

# Deep Learning Week 8

## **Further topics about deep learning**

---

Hanyang Artificial Intelligence Group



## **지금까지 배웠던 내용들**

**딥 러닝이란 무엇인가? 딥 러닝으로 무엇을 할 수 있을까?**

**인공지능에서 모델이란 무엇인가? 딥 러닝 모델은 어떻게 학습되는가?**

**딥 러닝 모델의 성능을 어떻게 평가할 수 있을까?**

**어떻게 시각적인 정보(이미지, 비디오)를 딥 러닝으로 다룰 수 있을까?**

**어떻게 사람이 사용하는 언어를 컴퓨터가 이해하게 할 수 있을까?**

## 오늘의 Topic

# Generative AI with deep learning

- 컴퓨터 비전과 자연어 처리 등 다양한 형태의 생성 모델

## Large Language Model

- ChatGPT의 개발이 가능하도록 한 초거대 언어 모델

## 딥 러닝을 활용한 서비스

- 클로바 노트로 살펴보는 딥 러닝 모델의 실제 서비스 적용 방식

# What is Generative AI?

## 새로운 데이터를 생성할 수 있는 AI

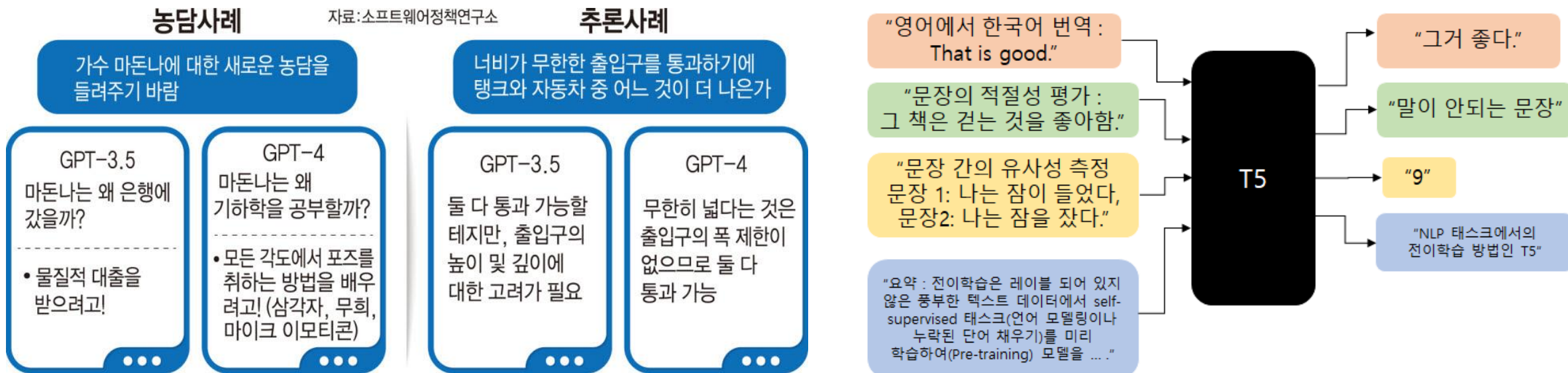
- 이미지, 텍스트 등 입력된 형태의 데이터를 이해해서 유용한 정보를 추출하는 기존 모델과 달리, 새로운 데이터를 생성할 수 있는 AI가 급격하게 발전하는 중
- 텍스트 -> 텍스트, 이미지 -> 이미지 처럼 기존 형식과 같은 데이터를 생성하거나, 이미지 -> 텍스트, 오디오 -> 텍스트 등 입력과 다른 형태의 데이터를 생성할 수 있는 모델도 존재!
- 인간의 고유 영역이라고 여겨졌던 창의적인 활동을 인공지능이 수행할 수 있게 됨



# Generative AI: text-to-text

## 입력된 텍스트를 다른 형태로 변환하거나, 이어서 써주는 AI

- GPT: 입력된 텍스트에 적절하게 이어질만 한 텍스트를 생성해주어, 소설 쓰기, 대화 생성 등 다양한 task에 활용될 수 있고, 최근엔 instruction tuning으로 자연어로 된 명령을 처리할 수 있게 됨
- T5: 입력 텍스트를 특정한 의미를 가지는 vector로 인코딩하고, 이를 바탕으로 새로운 텍스트를 출력하는 구조로 되어있어, 자연어 이해와 생성 둘 다 뛰어난 성능을 보여주며 효율적임

