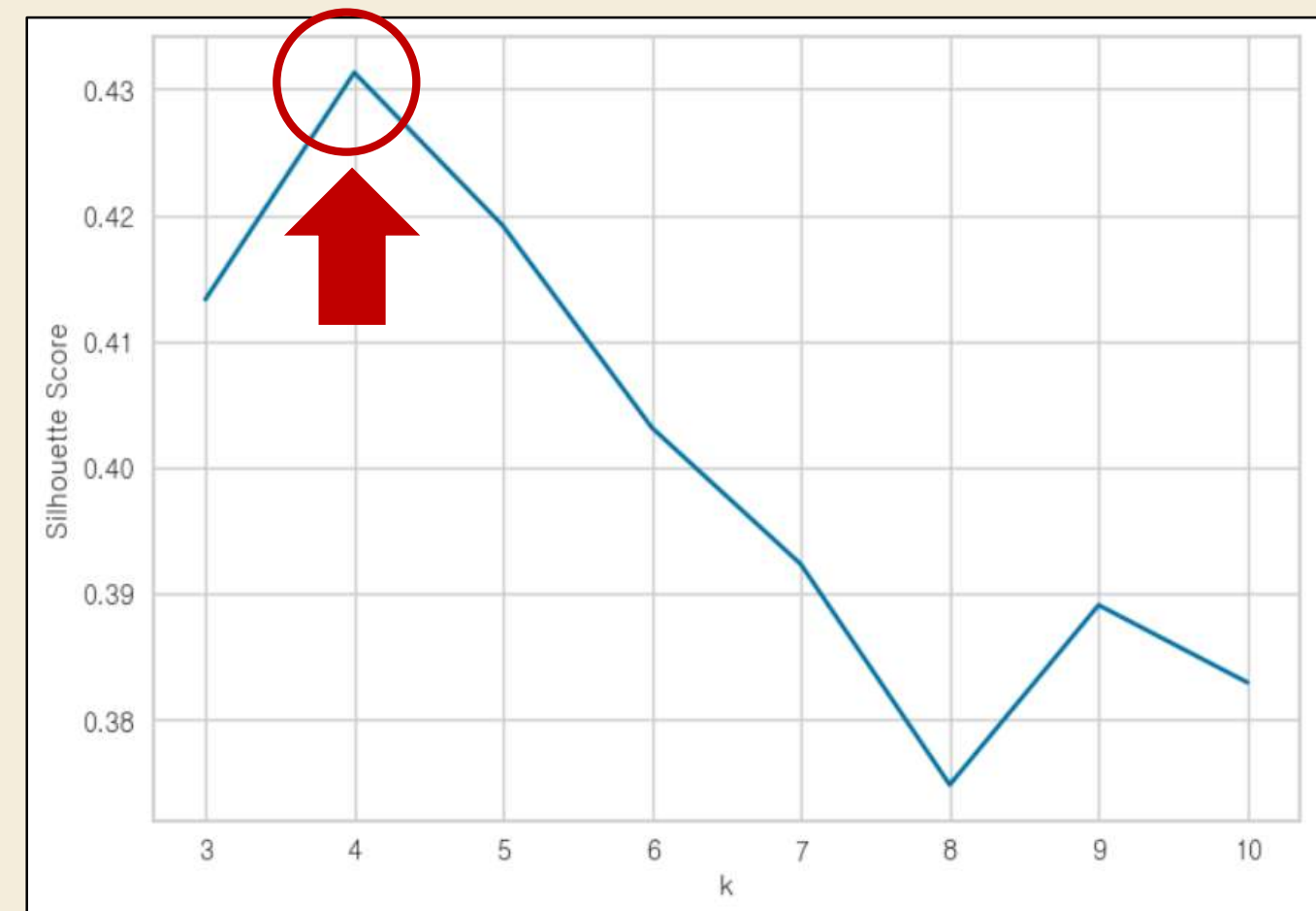


## 이너셔



- 각 데이터 포인트가 군집 중심에서 얼마나 가까운지 나타내는 값

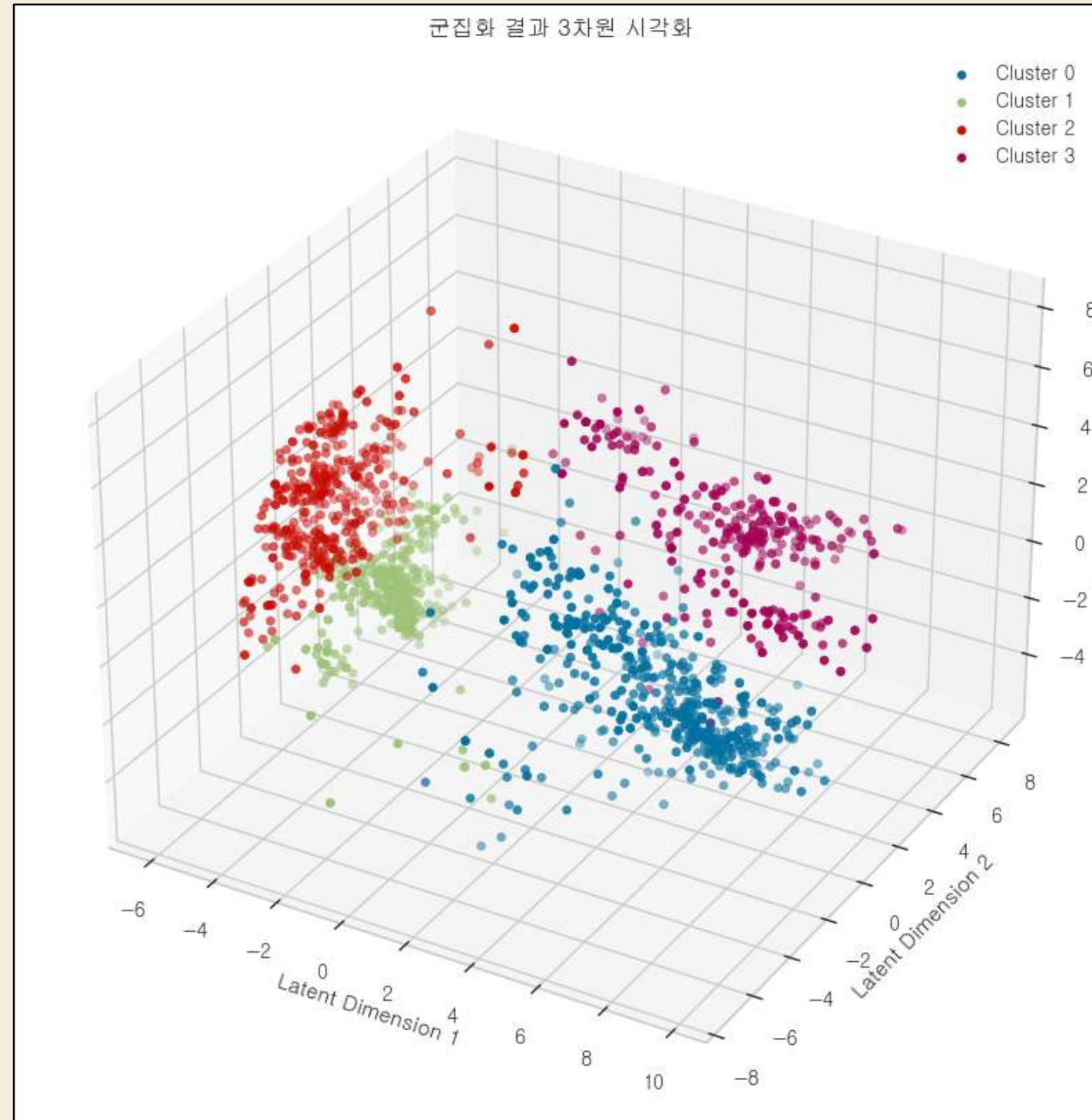
## 실루엣 계수



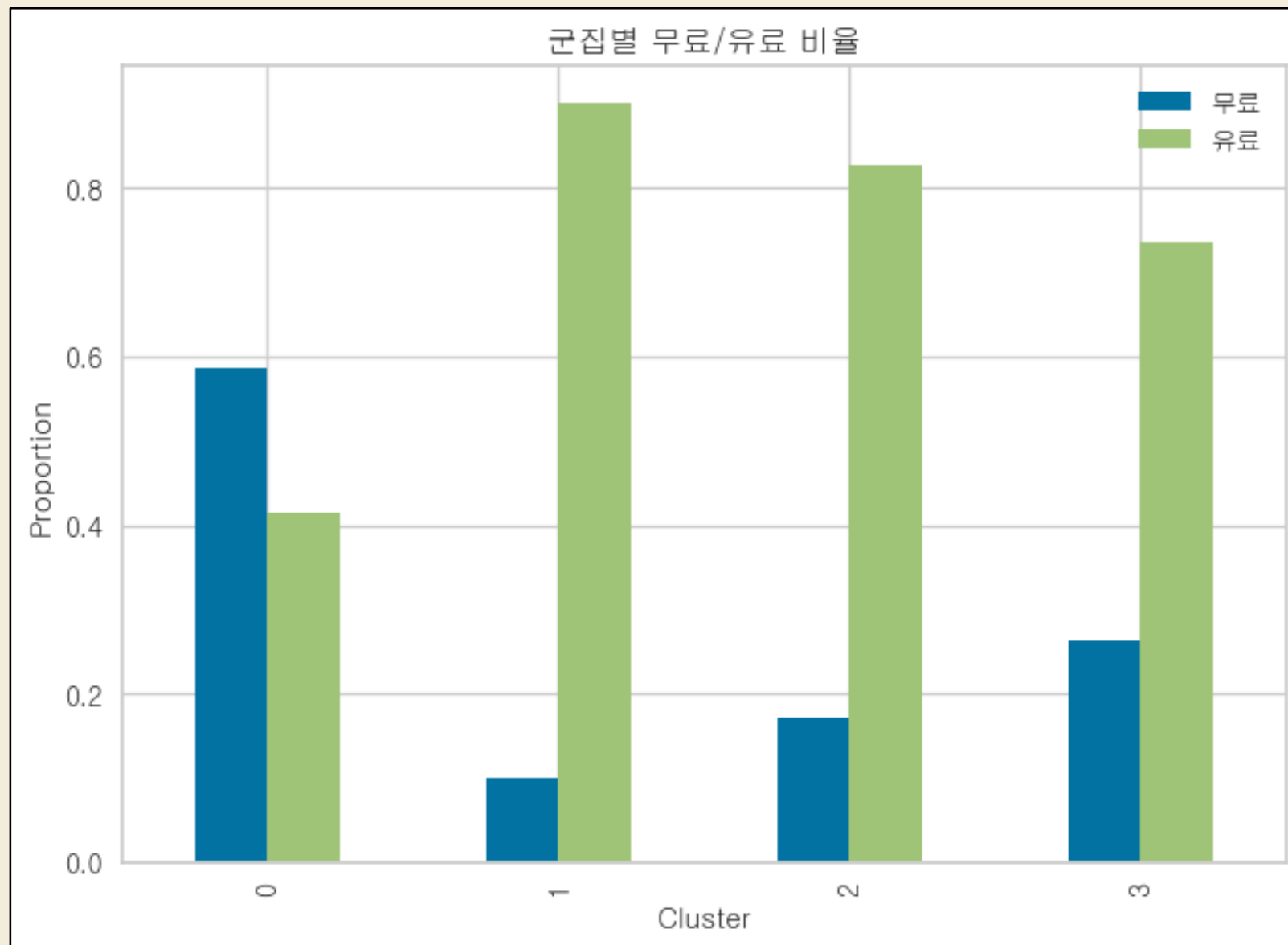
- 데이터가 군집 내부에서 얼마나 잘 뭉쳐있고, 다른 군집과 얼마나 분리되어 있는지 측정하는 지표



# 군집 분석



# 군집 분석



## Cluster 0 보유 태그

#가족, #검은머리, #동양판타지, #서양판타지, #성장, #아포칼립스, #연예계, #이능력, #이세계, #착각, #담가족, #검은머리, #동양판타지, #서양판타지, #성장, #아포칼립스, #연예계, #이능력, #이세계, #착각, #탑등반물 등

- 무료 소설 대다수가 포함된 cluster 0은 남성향 작품 외의 다양한 장르의 소설이 포함됨.

# 군집 분석

## 군집 분석 결과



	고유 태그	인기 조합	특징 요약
Cluster 0	#먼치킨, #사이다, #힐링, #각성, #빙의	(#먼치킨, #빙의), (#사이다, #힐링), (#먼치킨, #사이다, #힐링)	무료 작품 대다수가 포함되어 다양한 장르가 포함됨.
Cluster 1	#전쟁, #정치, #군주, #세계사, #조선	(#전쟁, #빙의), (#정치, #세계사), (#전쟁, #세계사)	역사 중심의 스토리.
Cluster 2	#운동선수, #노력가, #야구, #축구	(#먼치킨, #빙의), (#노력가, #사이다, #힐링)	스포츠와 성장 중심의 스토리.
Cluster 3	#생존, #중세, #조선, #한국사	(#전쟁, #빙의, #경영), (#천재, #힐링), (#회귀, #경영)	역사적 배경의 판타지와 생존 스토리.

