

# GPT 모델로 수필을 써보자



유승한  
GDSC Soongsil  
jorzor99@gmail.com

```
... == interbyorg : true  
if (Status = filterByStatus ? study.status === filterByStatus : true  
    matchStatus) {
```

```
function filterStudies({ studies, filterByOrg = false, filterByStatus = false }) {  
  return studies.filter(study => {  
    if (filterByOrg) {  
      return study.lead_organization === filterByOrg
```

빛과 색채의 연작을 감상하던 그 순간. 내 마음에 들어온 물 위로 까슬하게 늘어져 있는 수련의 모습은 생경하면서도 슬펐고, 아름다웠으며 고통스러웠던 기억들을 떠올리게 했다. 바로 나의 작은 변화들에 대한 것이었다. 그 시절에도 나에게 영향을 미치고, 기억할 수 없었던 것들을 만나게 되곤 한다. 그리고 그때를 떠올리면, 참 많은 추억들이 있다.



---

미술은 내 안에 있는 열정의 불씨를 일으켜 주었다. 나는 그것을 통해 나 자신을 완성해나갔다. 새로운 인생을 열렬히 꿈꿨고 이는 나의 삶을 바꾸는 계기였다. 내 인생의 변곡점에 가깝지 않나 싶었다. 그가 그린 그림처럼 세상을 매일매일 새로운 모습으로 살아갔고, 정해진 코스 외엔 한 번도 나오지 않았던 고래섬을 구경하며 바닷 바람과 맞부딪치는 물고기들을 만나거나, 혹은 사람이 살아가면서 만나는 불가사의한 바다를 함께 항해하였다. 하늘의 일부가 된 나뭇잎, 예쁜 꽃들을 보며 내 마음 속에 미소가 번졌다. 그렇게 나의 성격은 이전과는 다른 사람으로 차차 바뀌어갔다.



미술은 내 안에 있는 열정의 불씨를 일으켜 주었다.  
이어서 소설을 써줘



미술관에서 그림을 바라보던 그날, 내 안에 있는 열정의 불씨가 일어났다. 그림 속에서 나는 내 안에 있는 것들을 찾아낼 수 있었고, 그것을 통해 나는 무엇을 해야 할 지를 알게 되었다.  

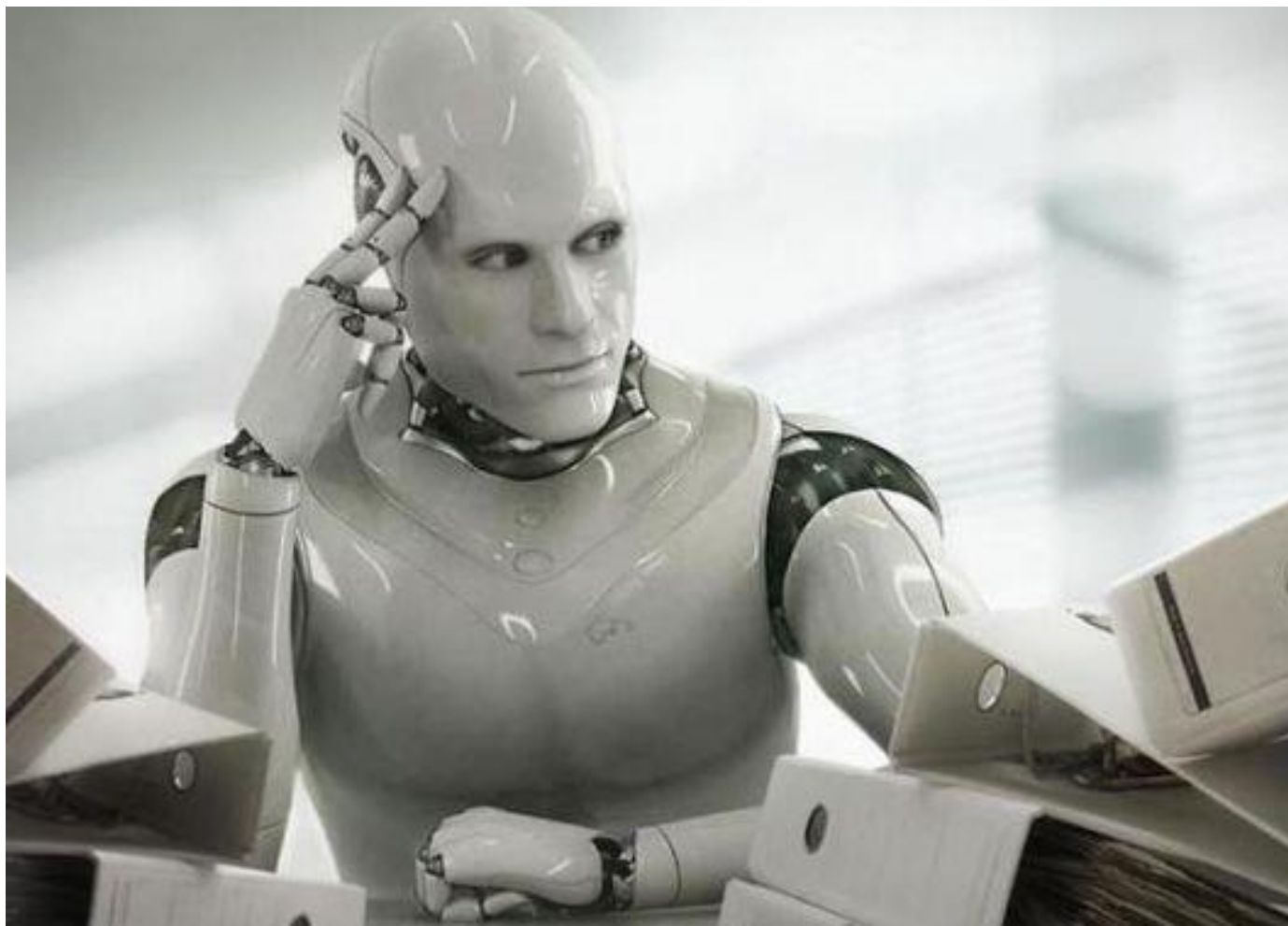
나는 작은 아파트에서 살면서도 미술관으로 자주 다녀왔다. 거기서 내가 본 그림들은 나에게 큰 영감을 줬다. 나는 내 안에 있는 열정의 불씨를 따라서 미술에 더욱 빠져들었다.

그렇게 미술에 대한 열정이 커지면서, 나는 작은 스튜디오를 임대하게 되었다. 거기서는 나의 미술적인 열정을 마음껏 표현할 수 있었다. 내가 그림을 그리면서, 내 안에 있는 무엇인가를 발견하고, 그것을 그림 속에 담아낼 수 있었다.