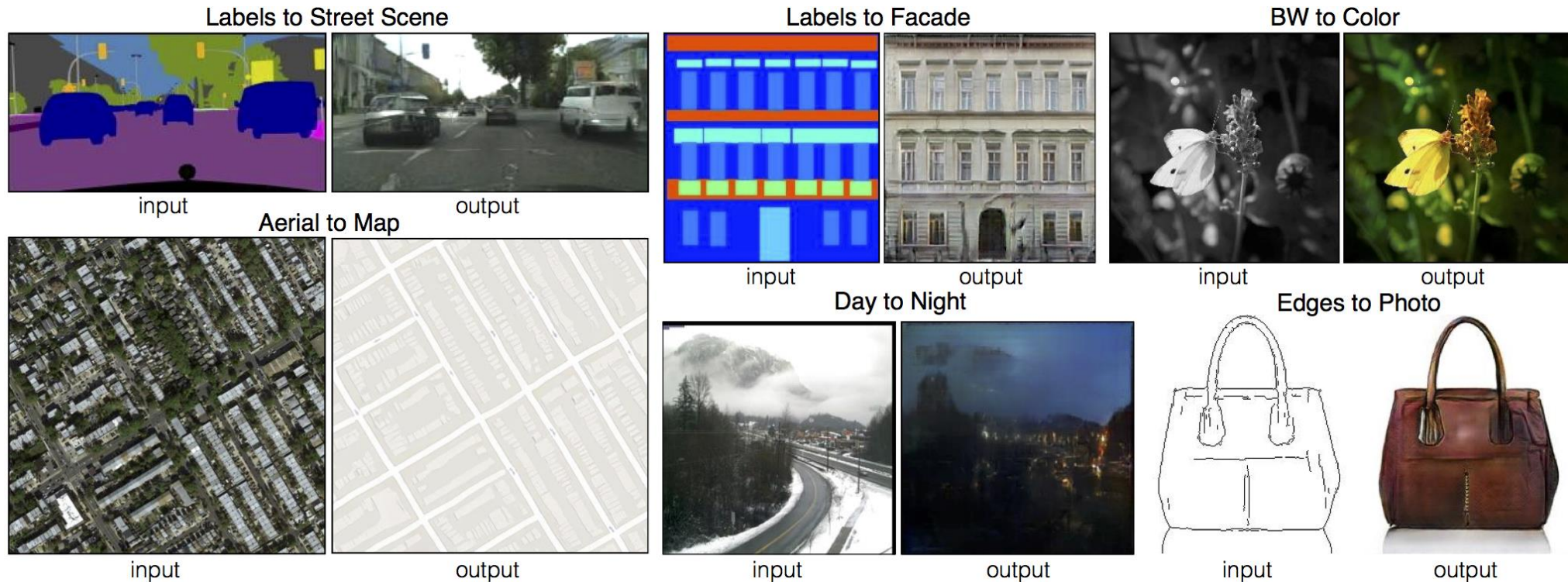




# Generative AI: image-to-image

## 입력된 이미지를 바탕으로 새로운 이미지를 생성하는 AI

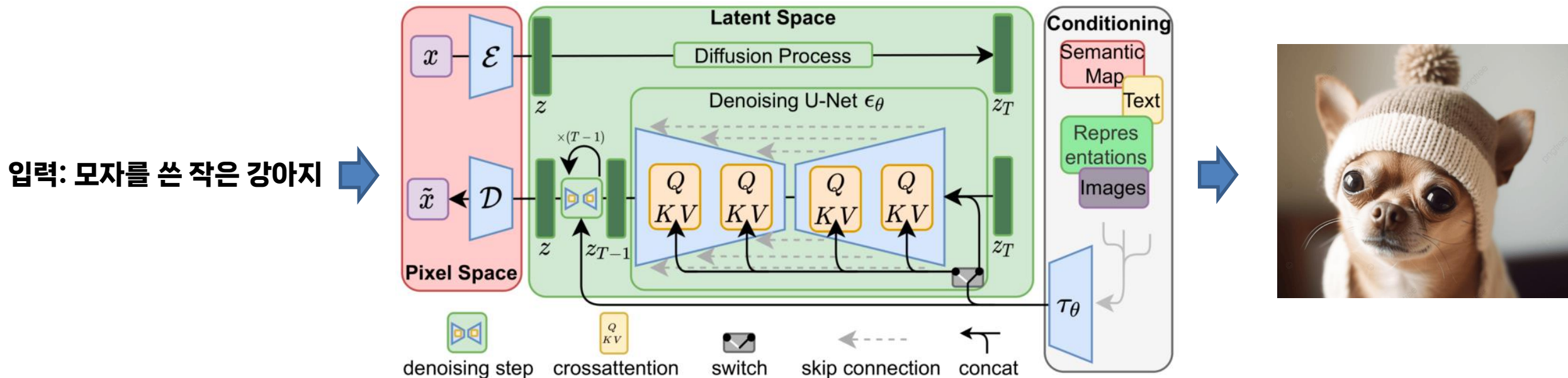
- 입력 이미지의 색상, 스타일 등을 원하는 형태로 변환하거나, 스케치를 기반으로 더 복잡한 이미지를 생성하는 등 다양한 이미지 생성이 가능한 모델
- Pix2Pix(StyleGAN 기반) 등 굉장히 뛰어난 성능의 모델들이 공개됨