# Modeling & Demonstration

딥러닝 모델링 및 시연



1. 서비스의 필요성
2. 사용 모델 및 선택 이유
3. 파인튜닝 과정
4. 서비스 기능 시연



# 단락 생성 서비스의 필요성



\*에디션지수가 노출되지 않는 경우는 다음과 같습니다.

- 총 회차수가 5화 이하인 경우
- 12주 이내 연재된 회차가 없는 경우
- 1화 연재 후 18주가 초과한 경우
- **작품의 평균 글자수가 4,500자에 미달되는 경우**
- 작품의 장르를 변경하는 경우
- 웹소설 게시글 원칙에 어긋난 작품

## ▶ 일반연재

자유연재에서 한 단계 올라간 연재 공간입니다.  
일정 기준 이상이 충족되어야 승급할 수 있습니다.  
일반 연재 승급을 위해서는 아래 세 조건을 모두 충족하여야 합니다.

1. 자유연재란에서 연재한 한 작품의 총 분량 75,000자 이상 (주장르가 로맨스인 경우 50,000자 이상)
2. 프롤로그를 제외한 회차당 최소 글자수 3,000자 이상
3. 연재 당일을 제외한 최소 7일 이상 연재

\* 이미 출간 계약이 된 글에 대해서는 계약 업체명 기재 및 원고파일을 첨부하여 주시면 승급이 가능합니다. 이 경우 업체를 통해 사실 확인 후 승급을 도와드립니다.

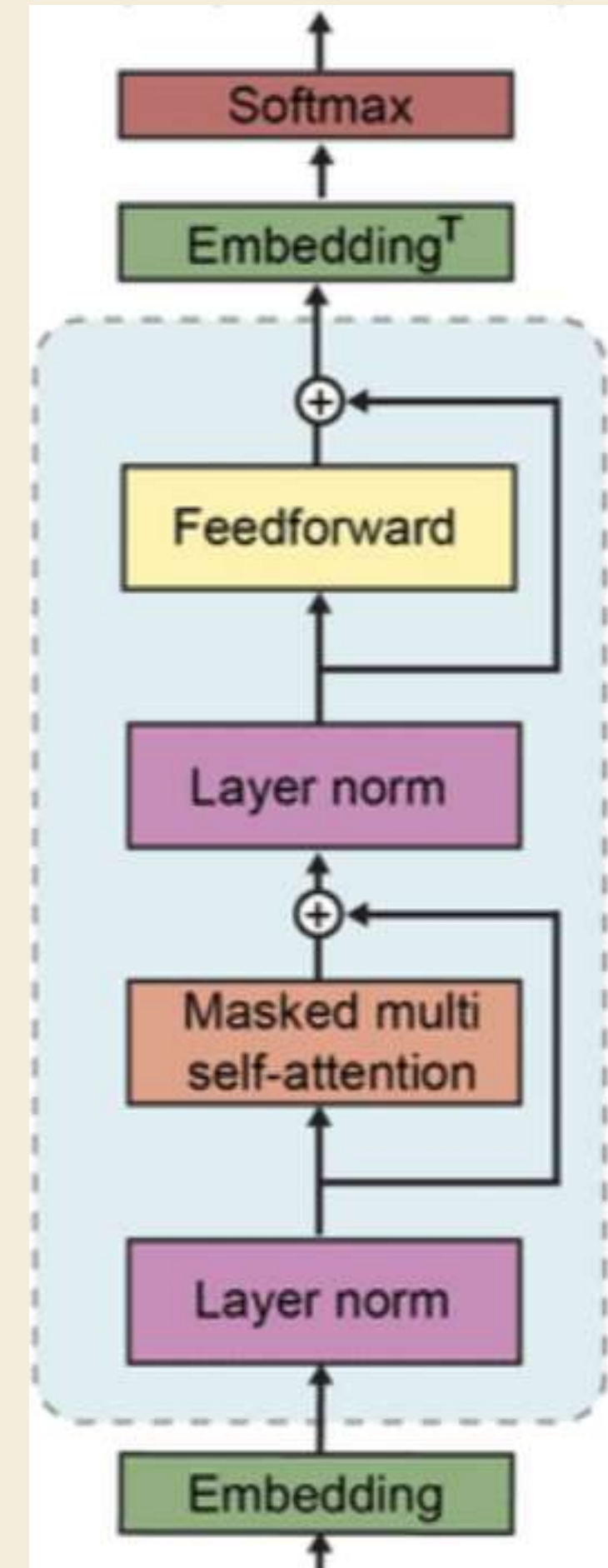
\* 신청하는 작가명으로 장르소설을 출판한 경력(전자책 포함)이 있다면 해당 작품을 완결하지 않더라도 신청이 가능합니다. 처음부터 일반연재로 시작할 수 있으며, 승급 이후 프로필에 영구적으로 일반연재가 유지됩니다. 출간작의 기준은 유료판매 여부이며, 장르소설이 아니거나 개인출판(자비출판)일 경우 인정이 되지 않습니다.