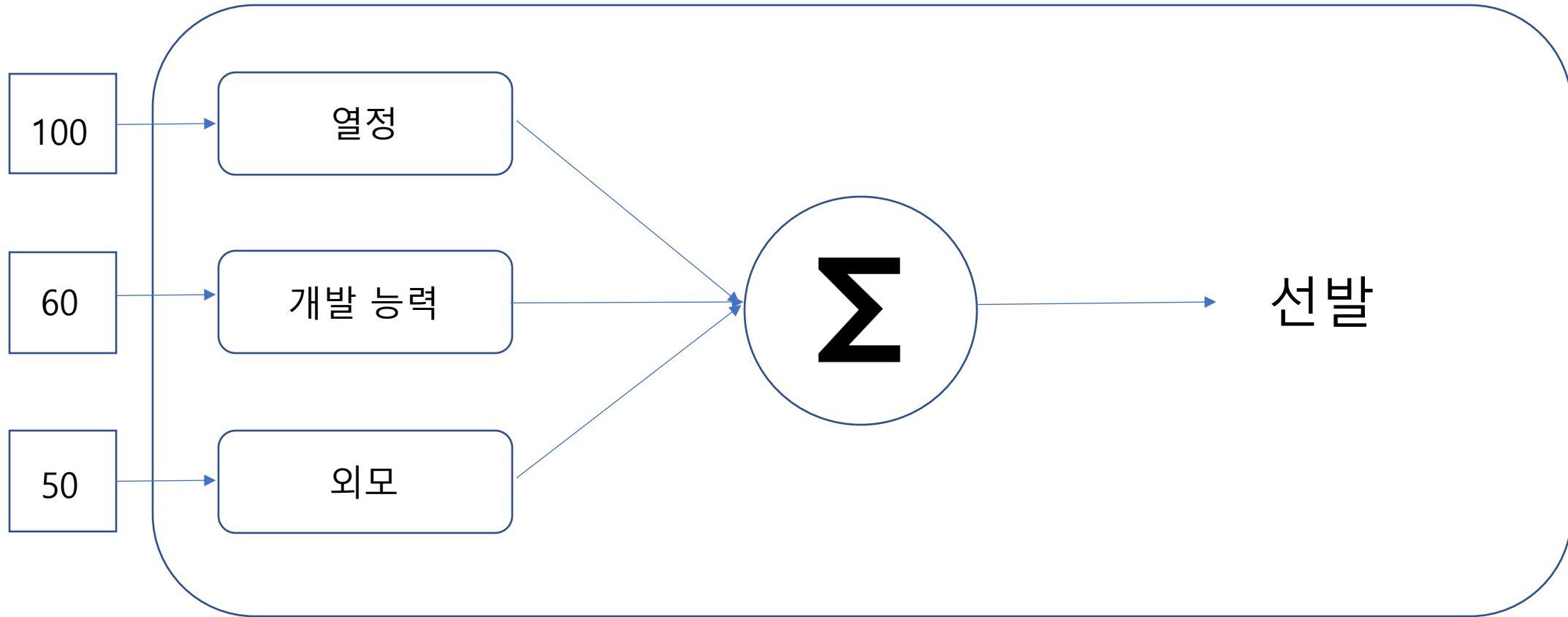
나는 작업을 하면서도 많은 어려움을 겪었다. 하지만 내 안에 있는 열정의 불씨가 나를 이끌어주었다. 그 불씨는 나를 지속적으로 동기부여하고, 새로운 아이디어를 불러일으키며, 내가 더욱 발전할 수 있도록 돕고 있었다.

그리고 나는 미술관에서 그림을 바라보며 시작된 내 열정의 여정을 기반으로한 소설을 써보았다. 이 소설은 나의 내면을 다루며, 미술과 열정의 중요성을 강조한다. 그리고 이러한 소설은 미술에 대한 열정을 갖고 있는 많은 사람들에게 위로와 동기부여를 제공할 수 있을 것이다.

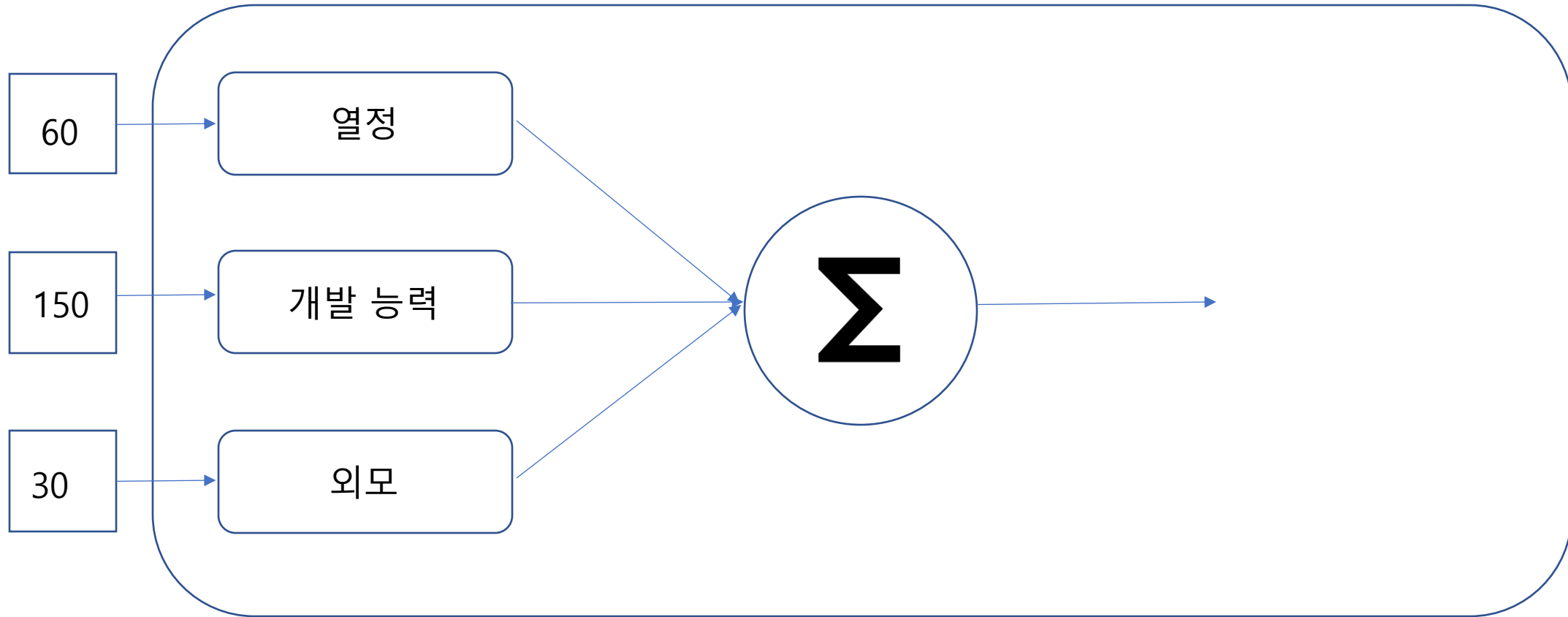
미술은 내 안에 있는 열정의 불씨를 일으켜 주었다. 그리고 그 불씨는 지금도 나를 이끌어주고 있다. 나는 미술을 통해 나 자신을 발견하고, 무엇을 하고 싶은 지를 알아낼 수 있었다. 나는 내 안에 있는 열정의 불씨를 지속적으로 살려가며, 미술과 함께 내 인생을 살아가고자 한다.