# Generative AI: text-to-image

## 사람처럼 요구사항을 읽고 그림을 그려주는 AI

- 모델의 입력으로 텍스트 데이터를 받아, 입력된 텍스트의 특징이 반영된 이미지를 생성해주는 모델
- Dall-E, Stable Diffusion 등 최신 모델은 굉장히 좋은 품질의 이미지를 생성할 수 있어 다양한 산업 분야에 적용될 수 있음



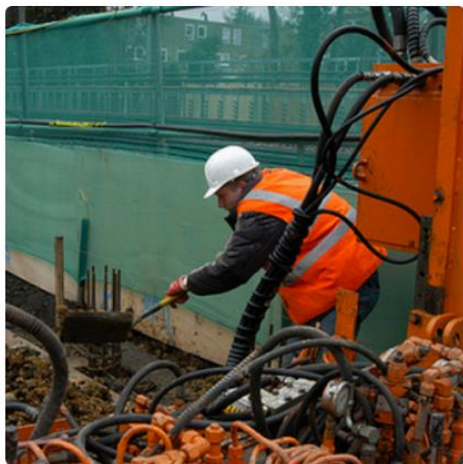
# Generative AI: image-to-text

## 그림에 대한 정보를 텍스트로 변환해주는 AI

- 입력된 이미지의 특징이나 등장 요소 등을 텍스트로 설명할 수 있는 모델
- 시각 장애인들을 위한 이미지 캡션 생성이나, 이미지를 이해할 수 있는 챗봇 기술에 활용될 수 있음



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."