- 에디션 지수란?
- 베스트 리그에서 현재 연재되고 있는 작품들에 대한 상대적인 지표



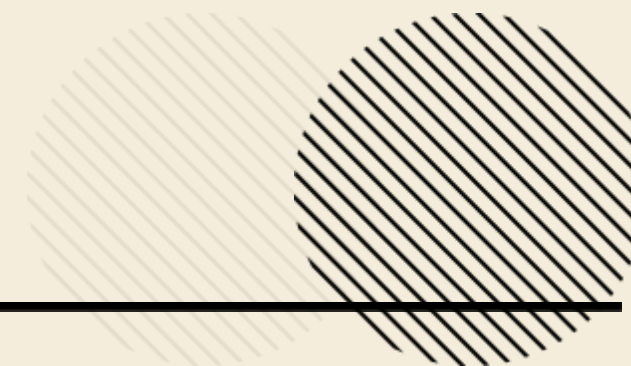
# 단락 생성(KoGPT-2) 모델

- Embedding: 입력 텍스트를 벡터화
- Masked multi self-attention: 각 단어가 문맥 내에서 다른 단어와 어떤 관계를 가지는지 학습
- Layer norm: 각 레이어의 출력을 정규화
- Feedforward: 학습된 문맥 정보를 기반으로 단어 예측
- Softmax: 단어에 대한 확률 분포 계산 후 최적의 단어 선택



GPT-2 decoder block

# 단락 생성(KoGPT-2) 모델

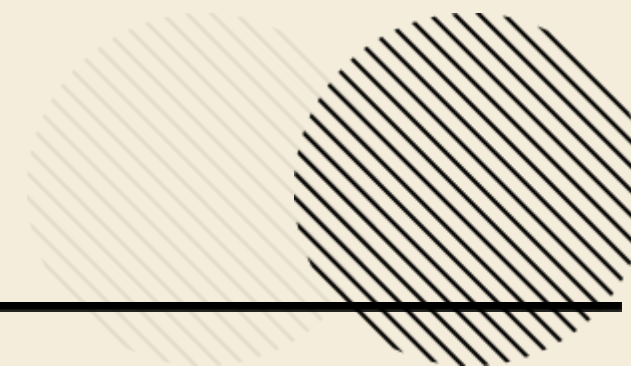


사용 데이터셋: 카카오페이지 + 네이버 시리즈 소설 데이터

```
df = pd.read_csv("./data/gpt_fine4_750.csv")
df.head()
```

		cleaned_content	genre
0	화 챔피언 경기 시작.시합을 알리는 레퍼리의 선언이 시작되자, 상대 선수가 슬금슬금...		fantasy
1	와아아아아아!기대를 저버리지 않는 챔피언의 모습에 관중석에서 커다란 환호성이 터져 ...		fantasy
2	대부분 자신들의 이익에 대해서만 늘어놓는, 같잖은 내용들이다!기생충 같은 것들!...		fantasy
3	피로로 인해 몸이 무거운 것이 아니었다.피곤함 때문에 느껴지는 피로와는 전혀 달랐다...		fantasy
4	한데 이상했다.분명 처음 보는 얼굴이었지만, 어딘가가 낯이 익었다.그 순간 갑자기 ...		fantasy

# 단락 생성(KoGPT-2) 모델



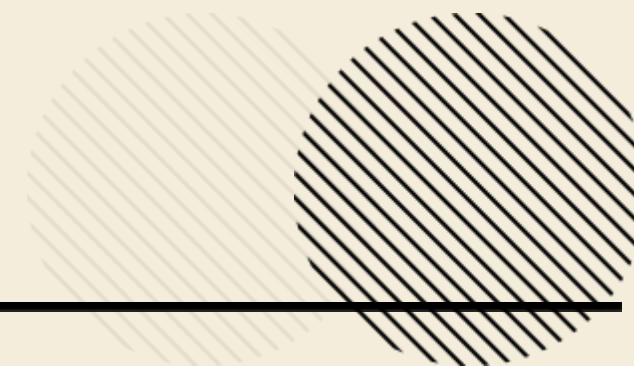
```
import torch
from transformers import GPT2LMHeadModel

model = GPT2LMHeadModel.from_pretrained('skt/kogpt2-base-v2')
text = '여주가 소설 속 세계에 들어온 후 이후 장면을 그려줘.'
input_ids = tokenizer.encode(text, return_tensors='pt')
gen_ids = model.generate(input_ids,
                        max_length=128,
                        repetition_penalty=2.0,
                        pad_token_id=tokenizer.pad_token_id,
                        eos_token_id=tokenizer.eos_token_id,
                        bos_token_id=tokenizer.bos_token_id,
                        use_cache=True)
generated = tokenizer.decode(gen_ids[0])
```

generated

'여주가 소설 속 세계에 들어온 후 이후 장면을 그려줘.</d> #20180713 **오늘의 아침<unk>\n오늘은 어제보다 더 춥습니다.\n아침 최저기온 -1도, 낮 최고 기온 3도**로 오늘도 평년과 비슷하겠고요.\n낮에도 기온이 크게 오르지 못해서 종일 쌀쌀한 날씨가 이어집니다.\n다만 내일은 다시 맑은 하늘이 드러나겠는데요,\n내륙을 중심으로 건조특보가 확대되니까 화재 예방에 유의하셔야 합니다.\n현재 전국이 대체로 맑지만 강원 영동과 경북 동해안에는 대기가 매우 건조한 상태입니다.\n산불 등 각종 화재가 발생하지 않도록 각별히 주의하시기 바랍니다.\n'

# 단락 생성(KoGPT-2) 모델



```
# 사용자 요청을 포함한 프롬프트
```

```
genre = "romance_fantasy"
```

```
user_request = ""
```

```
여주가 소설 속 세계에 들어온 후 이후 장면을 그려줘.
```

```
""
```

```
# 텍스트 생성
```

```
generated_text = text_generator.generate(genre = genre, user_request = user_request, max_length = 512, temperature = 0.8, top_k = 35, top_p = 0.85, repetition_penalty = 1.5)
```

```
print(generated_text)
```

그런데 어떻게 그게 현실일 수가 있단 말인지. 어쩌면 이게 현실이 맞을지도 모른다는 생각이 들었다. 아니, 이건 분명 현실이 아닐 거다. 원작에선 이런 일이 벌어지지 않았으니까! 하지만 지금 눈앞에 있는 상황에서는 그런 말도 안 되는 일들이 일어나지는 않을 것이다. ‘대체 왜 내가 이렇게 된 거지?’ 나는 혼란스러움에 침대에서 벌떡 일어났다. 눈을 크게 떴다가, 갑자기 허공에 나타난 반투명한 창 너머로 보이는 낯익은 광경에 멍하니 중얼거리다 다시 정신을 차렸다. 그러자 익숙한 풍경이 눈에 들어왔고..... 우르릉, 하는 굉음과 함께 무언가 부서져 내리는 듯한 천둥소리가 귀를 때려댔고, 곧이어 눈앞에서 커다란 물체가 날아와 나를 덮쳤기 때문이다. 좌익 하고 떨어진 물체에 부딪힌 충격은 생각보다 컸지만, 나는 반사적으로 몸을 비틀며 중심을 잡았다. 그리고 천천히 자리에서 일어섰는데..... 뭘가가 내 등을 강하게 찔렀던 모양이다. 깜짝 놀라 고개를 돌리니 새카만 로브 쓴 남자가 서 있었다. 검은 머리카락에, 그리고 하얀 피부색까지 모두 똑같이 생겼지만 분명히 사람이었다. 분명히 남자였는데도 너무 잘생겼다고 해야 할까? 내 눈이 점점 커졌다. 눈이 마주치자마자 남자의 입술이 열리며 웃음이 터질 것 같았지만, 애써 참으며 입을 다물려고 했다. 지금 중요한 건 이곳이 어디인지가 아니니까. 잠시 후, 남자는 한 손으로 얼굴을 가리고 있었는데, 마치 제 손을 꼭 붙잡는 것처럼 아주 자연스럽게 느껴지기까지 하였다.’ 라고 쓰여 있어 조금 이상했지만, 어쨌든 남자를 향해 미소 짓고 있어서 얼굴이 보이지 않았다. 그러나 어쨌든, 그는 여전히 내게서 시선을 떼지 않고 있었고, 그렇게 한참 동안 아무 일도 없었다는 듯 계속해서 말을 이어갔는데, “그래서, 이제 좀 살 만해졌네?” 남자의 말이 끝나기가 무섭게, 나도 모르게 입이 저절로 움직였다. 내가 아무리 노력해도 아무런 느낌도, 어떤 의미감도 느끼지 못한 채 그저 ‘살았다’는 말의 의미를 제대로 이해하고 있던 탓이었나 보네. 난 그만 헛웃음을 지르고야 말았고, 뒤늦게 상황을 파악한 순간 눈물이 날 뻔했지만 이미 늦어버린 뒤였기에 당황할 법 없었다.