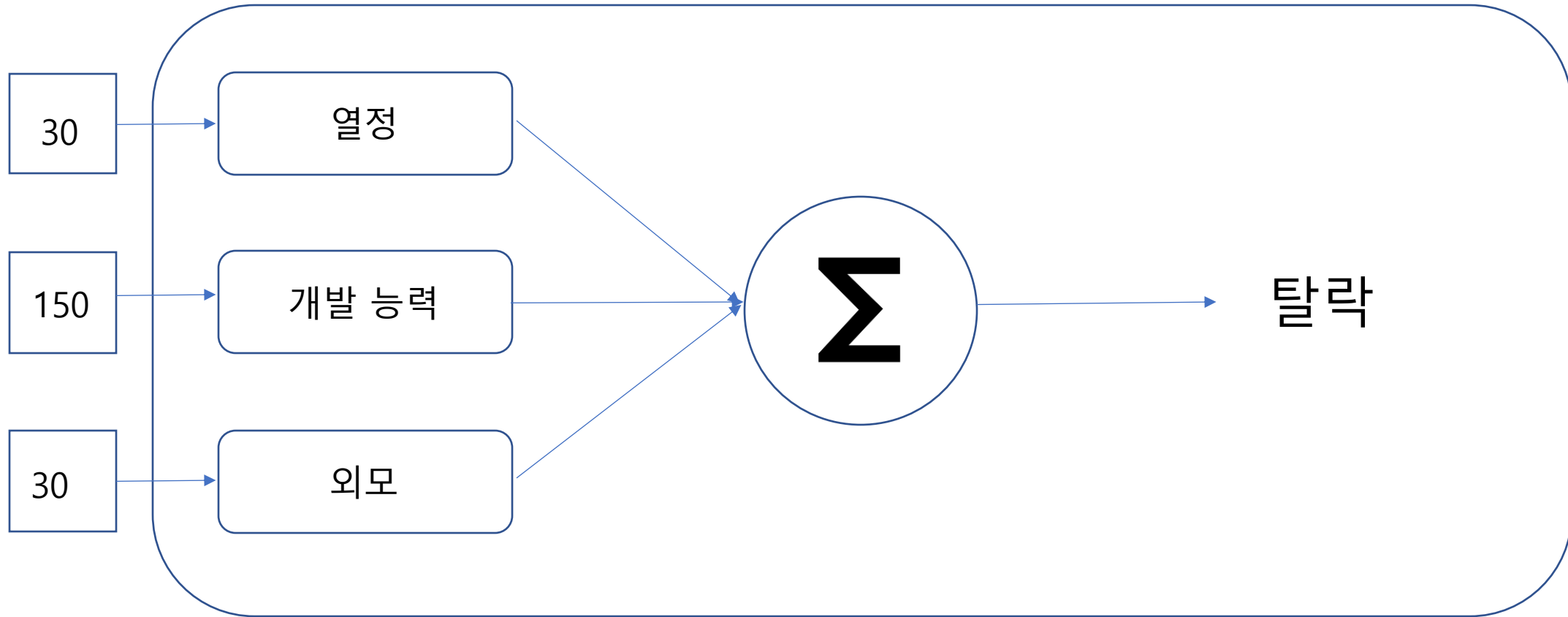
# PERCEPTRON



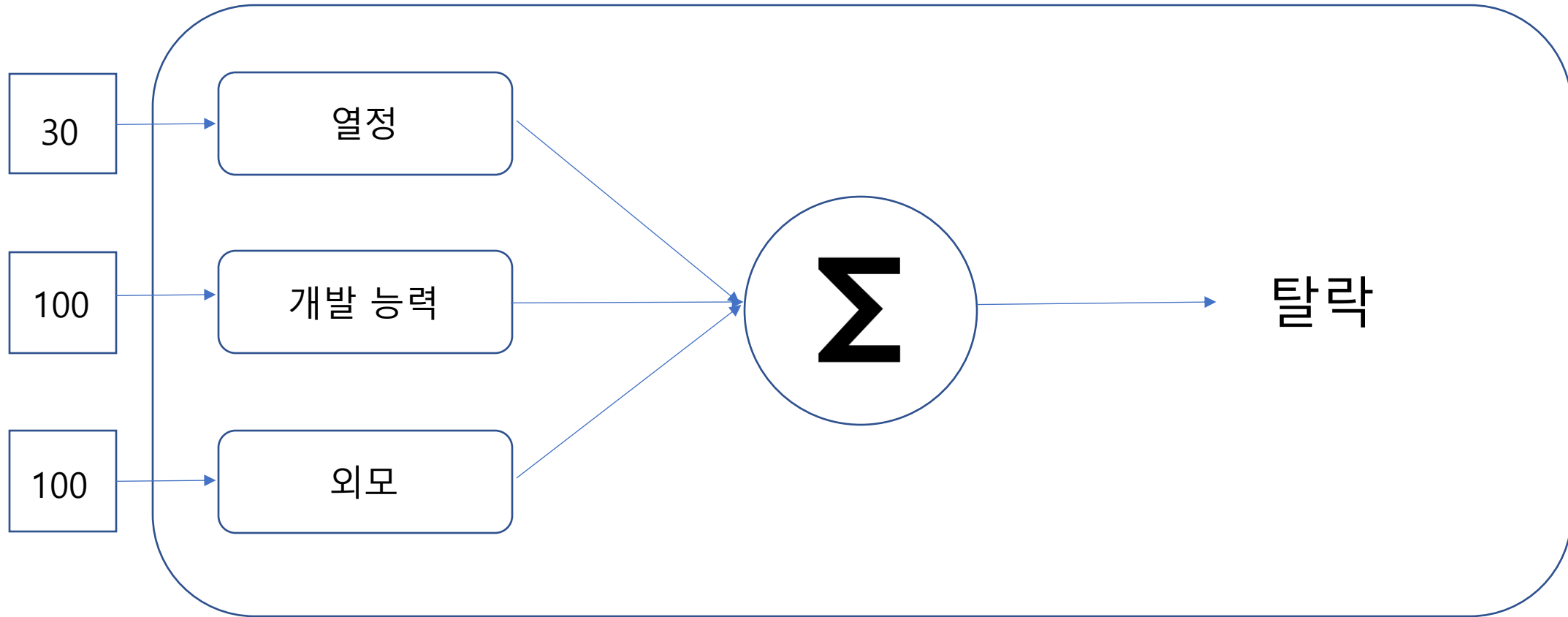
# PERCEPTRON



# PERCEPTRON



# PERCEPTRON

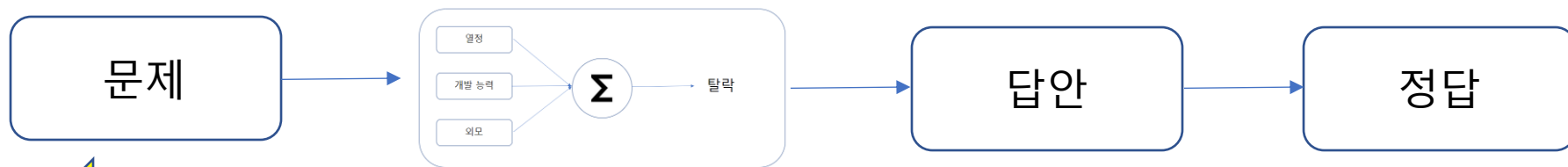




# AI가 공부하는 법

1. 문제를 받아서 결론을 도출.
2. 정답이랑 다르네?
3. 정답과 어떻게 다른지를 생각하고, 어떻게 해야 정답을 낼 수 있는지 고민한다.
4. 자시의 수치(parameter, weight)를 수정해서 다시 시도해본다.
5. 위 1~4 과정을 반복한다