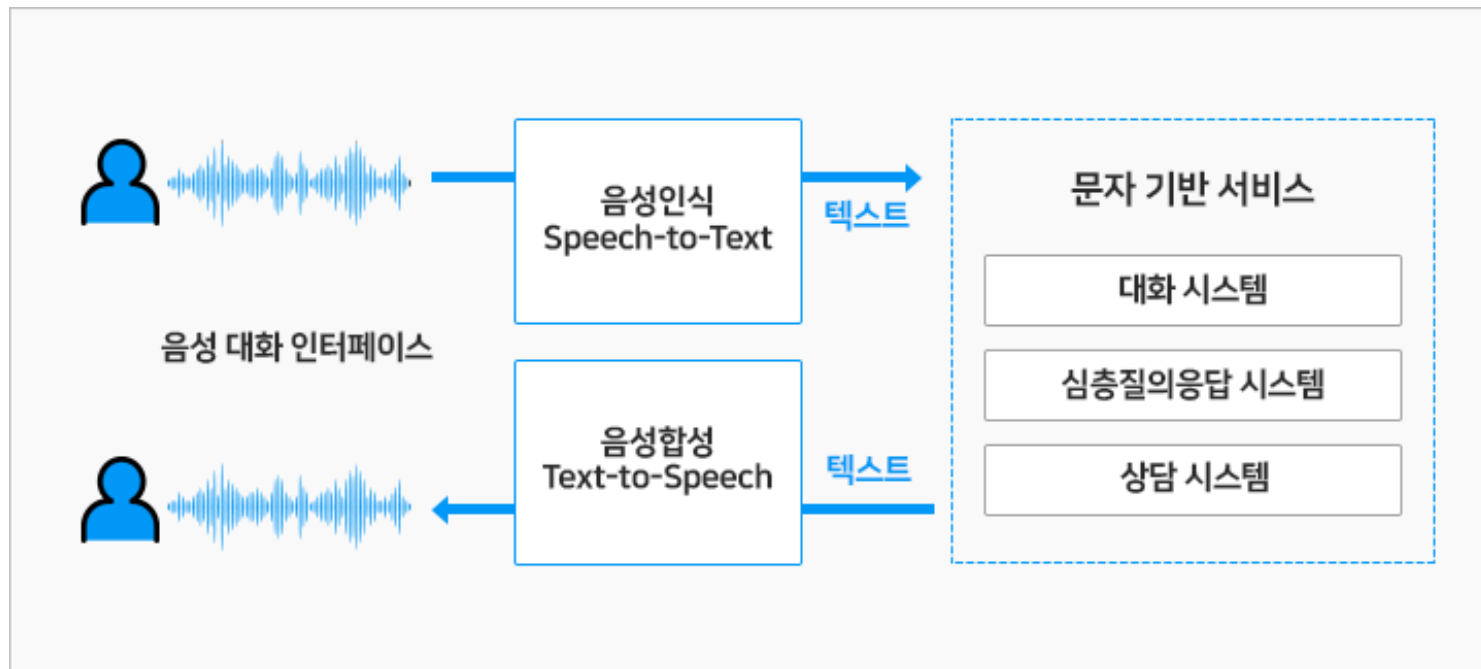




# Generative AI: speech-to-text & text-to-speech

## 음성 인식 & 음성 합성 모델

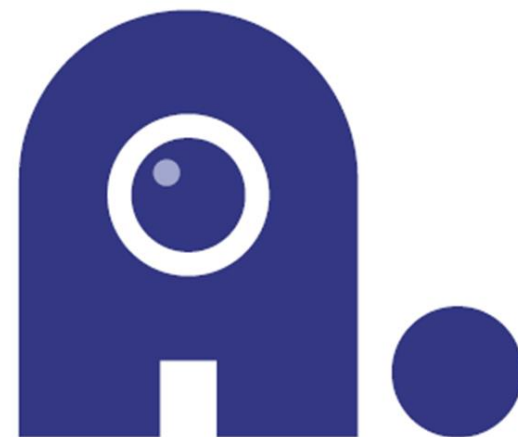
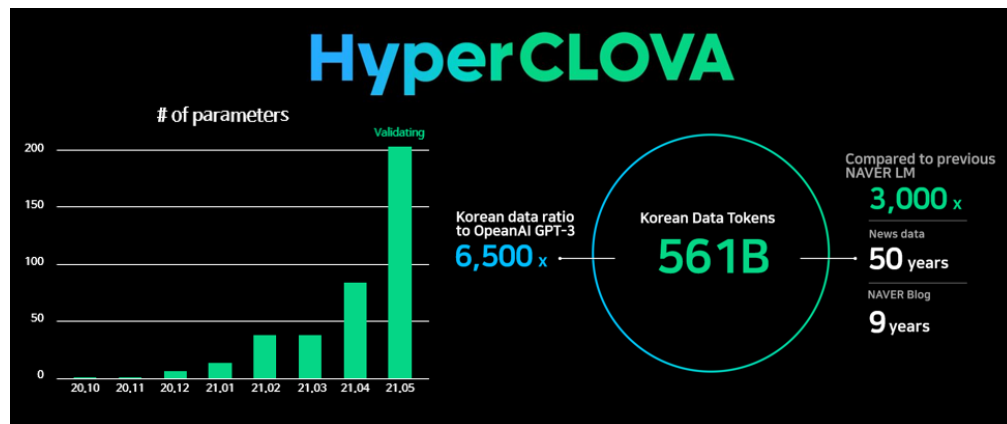
- 인간의 대화 음성을 인식하여 텍스트로 전사(transcript)하거나, 입력된 텍스트를 사람이 읽은 것 처럼 합성하여 음성을 생성할 수 있는 모델
- 자연어 처리 기술과 함께 사용하여 대화하듯이 컴퓨터와 음성으로 소통할 수 있다!



# LLM(Large Language Model)

## 초거대 언어 모델이 가져온 미래

- GPT-3 등 파라미터 개수가 수십~수백 Billion 이상인 모델
- 기존 모델과 구조적으로 큰 차이는 없지만, 사이즈와 데이터를 엄청나게 많이 늘린 것 만으로도 Few-shot learning, Meta learning 등 기존 모델과 차별화된 다양한 언어 능력을 갖추게 됨
- 언어 데이터로 pretraining을 거친 후 특정 task를 수행하도록 finetuning되었던 기존 모델과 달리, 단일 모델로 모든 task를 추가 학습 없이 수행할 수 있는 능력을 가짐



# LLM의 문제점

## 초거대 모델을 위한 고성능 분산 컴퓨팅 시스템

- 지나치게 큰 모델 사이즈로 인해 모델 학습, 추론 과정을 기존 컴퓨팅 환경에서 진행하기 어려움
- 여러 GPU를 동시에 병렬로 가동하여 거대 모델을 분산 처리하는 기술이 필요함
- 엄청나게 많은 자원이 필요하기 때문에 개인 사용자는 학습 또는 추론이 매우 어려움

## 초거대 모델 학습을 위한 대규모 데이터셋

- 대규모 언어 모델을 학습시키기 위해서는 모델 사이즈에 걸맞는 대규모 데이터셋이 필요함
- 사용자가 많고 데이터가 많이 축적된 언어(영어 등)와 달리 사용자가 적은 소수 언어의 경우 데이터의 규모와 품질이 부족하여 상대적으로 모델의 성능이 부족함

**대규모 자본과 시설, 데이터셋이 없는 우리는  
어떻게 해야 할까?!!**



# 누구나 쉽게 LLM을 사용할 수 있도록 하자!

## Open-source LLM

- OpenAI, Google, 네이버 등 기업에서 출시한 언어 모델은 대부분 모델 파라미터를 공개하지 않음
- Llama, Polyglot 등의 open-source 모델이 공개되어 LLM에 대한 접근성이 크게 향상됨
- 모델 사이즈는 기업용에 비해 상대적으로 작지만, 성능은 결코 뒤떨어지지 않아 매우 효율적임