# 소제목 생성 서비스의 필요성



UP 11화 ★ 0.0 댓글 0 조회 1 15:00	
10화 ★ 10.0 댓글 0 조회 5 2024.12.21	
9화 ★ 10.0 댓글 1 조회 1 2024.12.21	태그 #미스터리
8화 ★ 10.0 댓글 0 조회 7 2024.12.17	작품 회차 (7) 최신순 1회부터
7화 ★ 10.0 댓글 0 조회 5 2024.12.16	7화 ★ 0.0 댓글 0 조회 4 2024.12.24
6화 ★ 10.0 댓글 2 조회 14 2024.12.13	5화 ★ 0.0 댓글 0 조회 5 2024.12.20
	4화 ★ 0.0 댓글 0 조회 3 2024.12.19

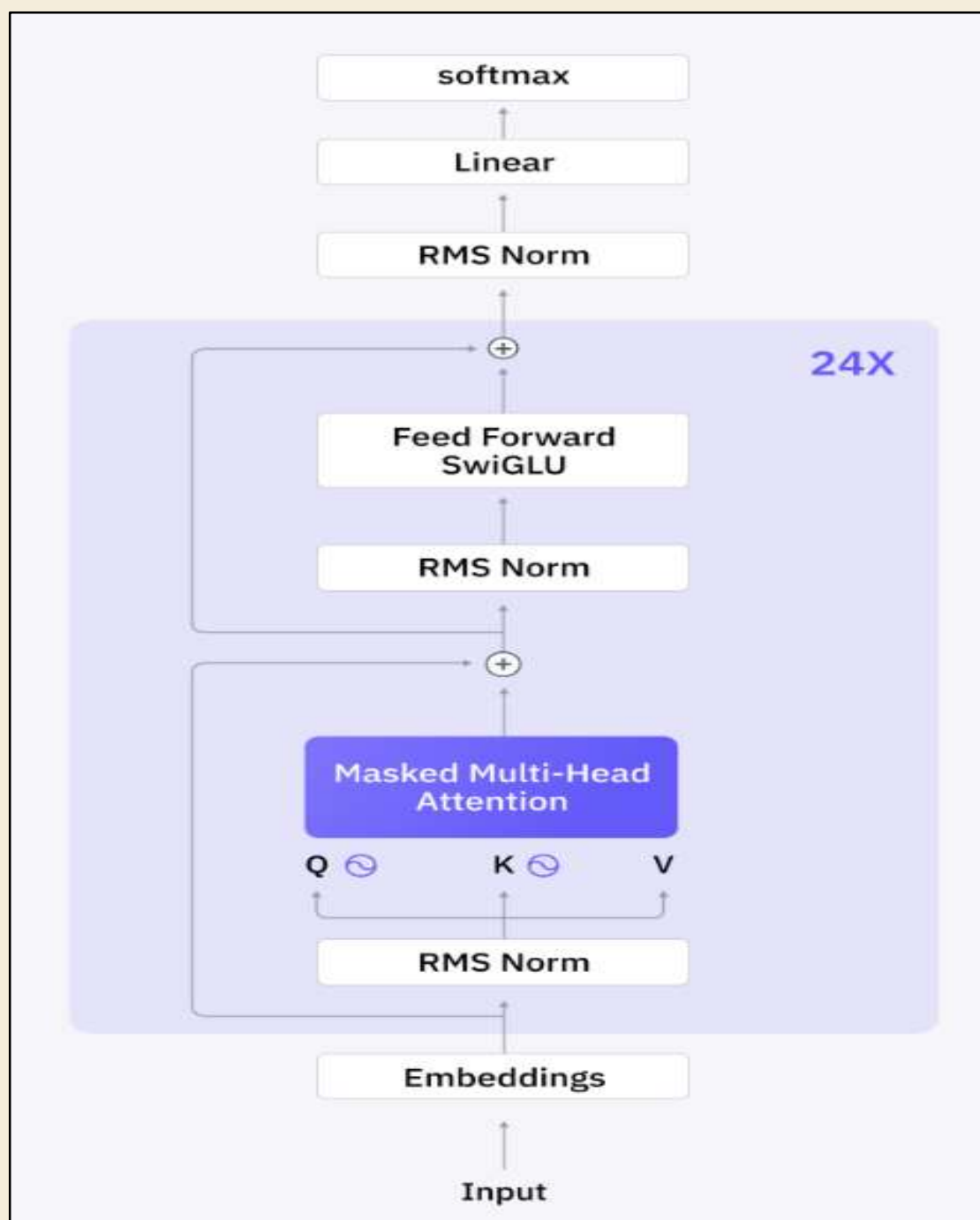
“

소설 제목은 신문 기사 제목과 같은 틀 짓기(framing)가 강하게 나타나지 않는다. 소설 텍스트의 내용은 사실(fact)이 아니므로 사회적 맥락이나 이데올로기가 중요한 역할을 하지 않는다. 그 대신 작가의 의도를 부각시키고 독자의 관심을 집중시키는 데 그 목적이 있다.

”

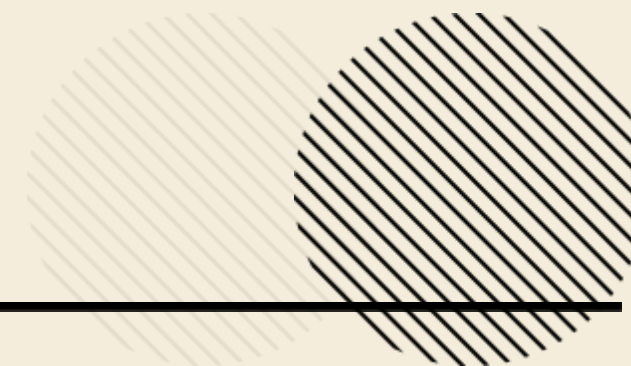


# 소제목 생성(LLaMA3) 모델

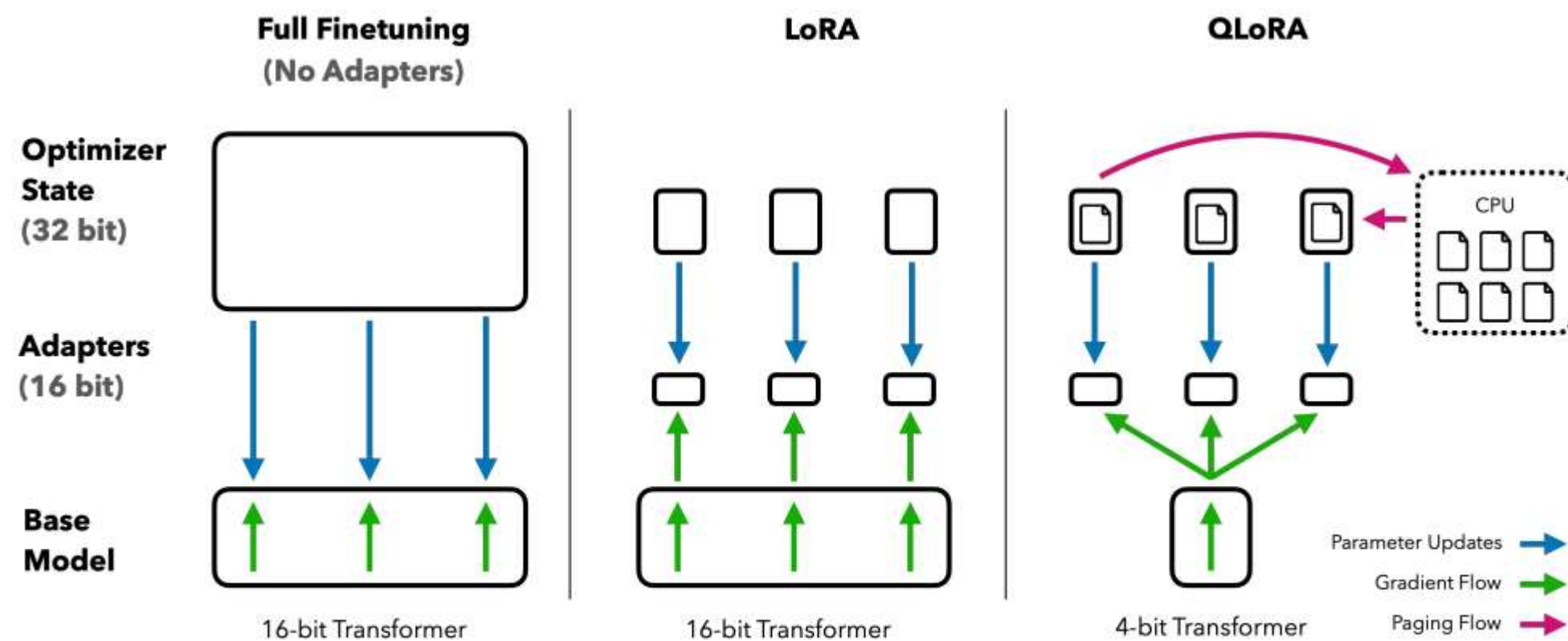


```
LlamaForCausalLM(  
  (model): LlamaModel(  
    (embed_tokens): Embedding(128256, 3072)  
    (layers): ModuleList(  
      (0-27): 28 x LlamaDecoderLayer(  
        (self_attn): LlamaSdpaAttention(  
          (q_proj): Linear(in_features=3072, out_features=3072, bias=False)  
          (k_proj): Linear(in_features=3072, out_features=1024, bias=False)  
          (v_proj): Linear(in_features=3072, out_features=1024, bias=False)  
          (o_proj): Linear(in_features=3072, out_features=3072, bias=False)  
          (rotary_emb): LlamaRotaryEmbedding()  
        )  
        (mlp): LlamaMLP(  
          (gate_proj): Linear(in_features=3072, out_features=8192, bias=False)  
          (up_proj): Linear(in_features=3072, out_features=8192, bias=False)  
          (down_proj): Linear(in_features=8192, out_features=3072, bias=False)  
          (act_fn): SiLU()  
        )  
        (input_layernorm): LlamaRMSNorm((3072,), eps=1e-05)  
        (post_attention_layernorm): LlamaRMSNorm((3072,), eps=1e-05)  
      )  
    )  
    (norm): LlamaRMSNorm((3072,), eps=1e-05)  
    (rotary_emb): LlamaRotaryEmbedding()  
  )  
  (lm_head): Linear(in_features=3072, out_features=128256, bias=False)  
)
```