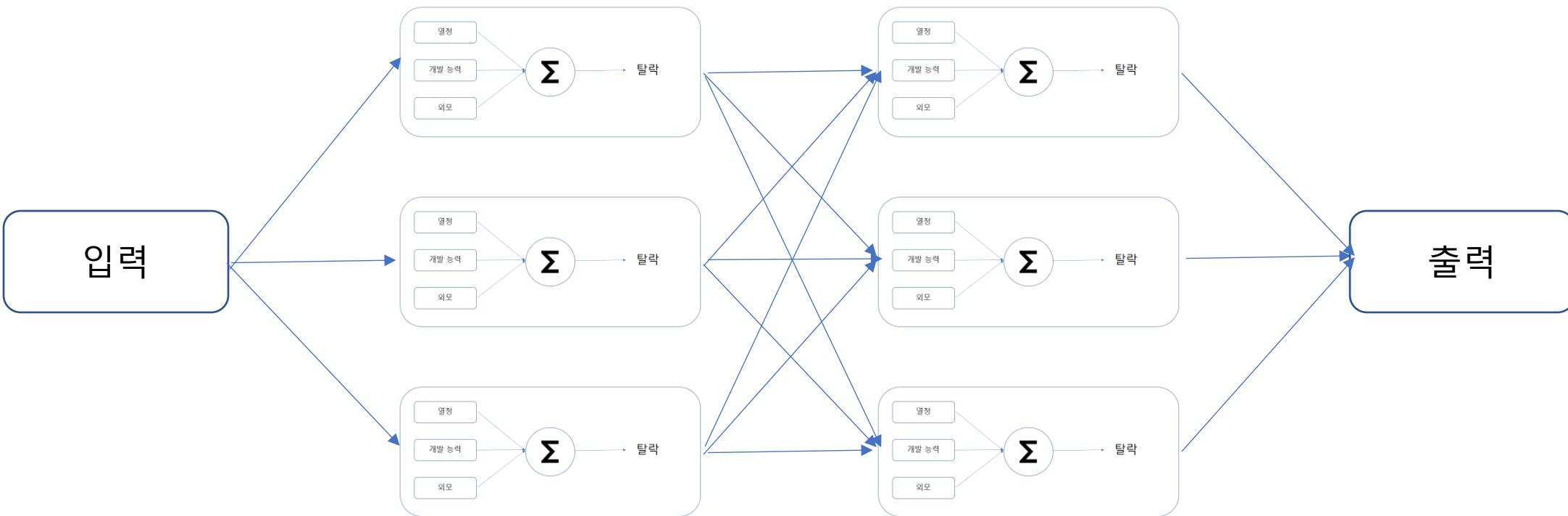
**Forward pass**



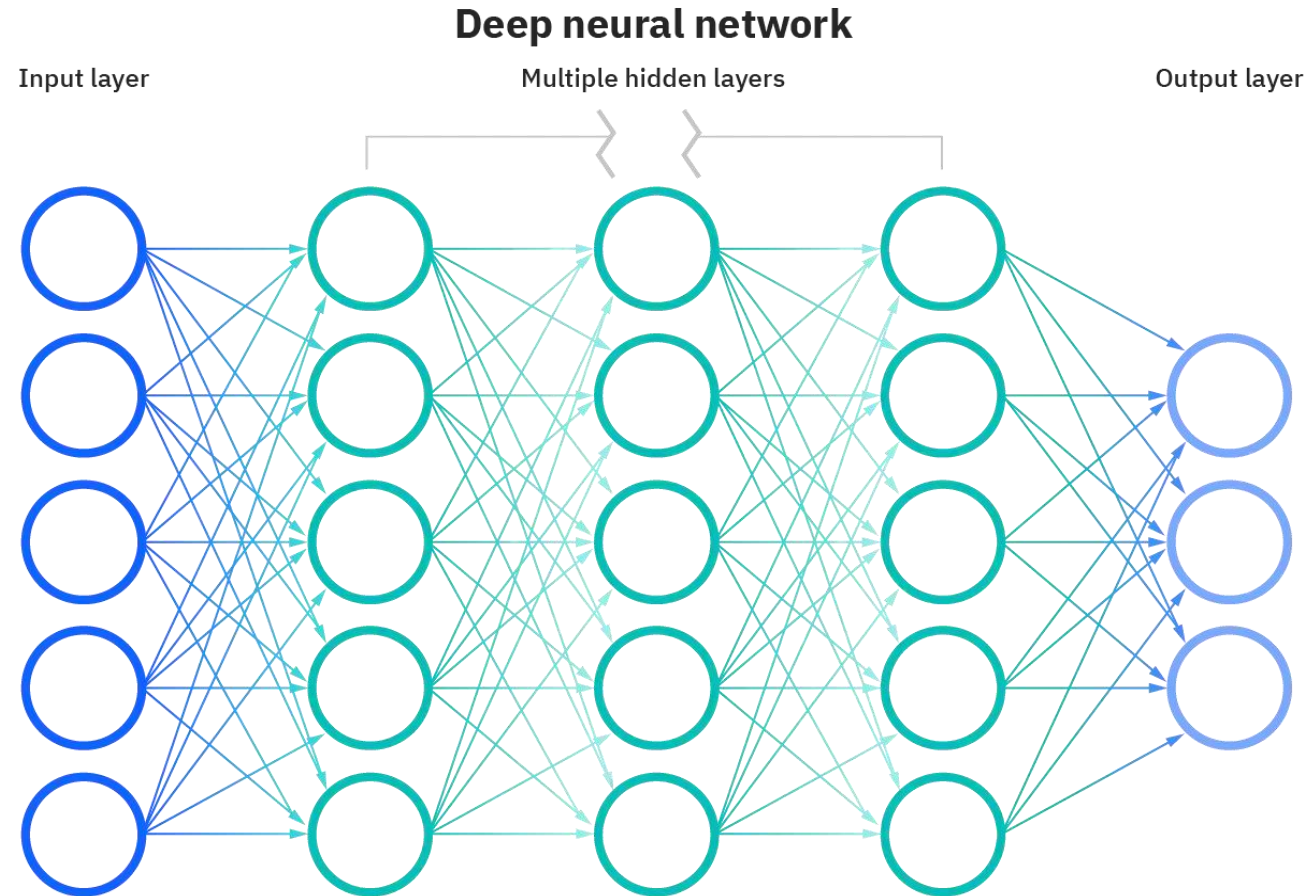
**Backward pass**

<https://mindslab.ai/>

# PERCEPTRON

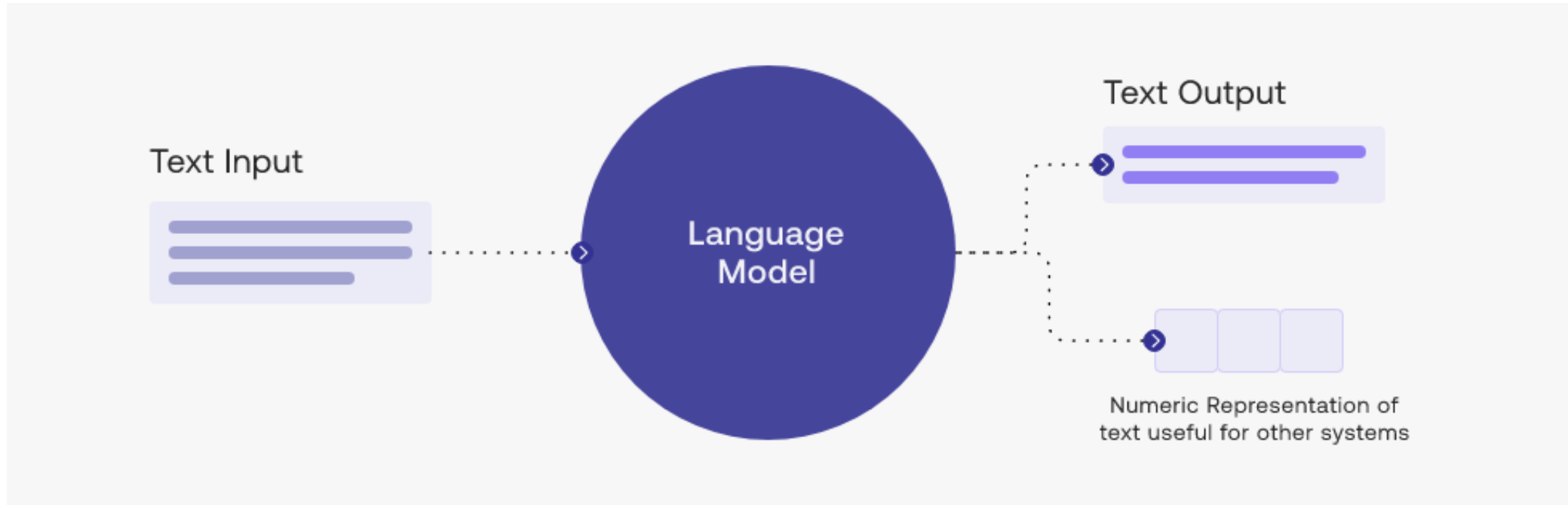


# Deep Neural Network



<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>

# Language Model



<https://docs.cohere.ai/docs/introduction-to-large-language-models>

# Language Model

단어들을 수직선에 위치시켜보자

교통 ↔ 경찰 ↔ 조폭 ↔ 폭력 ↔ 고통

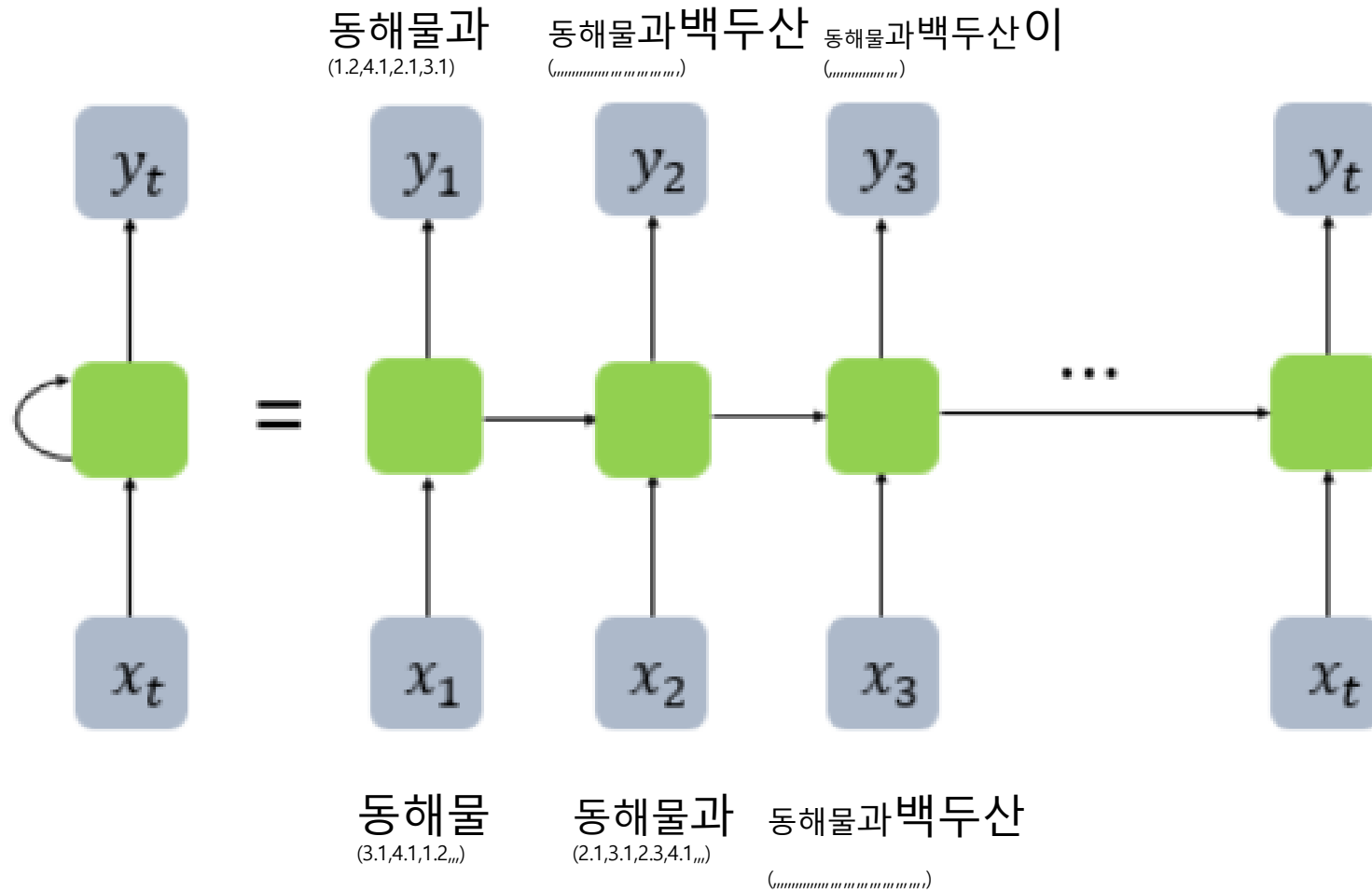
- 교통과 고통은 매우 닮았지만 의미적으로는 전혀 다르다

언어는 의미가 모여서 흐르는 Sequence

예나 ↔ 선정이 딸이에요

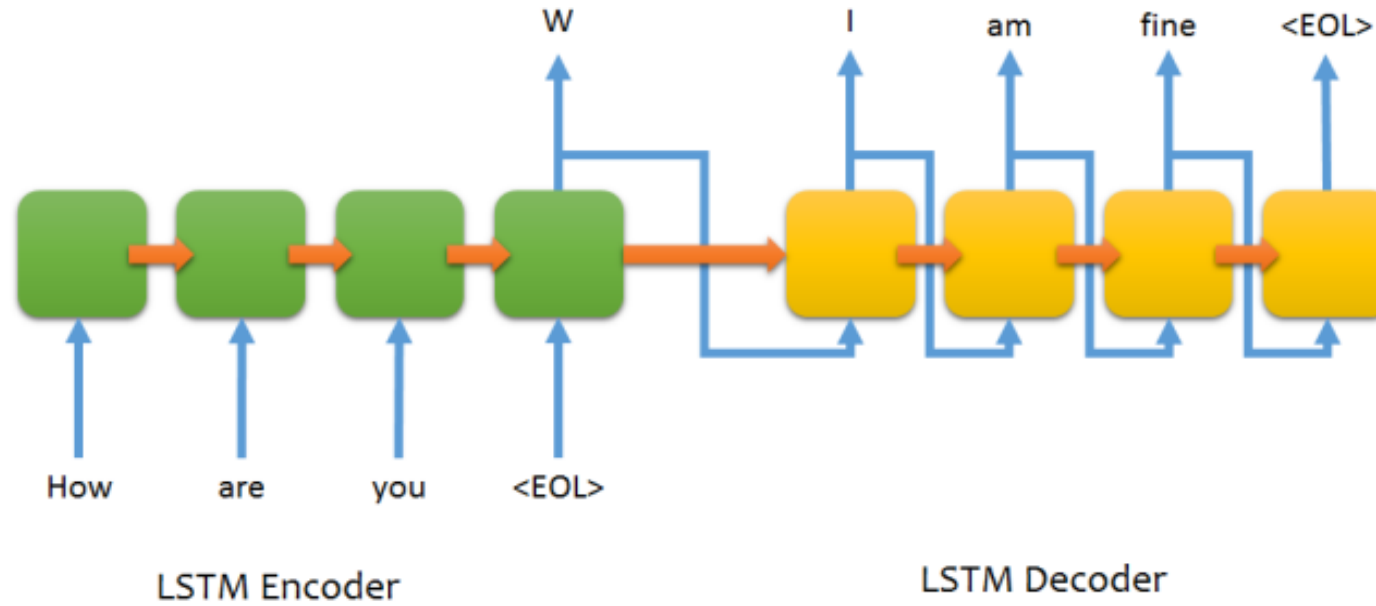
누구인가? ↔ 누가 지금 기침 소리를

# RECURRENT NEURAL NETWORK



<https://wikidocs.net/22886>

# SEQ2SEQ



<https://hugrypiggykim.com/2017/09/21/tensormsa-guide-seq2seq/>

# RNN의 문제

1. 입력이 길어질 시 처음 입력했던 데이터를 기억 못함  
(RNN의 고질적인 문제인 기울기 소실(vanishing gradient) 문제가 존재)

나 : 커서 뭐가 되고싶니?

RNN : 프로그래머가 되고 싶어요

나 : 그럴라면 게임을 많이 해야겠네