## 자원을 적게 사용하도록 하는 효율적인 딥러닝 기술들

- 대규모 언어 모델을 원하는 특정 분야에 특화되도록 학습시키려면 엄청나게 많은 자원이 필요함
- LoRA 등 모델을 학습 비용을 크게 절약하면서도 성능 저하가 거의 없는 학습 방식
- 기존 모델 파라미터의 precision을 낮춰 모델의 사이즈를 줄이는 quantization 방식

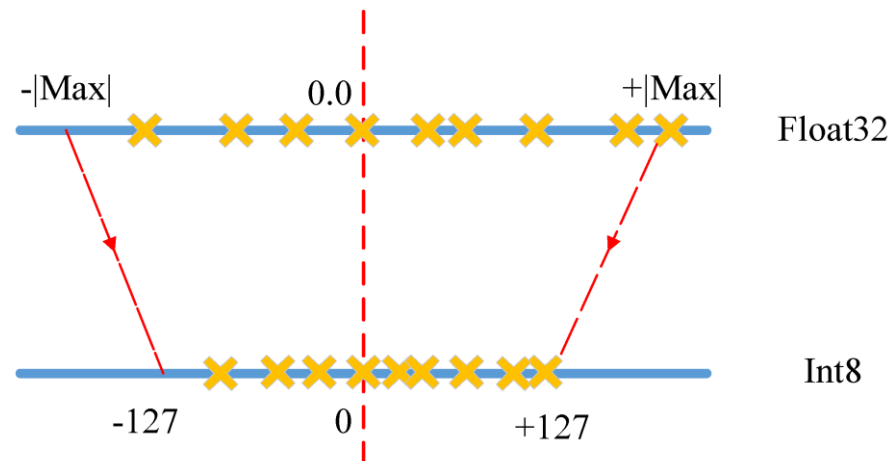


# 누구나 쉽게 LLM을 사용할 수 있도록 하자!

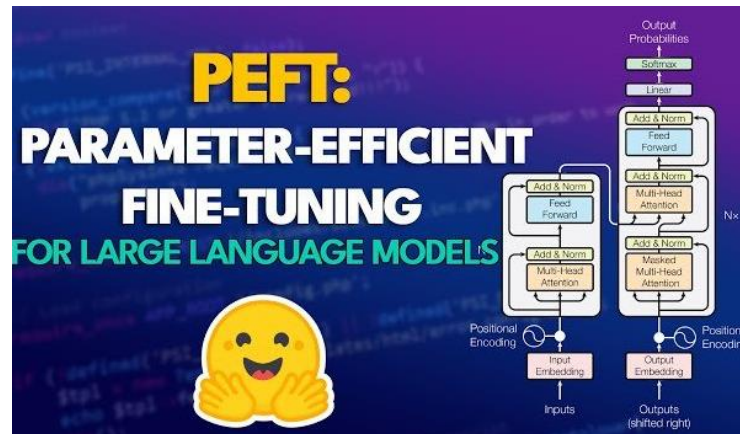
## Open-source LLM



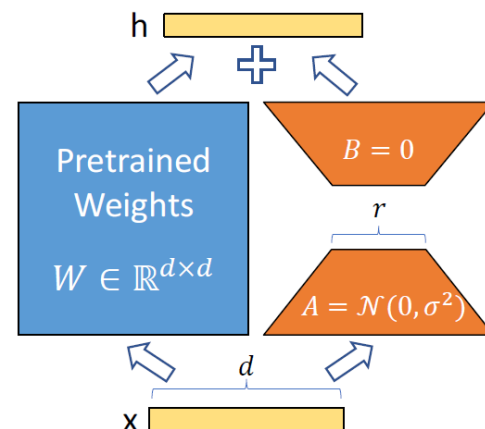
## Model quantization



## Parameter-efficient finetuning



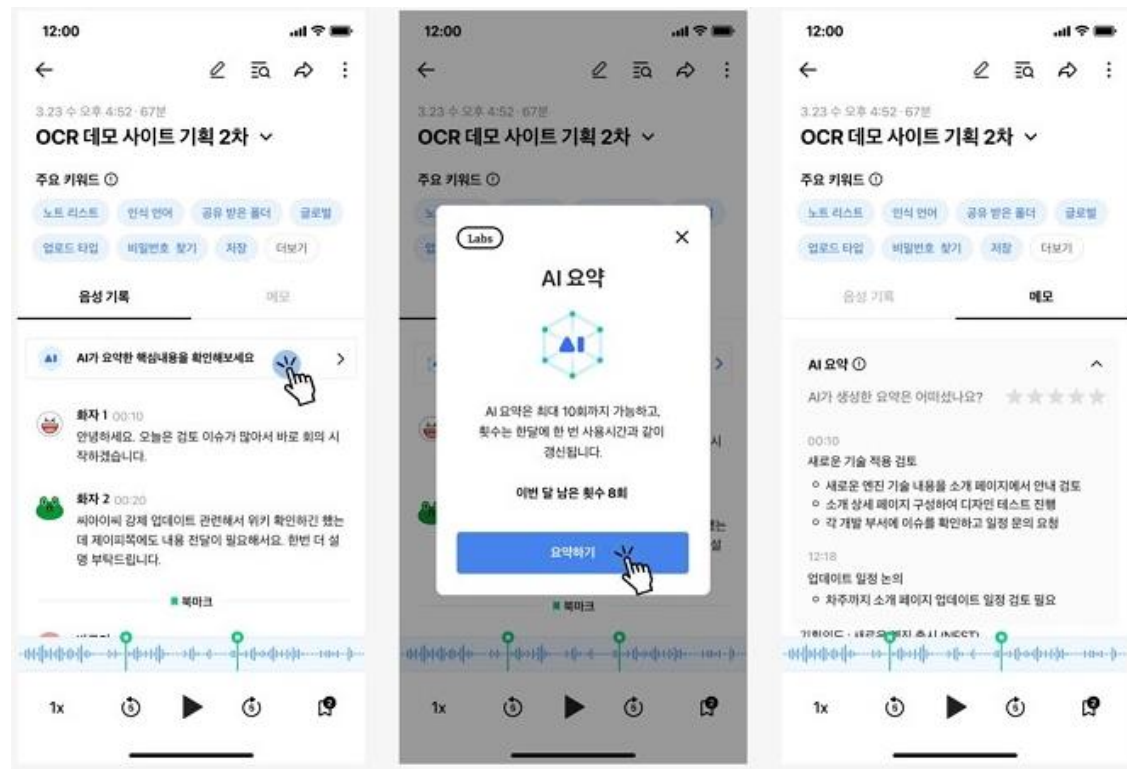
## LoRA(Low Rank Adaptation)



# 딥 러닝을 활용한 서비스

## 클로바노트를 통해 보는 딥 러닝 기술의 서비스 적용 방식

- 일반 사용자가 쓸 수 있는 서비스 중 딥러닝 기술이 성공적으로 적용된 사례
- 네이버에서 자체 개발한 음성 인식 엔진과 초거대 언어 모델(HyperCLOVA)이 사용됨
- 음성을 녹음하는 것 만으로도 텍스트로 변환된 기록과 주요 키워드, 자동으로 요약한 결과를 얻을 수 있음



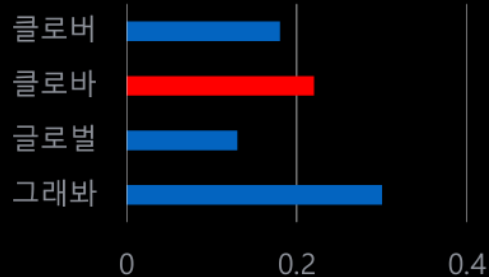
# 클로바 노트의 음성인식 과정

## 대화 음성 기록을 녹취록 텍스트로

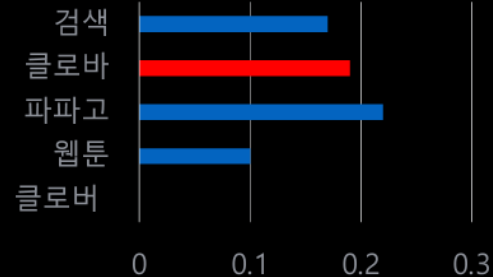