# 소제목 생성(LLaMA3) 모델



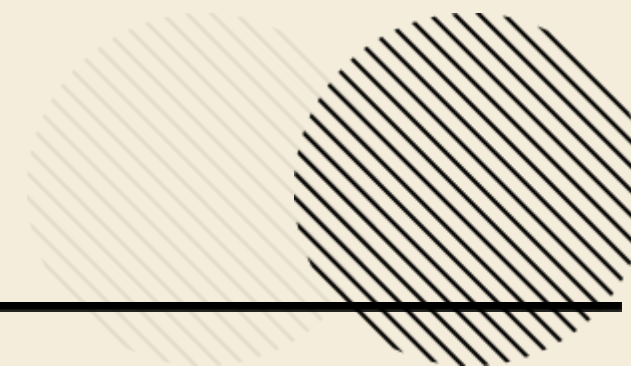
## QLoRA



**Figure 1:** Different finetuning methods and their memory requirements. QLoRA improves over LoRA by quantizing the transformer model to 4-bit precision and using paged optimizers to handle memory spikes.

- 양자화(Quantization) 기술 적용: 모델 파라미터를 8비트 이하로 양자화하여 메모리 사용량을 줄이고 계산 효율성을 높임
- 성능 저하 최소화: 양자화 기술을 사용하면서도 메모리 효율성과 모델 성능을 모두 유지
- 적은 메모리로 대형 모델 활용: 메모리를 절약하면서 대형 언어 모델을 제한된 하드웨어 환경에서 실행할 수 있음

# 소제목 생성(LLaMA3) 모델



## LoRA

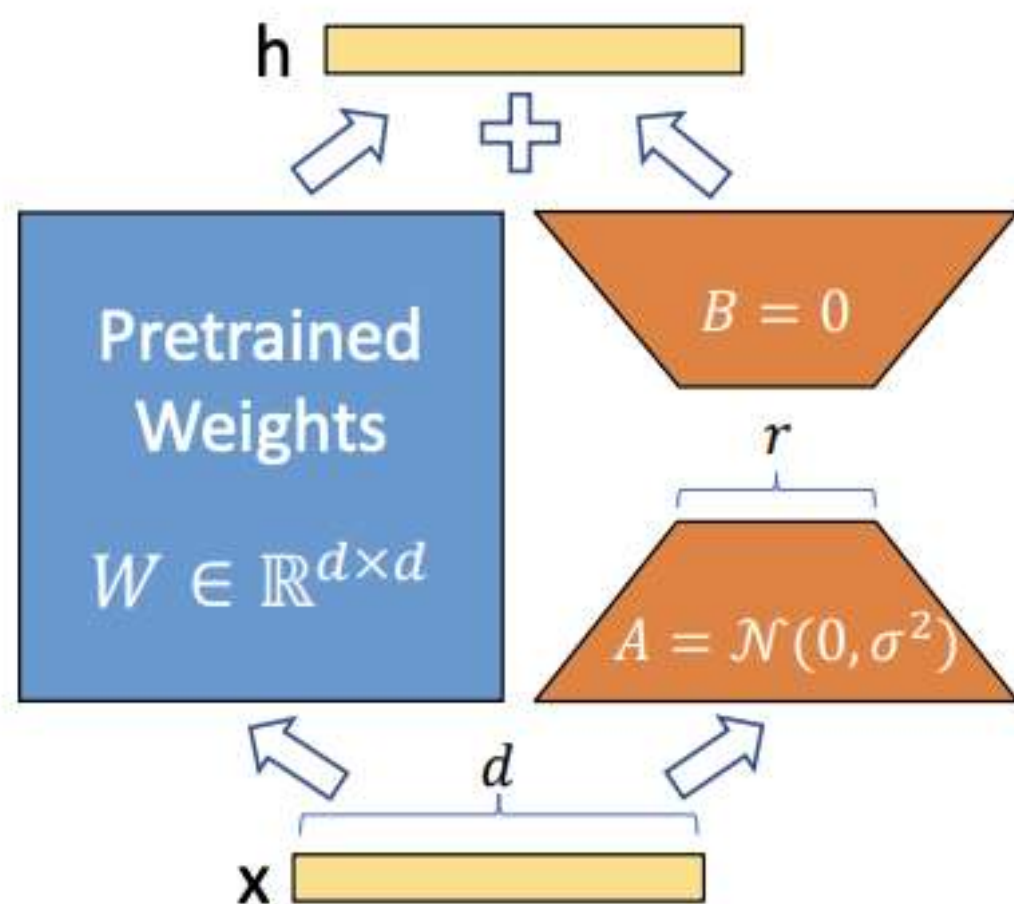
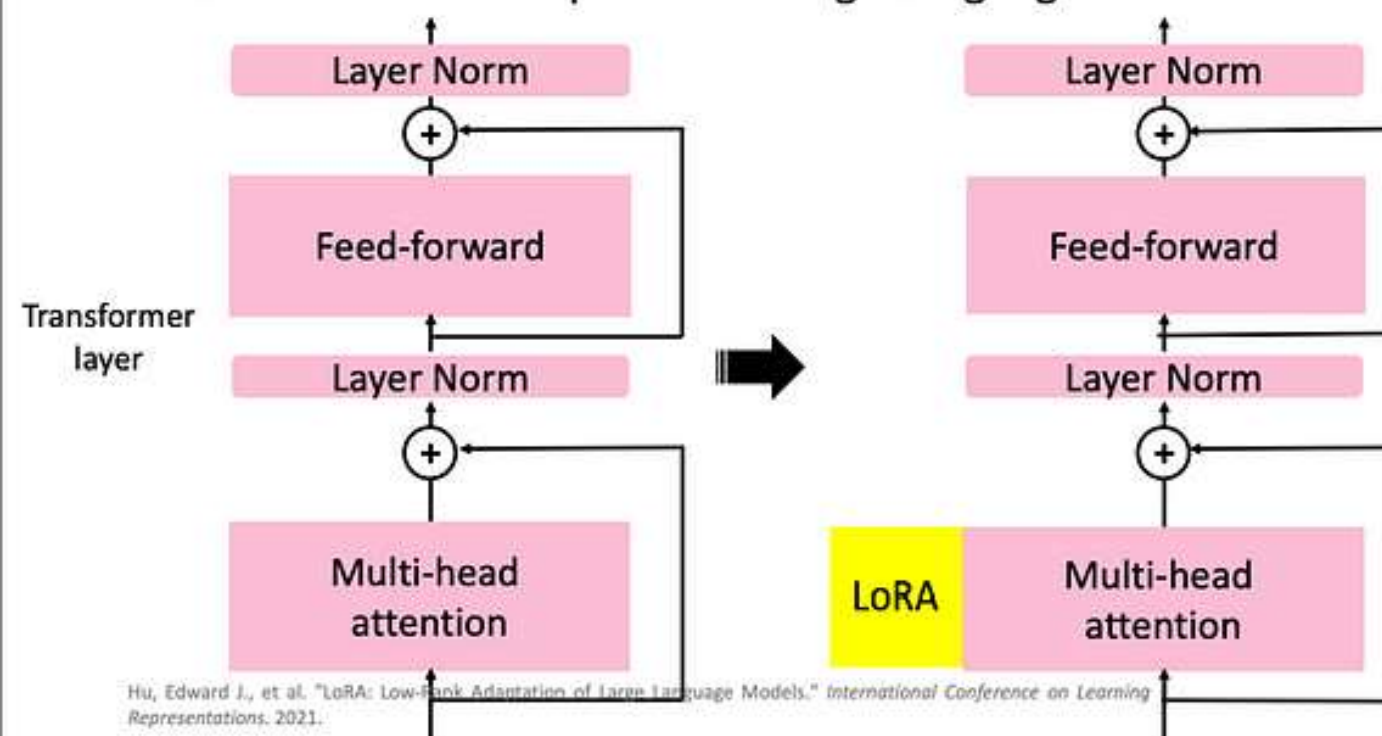


Figure 1: Our reparametrization. We only train  $A$  and  $B$ .

Slides credit: Cheng-Han Chiang, Yung-Sung Chuang, Hung-yi Lee, "AACL\_2022\_tutorial\_PLMs," 2022.

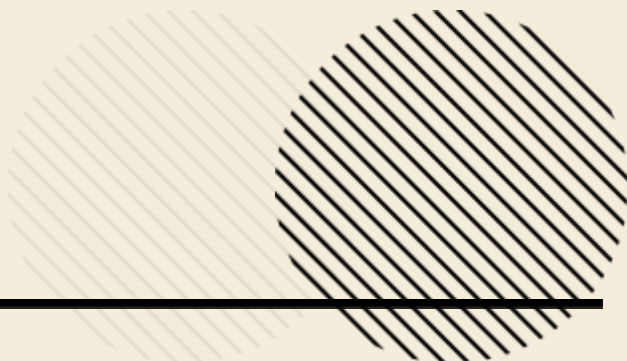
### Parameter-Efficient Fine-tuning: LoRA



- LoRA: Low-Rank Adaptation of Large Language Models



Hu, Edward J., et al. "LoRA: Low-Rank Adaptation of Large Language Models." *International Conference on Learning Representations*. 2021.