RNN : 요즘 젊은 사람들이 하라는 공부를 안하고 게임만 하고, 에휴 나때는 말이여

2. 각 단어의 의미를 정의하는데 있어서, 단어 간의 거리 혹은 순서보다 관련이 있는 특정 단어가 더 중요함  
(하나의 고정된 크기의 벡터에 모든 정보를 압축하려고 하니까 정보 손실이 발생)

나 : 네 이름이 뭐니?

RNN : 저는 게임을 좋아하는 AI 유튜버예요.

나 : 가장 좋아하는 게임이 뭐야?

RNN : 저는 유튜버 침착맨을 좋아해요.



# ATTENTION IS ALL YOU NEED



# GENERATIVE PRETRAINED TRANSFORMER

생성적 사전학습 트랜스포머

1. 생성?
2. 사전학습?
3. 트랜스포머?

# GENERATIVE PRETRAINED TRANSFORMER

생성적 사전학습 트랜스포머

1. 생성?

머리부터 발끝까지?

2. 사전학습?

1. 사랑스러워 : 0.88888
2. 오로나민 C : 0.10000
3. 핫이슈 : 0.01110
4. !!!!! : 0.00001

3. 트랜스포머?

# GENERATIVE PRETRAINED TRANSFORMER

## 생성적 사전학습 트랜스포머

1. 생성?

인간이 사용하는 자연어에 대한 일반적인 학습을 진행(PRE-TRAINING) 하고

2. 사전학습?

이후에 각 세부 TASK에 대해 추가학습(FINE-TUNING)을 하자!

PRE-TRAINING + FINE-TUNING = TRANSFER LEARNING

3. 트랜스포머?

# GENERATIVE PRETRAINED TRANSFORMER

생성적 사전학습 트랜스포머

1. 생성?

번역에 쓸 목적으로 SEQ2SEQ를

2. 사전학습?

ATTENTION만으로 구현한 ARCHITECTURE

3. 트랜스포머?