- 음성 인식 모델을 사용하여 구간별 발화자 구분 및 발화 내용 텍스트 변환
- 텍스트 변환 과정에서 더 정확한 결과를 위해 언어 모델의 예측값을 함께 사용



사용자의 마음을 이해하는 네이버 OOO



$$P_{AM}(Y_t|X, Y_{1:t-1})$$



$$P_{LM}(Y_t|Y_{1:t-1})$$



# 클로바 노트의 음성인식 과정

## 참고(DEVIEW 2023)

[https://devview.kr/data/devview/session/attach/\[144\]%ED%81%B4%EB%A1%9C%EB%B0%94%EB%85%B8%ED%8A%B8%EC%9D%98+%EB%91%90%EB%87%8C+NAV%ER%EC%9D%98+End-to-End+%EC%9D%8C%EC%84%B1%EC%9D%B8%EC%8B%9D+%EC%97%94%EC%A7%84+%EC%86%8C%EA%B0%9C.pdf](https://devview.kr/data/devview/session/attach/[144]%ED%81%B4%EB%A1%9C%EB%B0%94%EB%85%B8%ED%8A%B8%EC%9D%98+%EB%91%90%EB%87%8C+NAV%ER%EC%9D%98+End-to-End+%EC%9D%8C%EC%84%B1%EC%9D%B8%EC%8B%9D+%EC%97%94%EC%A7%84+%EC%86%8C%EA%B0%9C.pdf)

### 음성 기록

점유율 편집



참석자 1 00:00

티저로 바꿔 인티저 값을



참석자 2 00:03

콘 타이



참석자 1 00:04

플로팅 포인트 32비트로 바꿔주고 그 값을 또 이제 다른 레이어로 보낼 때는 31it를 8pit it로 다시 보내주는 과정을 리니어 레이어 안에 이제 추가를 해줘서 구현한 것으로 알고 있습니다.