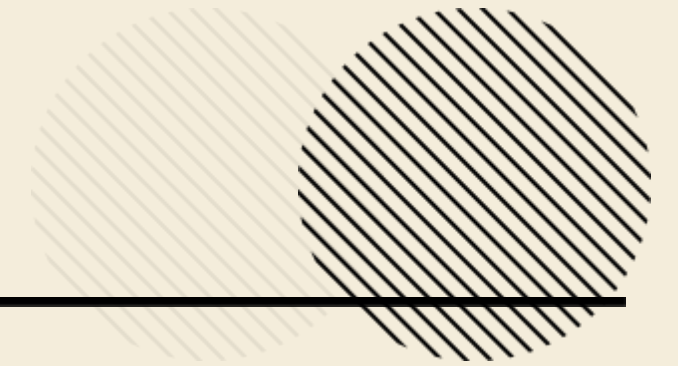
# 소제목 생성(LLaMA3) 모델



<div>cleaned_sub_title</div> <div>string · lengths</div> <div></div>	<div>cleaned_content</div> <div>string · lengths</div> <div></div>
소설 속 망나니	근신 처벌을 받은 에단은 별채로 돌아왔다 충분히 이해가 가는 처우였다 그간 저지른 패악과 사건들이 있으니 이 정도 처벌쯤은 달게 받아야 했다 그렇다고 하기에는 호화롭네 근신을 위해 마련된 별채였지만 시설은 무엇 하나 부족...
격투 천재	가주님이 입장하십니다 웅성거리던 인파가 양쪽으로 나뉘어 길을 텃다 마치 흥해가 갈라지는 것 같았다 소란이 잦아들고 사람들은 침묵했다 이것이 블란테 가문의 주인이 가지고 있는 입지였다 재밌네 흥미로웠다 시합을 뛰던 시절 언...
몬스터 토벌	그래 토벌대 말이더냐 ? 빈센트가 피식 웃으며 묻자 에단이 고개를 끄덕였다 그렇습니다 좋다 가주가 되어서 한 입으로 두말을 해서는 안 되겠지 토벌대를 꾸릴 권한을 주마 물론 그렇게 많은 인원을 차출해 나갈 수 없다는 것은 알고...
새로운 훈련	본격적인 체력 단련이 시작되었다 기술의 체득은 체력이 밑바탕이 된 이후에 쌓아 올려도 늦지 않는다 아무리 훌륭한 기술을 가지고 있어도 그 기술을 발현할 기반이 쌓여 있지 않다면 아무런 의미가 없었다 그런 의미에서 에단이 요구...
세력 구축	분위기가 냉각된다 소문의 속도는 빠르다 특히 폐쇄적인 집단에서 도는 소문은 더욱 빨랐다 에단이 카론을 이겼다 에단이 토벌대를 꾸린다 토벌대의 첫 인원으로 마구간 하인을 택했다 제정신이 아니다 자만이 하늘을 찌른다 이와 같...
천적의 만남	훈련은 탄력을 받았다 능력을 감추고 있던 네이드는 어쩔 수 없이 수습 기사들을 지도하고 있었다 이미 에단에게 비밀이 까발려진 터라 더 이상의 비밀 유지가 무의미하다고 판단한 것이다 인력이 부족했다 에단으로서는 쓸 수 있는...
<div><div>&lt; Previous</div><div>1</div><div>2</div><div>3</div><div>...</div><div>203</div><div>Next &gt;</div></div>	



# 소제목 생성(LLaMA3) 모델



## 프롬프트 형식

llama3 파인튜닝

```
def generate_prompts(example):  
    prompt_list = []  
    for i in range(len(example['cleaned_content'])):  
        prompt_list.append(  
            f"""<|begin_of_text|>  
            <|start_header_id|>user<|end_header_id|>다음 글의 제목을 지어주세요:  
            {example['cleaned_content'][i]}<|eot_id|><|start_header_id|>assistant<|end_header_id|>  
            {example['cleaned_sub_title'][i]}<|eot_id|>"""  
        )  
    return prompt_list
```

- <|begin\_of\_text|>: 텍스트의 시작을 나타냄
- <|start\_header\_id|>user<|end\_header\_id|>:  
사용자 입력의 시작을 나타냄
- <|start\_header\_id|>assistant<|end\_header\_id|>:  
모델 응답의 시작을 나타냄
- 다음 글의 제목을 지어주세요: 모델에 대한 지시 사항
- cleaned\_content[i]: 소설 내용
- cleaned\_sub\_title[i]: 서브 제목 (모델이 생성한 제목)

# 표지 생성 서비스의 필요성

- 작가 지망생에게 일러스트 구매는 부담이 커,  
표지 없이 연재하는 소설이 다수임

NAVER 웹소설 | 시리즈 | 웹툰

홈 슈퍼캐스팅 시리즈에디션 **베스트리크**

추천 **랭킹** 신규 승격작 **로맨스** 로판 판타지

★10.0 관심 611

HJ(소설쟁이) 총 87회  
★10.0 관심 631

실시간 랭킹

1 **UP** **이로운 밤**  
반석†  
★9.96 관심 681

2 **UP** **태오의 여자**  
윤이나23  
★9.99 관심 555

3 **UP** **열일곱 초희와스물다...**  
sahas  
★10.0 관심 723

4 **UP** **착한 오빠, 나쁜 오빠**  
봉자가  
★9.98 관심 427

5 **UP** **미칠 것 같아**  
멜로와인  
★10.0 관심 611

6 **UP** **도둑맞은 아내**  
뜨일러  
★9.98 관심 225

7 **UP** **교수님과 나쁜 짓**  
주은득  
★10.0 관심 241

8 **UP** **본부장님의 여대생 키...**  
골드선  
★10.0 관심 275

9 **UP** **미친개와의 계약 결혼**  
나오웅  
★10.0 관심 189

연재작 완결작

모든장르 작가,작품

1 2 3 4 5 6 >

독 몽둥이™ 이 상남자 천재 작가...

독 후라이드11 나 혼자 초월급 영력...

독 정킴 내 몬스터만 스렛 99...

독 드형컨66 농사만 지었더니 비...

독 글리세롤 무한빙의로 갓갯 개발...

독 버랑글 귀환했는데 게임형...

독 주문향 국가대표 침술명의...

독 폭파중 무사 컨셉 잡다가 아...

독 글럭럭 EX급 먼지킨이 되었다

독 조2 방랑 기사로 회귀했...

독 게티 천재의사는 두 여자...

독 아온 황보세가 역대급 무...

독 흑우우 재벌집 사생아로 사...

독 shoro 아만인으로 시작하...

독 동강강 고백했다가 까이고 ...

독 신용하 차원 넘는 마법사

독 파인블루 불멸의 톱스타

독 서인하 옆집에 대표님이 이...

독 하늘곰 좀비 물기 좋은 날

독 돈많은현랑 펜싱 금메달은 소드...

독 이하담담 아포칼립스에서 나 ...

독 소말자 버려진 재벌가 자남...

독 원태훈 아포칼립스 미국에...

독 광별 처가머슴 딱려치고 ...

탑스타의 19 비디오를 얻었다

무심코 컨티비, 그 주인공이 나타났다.

최근 주목 받는 작품

작가연재 일반연재 자유연재

아역 배우

나만, 연기가... 현대판타지 팔란

목숨을 건 미... 현대판타지 칼텍스

무공으로 다... 현대판타지 불릭통통

착하게 살수... 스포츠 백아™

검은머리 박... 현대판타지 아아연하게

전재 공화자의 역대급 영지 설계

현대 판타지

방사능 먹는 판타지

무협

이혼 후 거물

최전 후 초월급 판타지

암살노트로 국가권력급 각성자

초월급 우주... 퓨전 coldpi...