# GENERATIVE PRETRAINED TRANSFORMER

생성적 사전학습 트랜스포머

1. 생성? 처음부터 각 단어마다 다음 단어를 순서대로 예측하도록
2. 사전학습? 다량의 자연어 데이터를 통해 인간의 언어에 대한 기초지식을 미리 학습해둔
3. 트랜스포머? 라는 구조를 활용한 인공지능 언어모델

# GPT1 VS 2 VS 3

## GPT-1 VS GPT-2

- > Unsupervised Multitask Learners
- > 더욱 방대한 데이터 더욱 거대한 모델구조
- > 파인튜닝을 안하더라도 세부 task를 처리하는 능력을 지님

## GPT-2 VS GPT-3

- > parameter를 1750억개로 키운 매우 거대한 GPT
- > 진정한의미로 파인튜닝이 필요 없어짐.
- > Few-Shot Learners : 몇개의 예시만으로 문제해결이 가능해짐

# AIxBOOKATHON

**2021.11.16(화)~11.17(수)**  
성균관대학교 학술정보관  
온라인(Webex, YouTube) 실시간 생중계, 비대면

**제3회 대학생 AI x BOOKATHON 대회**

참가 자격 | 한국 대학교 학생  
3학년 및 4학년 학부생이나 대학 수험생 제외  
예산 상가 후 본선 참가비 선형

**일정 안내**

신청 및 접수	11.02(토) ~ 11.02(토)	접수 마감
AI 특강 및 예선	11.04(토) ~ 11.04(토)	AI 특강 및 예선 문제 출제
대회 사전 교육	11.11(토) ~ 11.11(토)	AI 알고리즘 활용 교육
사전 학습	11.11(토) ~ 11.11(토)	대외에 수월 AI 데이터 학습
복합본 대회 진행	11.16(토) ~ 11.16(토)	주제에 따른 과제 수행
결과 발표(시상식)	11.17(토) ~ 11.17(토)	상사 및 결과 발표(시상식)

※ 참가비 및 1인 이상 시금교육 활동  
※ 사전교육(11월)을 이수한 후 본선 대회에 참가할 수 있음  
※ 대회 참가비 및 대회료  
※ 대회 참가비(1인)은 누구나 참여 가능(성균관대학교, 대학에 종속하여 미인정된 사립대 제외)

**시상 내용** | 대상(200만원), 우수상(100만원), 장려상(50만원)

**참가 신청** | 학술정보관 홈페이지(www.sku.ac.kr) 또는  
QR 코드 스캔 후 신청서 제출  
신청 URL: <http://naver.me/9K2m5d7k>  
※ 신청 신청에 따라 참가비 및 접수비 없음

**대회 참가 특징**

- AI 특강 참여자 중 상품 추천
- 본선 대회 참여자 전원 기념품, 식대 지원
- 입상자 네이버 입사 서류 전형시 가점 부여

문의 | [samsung@skku.ac.kr](mailto:samsung@skku.ac.kr)

주관 | 성균관대학교 학술정보관

후원 | NAVER, SKKU AI, SKKU AI, SKKU AI

**2023.01**  
성균관대학교 학술정보관  
온라인(Webex, YouTube) 실시간 생중계, 비대면

**제4회 대학생 AI x BOOKATHON 대회**

참가 자격 | 한국 대학교 학생  
3학년 및 4학년 학부생이나 대학 수험생 제외  
예산 상가 후 본선 참가비 선형

**일정 안내**

신청 및 접수	11.02(토) ~ 11.02(토)	접수 마감
AI 특강 및 예선	11.04(토) ~ 11.04(토)	AI 특강 및 예선 문제 출제
대회 사전 교육	11.11(토) ~ 11.11(토)	AI 알고리즘 활용 교육
사전 학습	11.11(토) ~ 11.11(토)	대외에 수월 AI 데이터 학습
복합본 대회 진행	11.16(토) ~ 11.16(토)	주제에 따른 과제 수행
결과 발표(시상식)	11.17(토) ~ 11.17(토)	상사 및 결과 발표(시상식)

※ 참가비 및 1인 이상 시금교육 활동  
※ 사전교육(11월)을 이수한 후 본선 대회에 참가할 수 있음  
※ 대회 참가비 및 대회료  
※ 대회 참가비(1인)은 누구나 참여 가능(성균관대학교, 대학에 종속하여 미인정된 사립대 제외)

**시상 내용** | 대상(200만원), 우수상(100만원), 장려상(50만원)

**참가 신청** | 학술정보관 홈페이지(www.sku.ac.kr) 또는  
QR 코드 스캔 후 신청서 제출  
신청 URL: <http://naver.me/9K2m5d7k>  
※ 신청 신청에 따라 참가비 및 접수비 없음

**대회 참가 특징**

- AI 특강 참여자 중 상품 추천
- 본선 대회 참여자 전원 기념품, 식대 지원
- 입상자 네이버 입사 서류 전형시 가점 부여

문의 | [samsung@skku.ac.kr](mailto:samsung@skku.ac.kr)

주관 | 성균관대학교 학술정보관

후원 | NAVER, SKKU AI, SKKU AI, SKKU AI

**대회 참가 특전**

- AI 특강 참여자 중 상품 추천
- 본선 대회 참여자 전원 기념품, 식대 지원
- 입상자 네이버 입사 서류 전형시 가점 부여



# AIxBOOKATHON - 데이터수집

## <수집한 데이터>

1. 글틴
2. 대표 에세이 문학회
3. 브런치
4. 동아 신춘문예
5. 모두의 말뭉치(비출판물 데이터)

## <데이터 수집 시 고려사항>

1. 학습을 위해 최대한 많이 수집하기
2. 어떤 종류의 데이터를 수집할지
3. 저작권이 위배되지 않는지

# AIxBOOKATHON - 데이터전처리

정규 표현식을 이용하여 학습에 방해될 것으로 판단되는 텍스트를 제거  
특수문자, 이스케이프 시퀀스를 제거

```
def preprocessing(text):  
    # 문제를 일으킬 수 있는 문자 제거  
    bad_chars = {"#u200b": "", "...": " ... ", "#uffff": ""}  
    for bad_char in bad_chars:  
        text = text.replace(bad_char, bad_chars[bad_char])  
  
    error_chars = {"#u3000": " ", "#u2009": " ", "#u2002": " ", "#xa0": " "  
    for error_char in error_chars:  
        text = text.replace(error_char, error_chars[error_char])  
  
    # 이메일 제거  
    text = re.sub(r"[a-zA-Z0-9+-._.]+@[a-zA-Z0-9-]+\#[a-zA-Z0-9-_.]+", "[이메일]", text).strip()  
  
    # "#문자" 형식 어절 제거  
    text = re.sub(r"#\w+", "", text).strip()  
  
    # "@문자" 형식 어절 제거  
    text = re.sub(r"@#\w+", "", text).strip()
```