참석자 3 00:20

설명이 전혀 안 되는데 다이내믹한 부분이 어디인 거예요.  
다이내믹하게 레지 그런 얘기했잖아



참석자 1 00:24

예 맞습니다. 그게



참석자 2 00:25

그거는 라이브러리에서 기본 값으로 정해져 있는 걸로

### 참석자 대화 점유율



참석자 3

52.5%

참석자 2

43.1%

참석자 1

3.7%

참석자 4

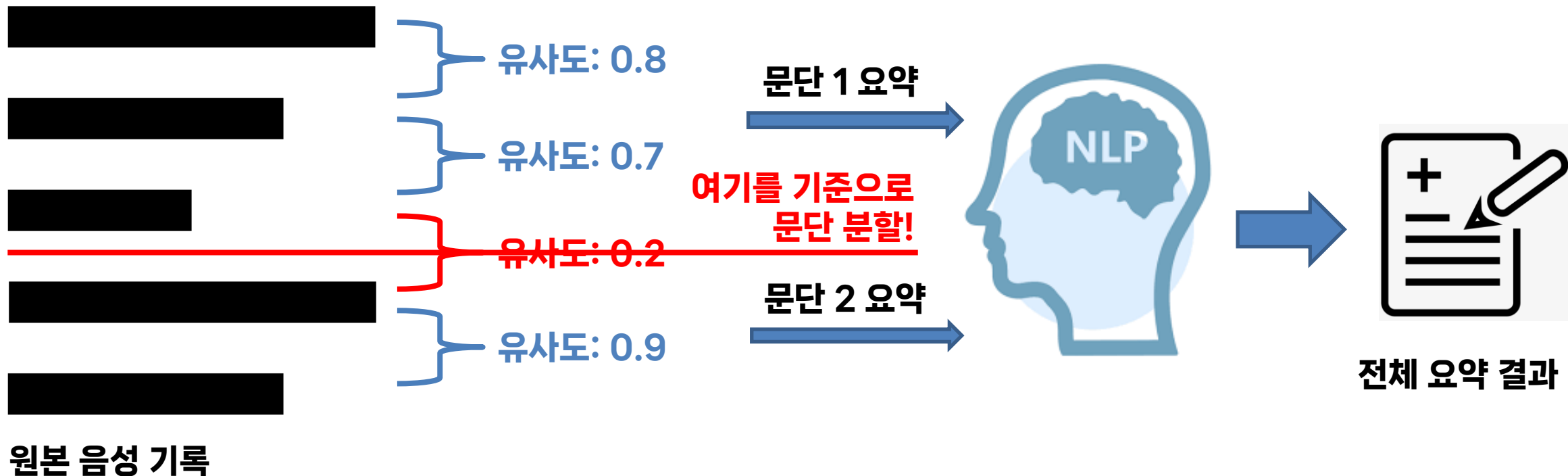
0.7%



# 클로바 노트의 자동 요약 과정

## 요약을 위한 의미 기반 구간 나누기

- 원문 요약을 위해 원본 입력을 적당한 길이로 나눠야 할 필요 발생
- 각 문장 사이의 유사도를 계산하여, score가 낮은 부분을 기준으로 분할



# 클로바 노트의 음성인식 과정

## • 참고(DEVIEW 2023)

[https://devview.kr/data/devview/session/attach/\[245\]HyperCLOVA%20%EC%9A%94%EC%95%BD%20%EC%84%9C%EB%B9%84%EC%8A%A4%20Long-Form%20Open-Domain%20Dialogue%20Summarization\\_0228.pdf](https://devview.kr/data/devview/session/attach/[245]HyperCLOVA%20%EC%9A%94%EC%95%BD%20%EC%84%9C%EB%B9%84%EC%8A%A4%20Long-Form%20Open-Domain%20Dialogue%20Summarization_0228.pdf)

### 음성 기록

🕒 점유율

편집

이성진 02:23

잘하는 그런 모델이라고 할 수가 있습니다.

그리고 또 이거는 중요한 얘기는 아니지만 이 버트라는 이름이 바이디렉셔널 인코더, 리프레젠테이션, 프렐 트랜스포머 이거 줄인 말인데 세서미 스트레이트라는 애니메이션 캐릭터 이름이에요.

그래서 이 버트가 나오기 전에 같은 애니메이션에 있는 모라는 캐릭터가 있어요.