EX급 드래곤... 현대판타지 공수자리

비밀번호

☐ 로그인 상태 유지

간편 회원가입 아이디/비밀번호 찾기

바로 사용가능한 5,000골드 지급

SNS 계정으로 로그인

N G f a P

단축키 안내

H 처음으로 B 내서재로

F 내선호작으로 M 무료웹소설로

N 작품목록열기

나 혼자서 유료화 축하는 골드로 할게요

100,000 골드 받기 >

MD추천 더보기+

무협 윤희무사 서마

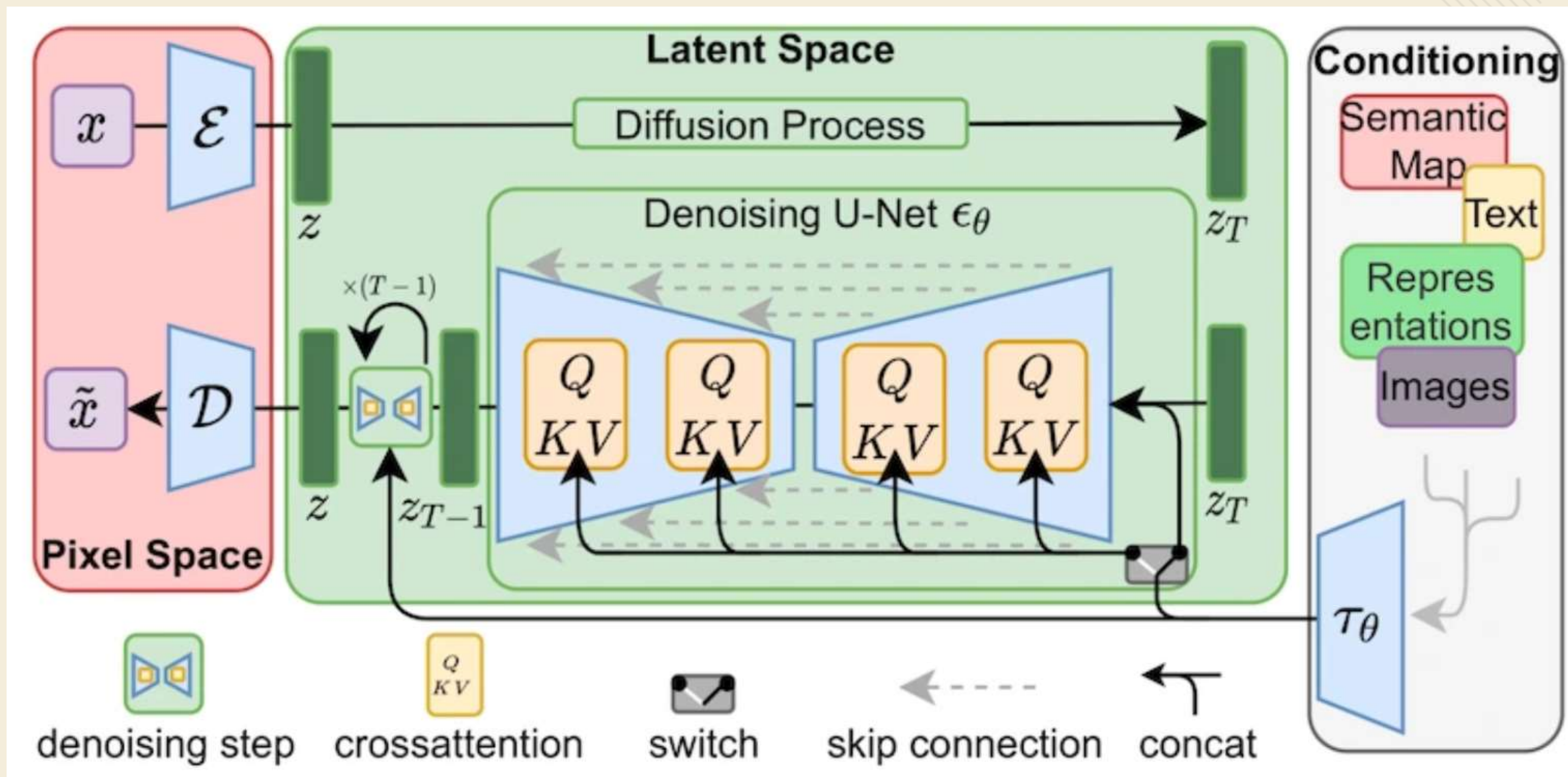
스포츠, 현대판타지 선수 능력이 다 보임 탈주병

현대판타지 K-문화의 제왕 고별

현대판타지, 판타지 금룡의 마법사 캠퍼아빠



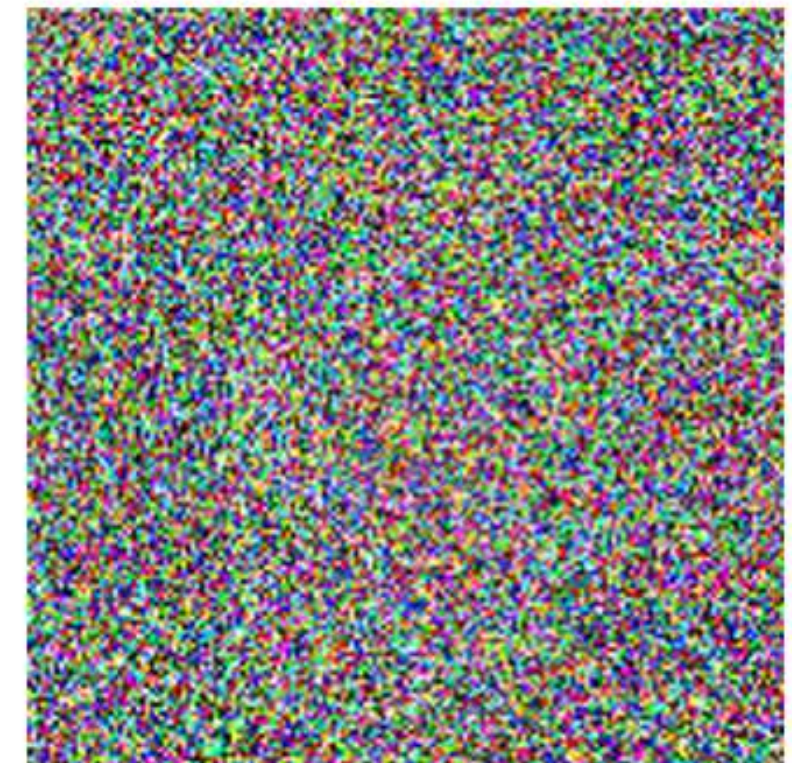
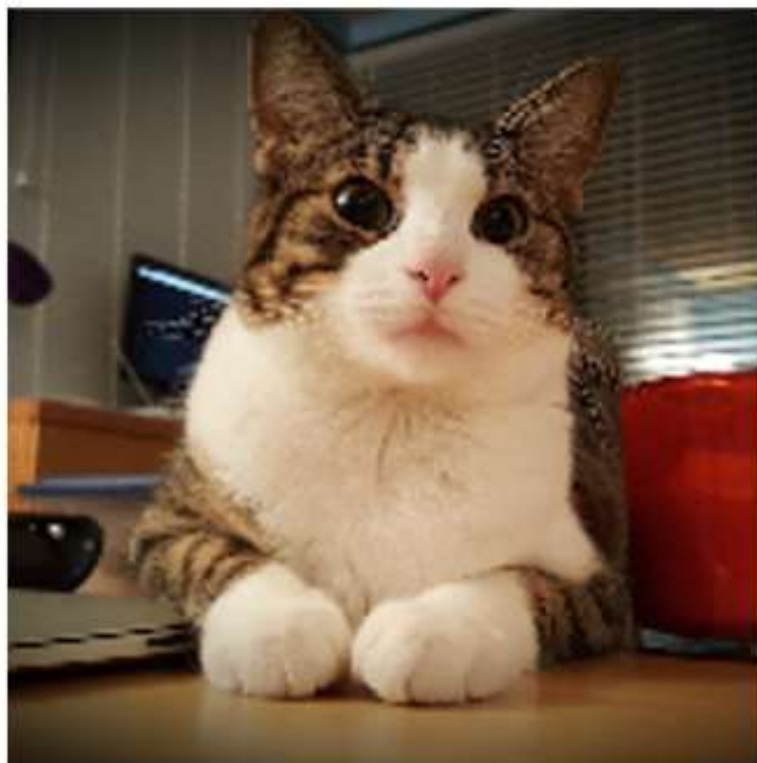
# 표지 생성(Stable Diffusion) 모델





# 표지 생성(Stable Diffusion) 모델

---





# 표지 생성(Stable Diffusion) 모델



## DreamBooth vs Textual Inversion vs LoRA

	특징	장점	단점
DreamBooth	특정 개체나 스타일에 대한 파인튜닝	적은 수의 이미지로 특정 개체 학습 가능, 매우 정확한 이미지 생성	일반화 능력 부족, 개체별로 별도의 파인튜닝 필요
Textual Inversion	텍스트 프롬프트에 대한 고유한 임베딩 학습	빠르고 효율적 학습. 스타일을 쉽게 추가 가능	기존 모델의 성능에 의존. 세밀한 스타일 제어 한계
LoRA	모델의 일부 파라미터만 업데이트하는 저자원 학습 기법	자원 절약. 빠른 실험 가능. 모델의 안정성 유지	특정 개체나 스타일 학습에 대한 성능이 떨어질 수 있음

# 표지 생성(Stable Diffusion) 모델

## 파인튜닝 전후 비교

```
pipeline.safety_checker = None

pipeline_output = pipeline(
    prompt=["검은 정장의 남자와 드레스를 입은 여자"],
    num_inference_steps=25,
    generator=torch.manual_seed(101)
)

pipeline_output.images[0]
```