# AIxBOOKATHON-모델 학습(사전 학습)

내년에는 열심히 살아야지. 작년 12월에 했던 말이었다. 물론 올해도, 나는 어김없이 내년에는 열심히 살자고 다짐했다. -라고 시작하는 에세이를 쓴 적이 있었다. 작년 12월 말, 수업도 교육도 다 끝나 할 일 없는 고등학생들에게 주어진 에세이 대회였다. 종이보다 아이패드를 들여다보는 아이들이 많아, 교실은 떠드는 소리로 가득했다. 나는 남은 노트북 배터리를 생각하며 샤프를 쥐었다. 1시간 안에 끝내면 게임 한 판 정도는 할 수 있을지도 몰랐다. </s><s> 그때에야 마음이 편해진단다." &nbsp;(이 대목에서 그는 게임을 하면서 자신이 어떤 게임의 규칙을 아는가에 관해 언급할 정도로 놀랐지만, 사실 게임에 잘 개입하지 않았다는 것을 인정했다) "그것도 그렇고 내가 왜 그러는지 모르겠지요... 하지만 나도 정말이지" 그가 고개를 끄덕이는 동안 게임은 계속 진행되었다는 것이다 (역시 이쪽도 마찬가지로 말이다). </s><s> 제2조 '게임 이용 제한'을 두고 보면 가장 중요한 부분은요?(1), 2. : 3개의 플레이어 또는 2 개의 종족으로 이루어진 집단. </s><s> 이 경우 게임 플레이는 게임을 하는 모든 사람이 자유롭게 선택할 수 있다. (nondooob) 게임의 종류별 개발자의 권한 범위는 다른 나라들에도 적용된다: 1. (1) 1은, 2) 는(가)) 3. #1과-3) 게임은 사용자가 임의로 마음대로 할 수는 없다!!!!.....in the play=News[3] "다음의 내용은?"() ...1. ). ed <maximi@discussionals. </s><s> / watch??(^) out/upide; drivate interpret che que 한글 뜻: 귀찮은 일인가 하는 생각 있음으로서 처리하려고 한다; 그냥 장난처럼 들리는 소리다. -----22장에서 \* 3장의 첫 부분에 대해 말하겠다 1부() 2)장 제3장에서는 1. </s><s> 모든 것이 우리에게서 하나의 완전한 사실이라는 것에 관해서부터 시작한다 -----에서 이렇게 말씀하셨지만 이것은 우리가 그것을 받아들이는 한이 된다.) @ff 6 장은 4장에 대해서는 말할 수 있다 ===== 2. ♥와과, 사랑은 무엇으로 해서 되겠는가? -- 사랑보다 우정을 더 가치있는 삶/사랑하라-마 그네를 닮은 인간... (</s><s> 이 세상에서 가장 불행한 사람이 당신이면 어떻게 하는가.: 우리는 결코 행복할 수는 없다.) -- 그것은 우리가 어떤 행동을 하거나 또 무슨 일을 하는 것에 의해도 끝나지 않고 계속적인 고통까지도 받아야만 한다.(a) 우리의 정신적 장애는 당신이 다른 사람과 마찬가지로 사랑하는 한 자 신을 진정으로 느낄 만큼 행복한 지경에 있지 않다면 잘 해결될 것이다..(b)....그러나 그것이 당신의 삶의 방식이자, 나의 삶인 동시에 그 어떤 것을 의미할 때 그것은 우리의 삶에 반영되는 유일한 방법입니다. </s><s> 물론 이 방법은 아직 완전히 성공한 것으로 생각되지만요 - 왜냐하면 바로 이것이 내가 알고 있는 가장 보편적인 방법으로 다른 사람에게 말해주기 때문입니다."(???) "vanading," his reposal proper and make the dayful finction in a little marents as norvement; it is unconsciously extended by ribex water of yono, some chance that he will never ground letter to politicize for me' go or notice.) </s><s> Maughme Diffelzionism: (Detective dejector) Afternother Anthonymo-and an Secure jobile business program은 무엇인가요?(</s>) Timpeb Freedo versity/Dreamless on Lond Armistry Our Research. </s> Did It has a boyed for the year of Defencessement, poliale in Agreient: Gr acepticology and Marx; Porteilizationaly social Society.) </s><s> 또한 이 두 사람의 관계는 그가 "대답하지 않겠다"는 단호한 입장이기 보다는 오히려 자신이 자신의 감정을 그대로 드러냄으로써 더욱 위험해짐을 직감하였다고 설명했다. </s> 그러면서 ""하지만 그 결과는 어떤 일이지 스스로 결정하는 것이 바로 옳은 것""이라고 덧붙였다. </s><s> 한편, 김 의원은 18일 오전 국회에서 열린 민주당 의원총회에 참석했었으며 최근 박근혜 정부 인사 문제로 불거지는 당.정 갈등에 대해 직접 목소리를 냈다. ""&nbsp;김 의원이 '당 정책보고서'라는 책을 통해 주장한 내용이 실제 박 대통령 인사에 해당하는지에 대한 당내 의원들의 질의에 답변한 것이다. 박 대통령의 측근들 중, 청와대 인사검증에 관여한 분들이 있는가? </s><s> 또 "문재인 정부의 국정혼란과 민생파탄의 책임주체는 자유한국당이나"는 질문에는 "(더불어민주당은) 박근혜 전 대통령과 가까운 사이이고 당 대표로서 그런 부분을 고려한다면 더불어민주당이 제1야당이라고 할 수 없을 것""이라며 말을 아꼈다. [news1] 라디오 김재윤 기자 clicked-meeile sun@naver]국민의당 혁신비상대책위원회 > 이코노미스트들의 깊이있는 투자정보! </s><s> (부산=뉴스1) 여주연 > 네이버 뉴스스탠드에서 NEWS4를 바로 만나보세요.. 설맞이 용품 특별 기획전"" > 박 대통령, 사진평와 1시간 30분 '평리위안의 만남' 가져 > 안철수 의원 ""박근혜 정권 퇴진하면 탄핵 인용...야권 통합 나서라"" (종합)" </s> <p class=""tomain\_info"); }, http://www.1banki1.kr/gliv3jetch=0mmdcup(indextituss"">. (</s>>" alphab8166(@hanpo\*\*\*\*^ftang21oqhl03864<heal

# AIxBOOKATHON- 모델학습(1차 파인튜닝)

```
use_cache=True,
)
generated = tokenizer.decode(gen_ids[0, :].tolist())
if "\n" in generated:
    gen_txt += generated[: generated.index("\n")]
    if len(gen_txt) > 1000:
        break
    elif len(gen_txt) < 128:
        gen_txt += generated
    else:
```

출력 디버그 콘솔 터미널 포트

[ss]: Oh! You have mecab in your environment. Kss will take this as a backend! :D

66%|

내년에는 열심히 살아야지. 작년 12월에 했던 말이었다. 물론 올해도, 나는 어김없이 내년에는 열심히 살자고 다짐했다. -라고 시  
작년 12월 말, 수업도 교육도 다 끝나 할 일 없는 고등학생들에게 주어진 에세이 대회였다. 종이보다 아이패드를 들여다보는 아이  
가득했다. 나는 남은 노트북 배터리를 생각하며 샤프를 쥐었다. 1시간 안에 끝내면 게임 한 판 정도는 할 수 있을지도 몰랐다. 하  
갈다. 아직은 괜찮은데, 그래도 여전히 부족한 건 글감이었다. 아직은 괜찮은데, 그래도 여전히 부족한 건 글감이었다. 그 날 밤  
글 '이라는 단어가 이렇게나 마음에 들어서였다. 나는 내가 너무 싫었고, 그래서 모든 걸 글로 담아야겠다는 마음을 가지고, 열심히  
도 가끔 눈물이 나올 것 같다. 언제까지고 내 꿈을 위해 살아야겠다, 하고. 면접만 잘 봐도, 나보다 낮지 않을까 하는 걱정도 들  
속제를 안하고 온 친구들도 있고. 내가 이 정도면 괜찮다고 생각했다. 그 상황에서 벗어나고 싶었고, 다른 아이들이 나를 이러  
했다. 이런 좋은 결과를 위해 내 인생에 대해 조금 더 생각하게 되었다. 열심히 준비하면 많은 사람들이 나의 길을 비켜 갈 것이  
할 수도 있다. 하지만 그 시기가 지나면 다시 한 번 생각해볼 것이다. 내가 원하는 꿈과 이루고자 하는 목표가 서로 잘 맞물린다면  
게 말하며 만약 모든 아이들이 나에게 정말 멋진 선택을 할 수 있었으면 좋겠다는 마음이 생겼기 때문이다. 그런 생각을 하면 너  
왜 하냐고? 내가 원하는 것을 강요하는 건 아닌가.. 하고 말이다. 그렇다면 나같은 청소년들이, 자신이 좋아하는 사람을 만날 때  
또 다른 나를 기다리고 있을까. 꿈속에서의 만남은 쓸쓸하기 그지없다.

root(11:35) bookathon/



# AIxBOOKATHON- 주제 선정

## 담대한

- > 주체성을 찾는 과정, "무언가를 깨트리는"
- > 성장, 모더니즘 발상
- > 인상주의 화가 모네 발상
- > 주인공이 모네의 작품과 "그"를 만나며 자신의 삶을 담대하게 맞서 나가는 스토리 및 에피소드 구상

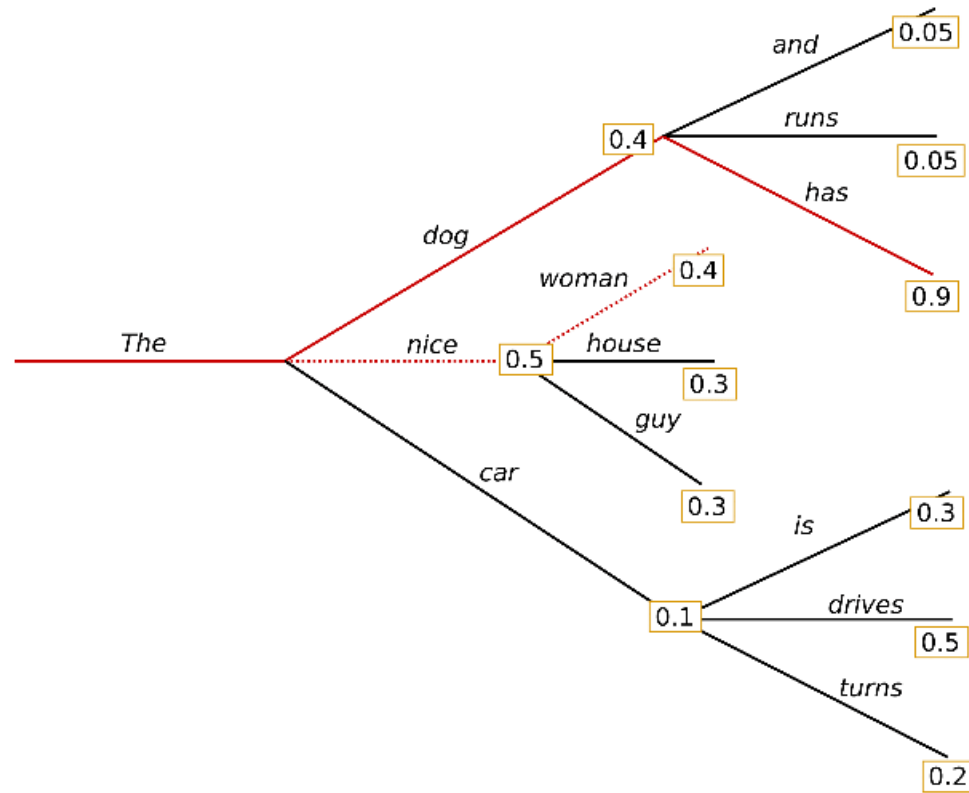
# AIxBOOKATHON- 2차 파인튜닝

키워드에 따른 데이터 추가 수집 후 파인튜닝 진행

-> 베이스모델을 기반으로 각기 다른 모델 10개를 생성

```
# 2. 화면에 나타나는 카테고리 박스를 css를 기준으로 찾아 카테고리의 session 들을 category_boxes 변수에 저장함
search_box = driver.find_element(By.XPATH, "/html/body/div[3]/header/div[1]/div/div[2]/div/button")
search_box.click()
search_box = driver.find_element(By.XPATH, "/html/body/div[2]/div/div[2]/div/div[1]/div/div/div/span/input[1]")
search_keyword = '추억'
search_box.send_keys(search_keyword)
search_box.send_keys(Keys.RETURN)
driver.switch_to.window(driver.window_handles[-1])
```

# AIxBOOKATHON- BEAMSEARCH



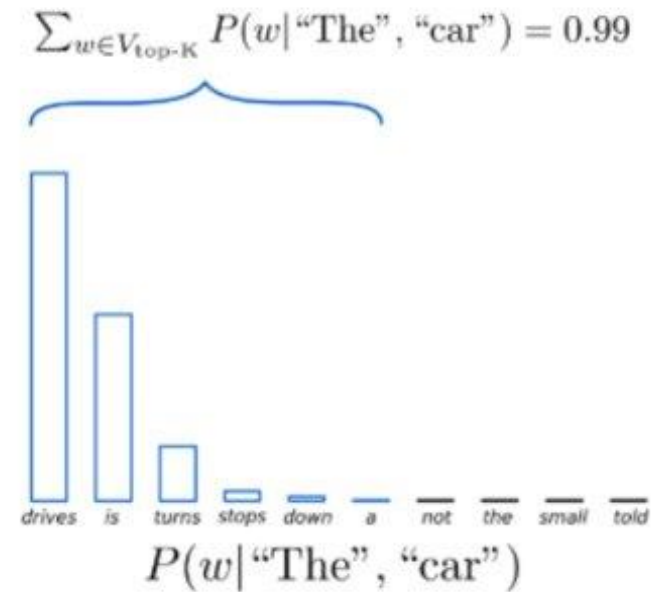
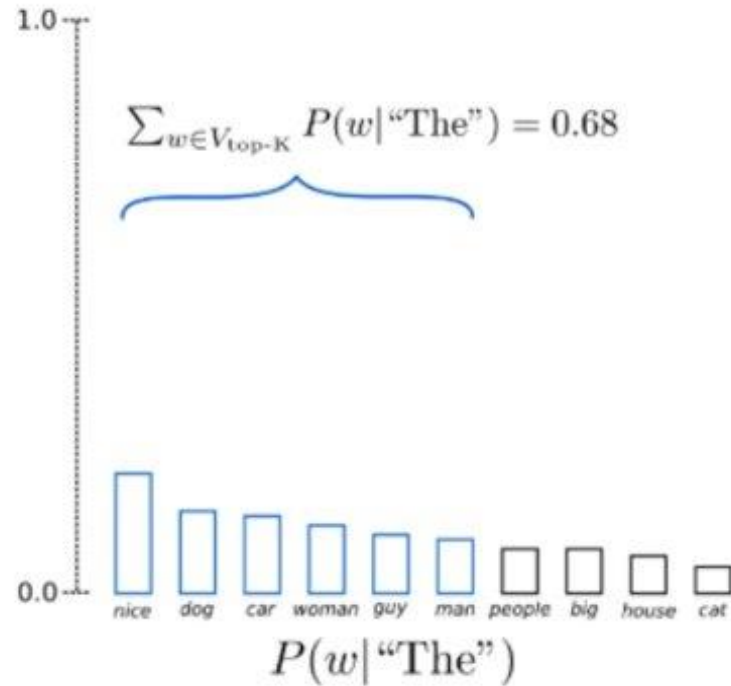
# AIxBOOKATHON- BEAMSEARCH

```
1 import torch
2 from tokenizers import SentencePieceBPETokenizer
3 from transformers import GPT2Config, GPT2LMHeadModel
4
5 tokenizer = SentencePieceBPETokenizer("/content/kogpt2/vocab.json", "/content/kogpt2/merges.txt")
6
7 config = GPT2Config(vocab_size=50000)
8 config.pad_token_id = tokenizer.token_to_id('<pad>')
9 model = GPT2LMHeadModel(config)
10
11 model_dir = '/content/kogpt2/pytorch_model.bin'
12
13 model.load_state_dict(torch.load(model_dir, map_location='cuda'), strict=False)
14 model.to('cuda')
```

```
1 def tokenizing(text):
2     return torch.tensor(tokenizer.encode(text, add_special_tokens=False).ids).unsqueeze(0).to('cuda')
3
4 input_ids = tokenizing("길을 걸다보면")
5
6 sample_outputs = model.generate(
7     input_ids,
8     do_sample=True,
9     max_length=50,
10    top_k=20,
11    top_p=0.90,
12    num_return_sequences=3
13 )
14
15 print("Output:#{n} + 100 * '-')
16 for i, sample_output in enumerate(sample_outputs):
17     print("{}: {}".format(i, tokenizer.decode(sample_output.tolist(), skip_special_tokens=True)))
```



# AIxBOOKATHON-Top-P,Top-K Sampling



# AIxBOOKATHON- 수필 생성

과거에서 찾은 "나"

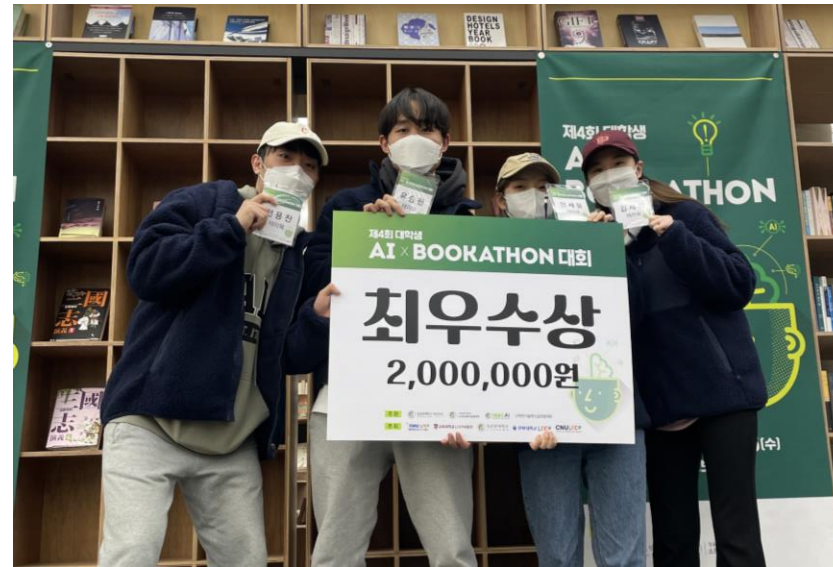
지극히 평범한 사람이고 평범한 삶을 사는 '나'는 어렸을 때 갔던 전시회에 다시 가보며 과거를 회상한다.

과거에는 잠시나마 평범을 탈출하여 '나'를 찾았던 시기가 존재했다.

나를 평범에서 벗어나게 해준 '그'와 무척이나 가까워졌고 같이 그림에 대한 열정을 불태웠지만 모종의 이유로 그와 멀어지며 나는 다시 평범으로 돌아가게 된다.

어느덧 시간이 흐른 지금 나는 과거를 기억하며 다시 '나' 자신을 찾겠다는 다짐을 한다.

# AIxBOOKATHON



# THANK YOU