0% | 0/25 [00:00<?, ?it/s]



```
lora_scale = 0.4

pipeline.safety_checker = None

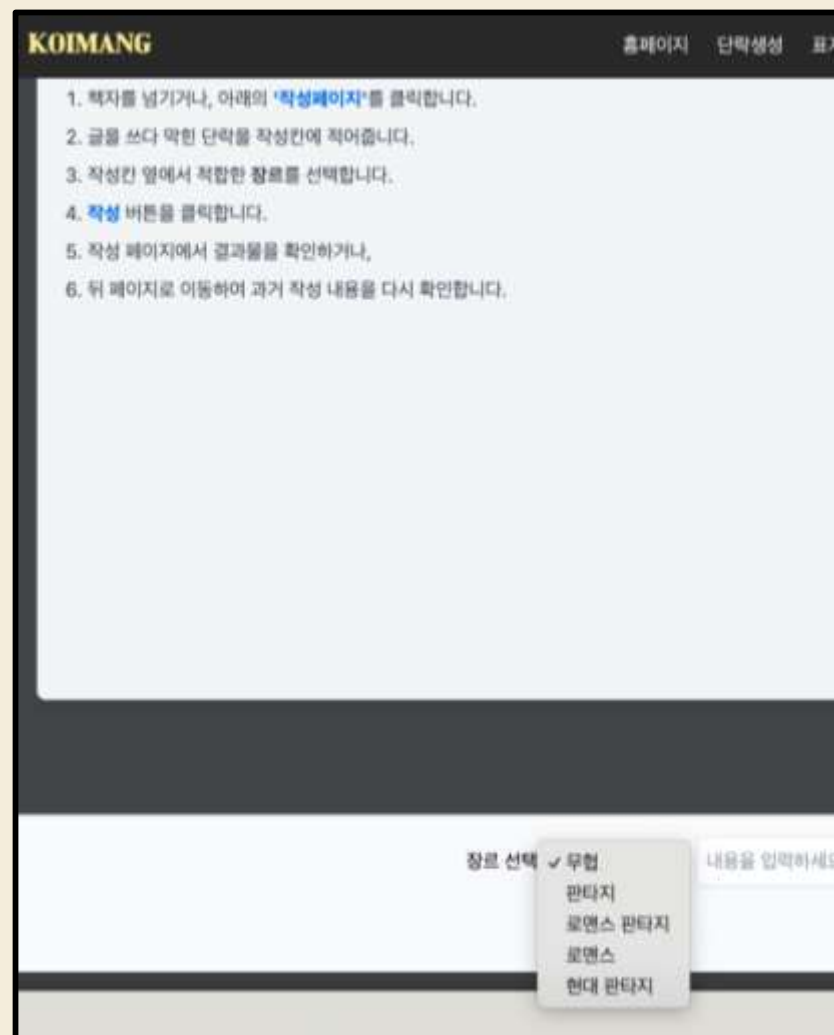
pipeline_output = pipeline(
    prompt=["검은 정장의 남자와 드레스를 입은 여자"],
    num_inference_steps=25,
    cross_attention_kwargs={"scale": lora_scale},
    generator=torch.manual_seed(101)
)

pipeline_output.images[0]
```

0% | 0/25 [00:00<?, ?it/s]

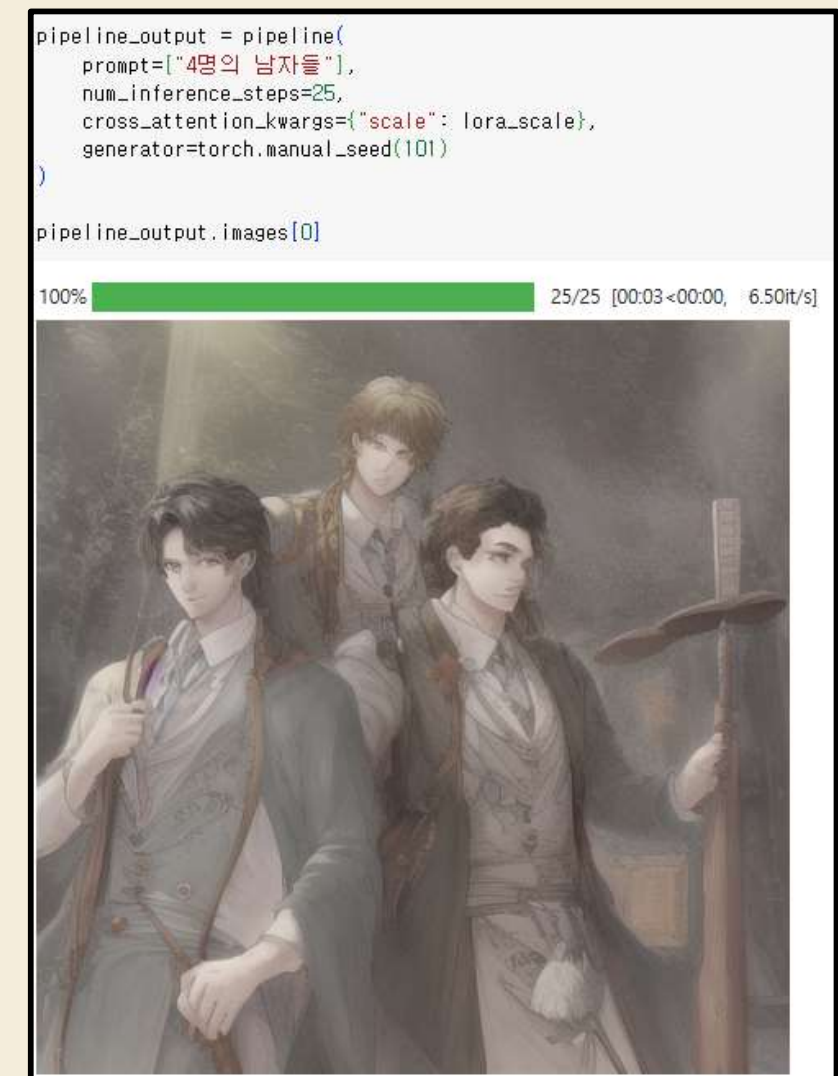
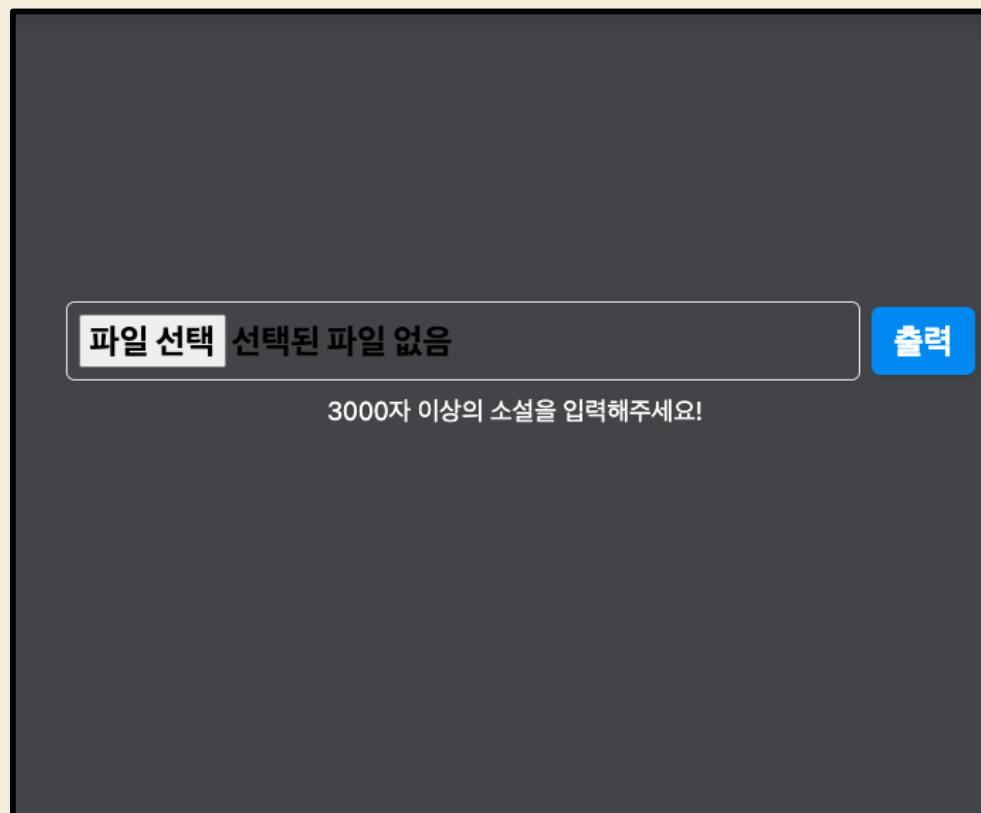


# 개선 방안



- 단락 생성 모델: 더 다양한 장르 추가 학습

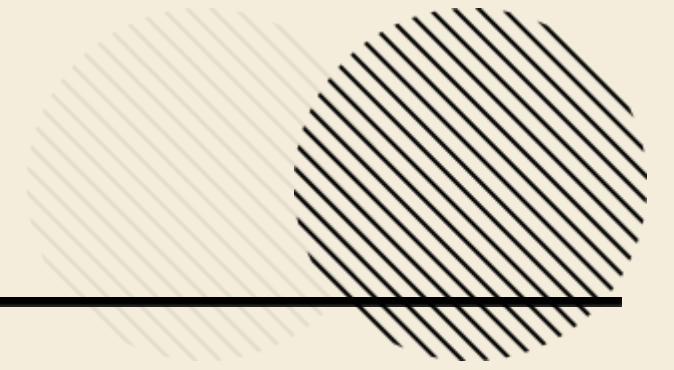
- 제목 생성 모델: 장르를 추가해 해당 장르에 어울리는 제목 생성



- 표지 생성 모델: 더 다양한 프롬프트 학습 필요

# 질의응답

---



## Q & A

---



2024.12.27.(금)

# KOIMANG NOVEL HUB

발표를 마무리합니다, 감사합니다 ☺