엘모라는 이름으로 누군가 이제 어거지로 끼어맞춰서 언어 모델 이름을 지었더니 그게 이제 되게 웃겼는지 그다음에 버트도 나오고 심슨에 나오는 바트도 나오고 같은 세서미 스트레이트에 등장하는 롱 버드라는 캐릭터도 나오고 되게

이성진 02:57

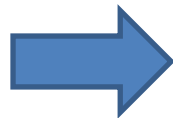
아무튼 좀 웃기게 짓는 게 그게 트렌드가 됐는데 저도 tv에서 본 적은 없어요.

저도 tv에서 본 적은 없는 옛날 캐릭터인데 아무튼 캐릭터 이름을 가지고 모델을 짓는 게 유행이 됐다. 이런 얘기를 할 수가 있습니다. 그래서 이제 아까 언어 모델이라고 얘기를 했는데 언어 모델이 구체적으로 뭐냐 이 언어 모델은 우리가 이제 인공지능 전체에서 제 얘기를 할 때 모델은 어떤 function이라고 얘기를 했었죠

● ● ●

이성진 37:07

교류를 하는 목적에서 뭔가 서로 우리는 어떤 거 하고 있나 그런 것들을 좀 발표하고 또 평소에 내가 공부하기 어려웠던 다른 분야의 지식들을 또 볼 수 있는 그런 기회니까 잘 참여하시는 분들은 오셔서 또 재밌게 잘 보시고 또 가면 좋을 것 같아요.



00:00 요약 구간

자연어 처리의 목표

- 자연어 처리의 가장 중요한 목표는 자연어 텍스트로 된 데이터를 컴퓨터가 이해할 수 있는 표현 벡터 형태로 변환하는 것임
- 자연어 처리의 가장 핵심적인 목표 중에 하나는 단어들 사이의 관계를 벡터 계산으로 표현하는 것임
- 시퀀스 2 시퀀스 모델은 인풋으로 어떤 텍스트를 받아서 컴퓨터가 이해할 수 있는 유용한 정보로 표현을 한 다음 그걸 이용해서 다시 출력을 하는 모델임
- 버트와 gpt는 트랜스포머에서 인코더 구조만 따온 게 버트고 디코더 구조만 따온 게 gpt임

03:21

언어 모델의 개념

- 언어 모델은 주어진 문장에 대한 확률을 계산할 수 있는 함수라고 생각을 할 수 있음
- 언어 모델이 얼마나 자연스럽고 적절한가에 따라서 확률을 계산할 수도 있음
- 언어 모델은 문장 뒤에 이어질 가장 자연스러운 텍스트에 대한 확률을 구하는 모델이라고 생각을 하면 됨

07:07

마스크드 랭귀지 모델링

- 마스크드 랭귀지 모델링은 주어진 입력에서 특정 부분을 마스킹을 해서 지워서 안 보이게 했을 때 그 부분에 들어갈 단어들의 확률 분포를 리턴하도록 학습된 모델임
- 마스크드 랭귀지 모델링은 이미 학습된 모델의 능력을 재사용하는 과정임
- 마스크드 랭귀지 모델링을 엄청나게 많은 텍스트에 대해서 사전 학습을 시킴
- 프리체이닝이 완료된 모델을 내가 원하는 테스트에 파인튜닝을 하는 과정도 거치게 됨

# 클로바 노트의 소제목 생성 과정

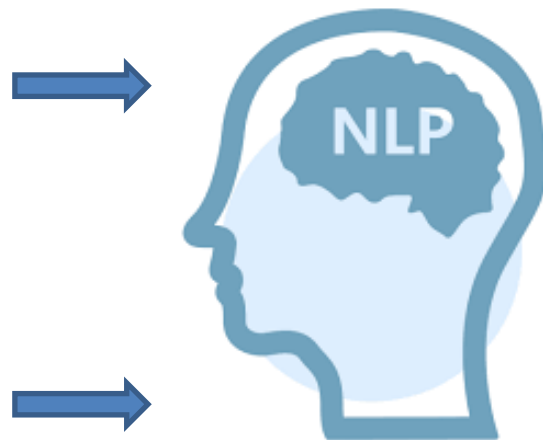
## 각 구간별 요약문 기반 제목 생성

- 원문을 분할한 각 구간별 요약문을 한 번 더 요약하여 간단하게 핵심을 확인할 수 있도록 제목 생성

### 원본 요약문

- 언어 모델은 주어진 문장에 대한 확률을 계산할 수 있는 함수라고 생각할 수 있음
- 언어 모델이 얼마나 자연스럽게 적절한가에 따라서 확률을 계산할 수도 있음
- 언어 모델은 문장 뒤에 이어질 가장 자연스러운 텍스트에 대한 확률을 구하는 모델이라고 생각할 수 있음

- gpt 모델은 입력된 텍스트에 가장 잘 어울리는 뒤에 이어질 단어들을 찾는 조건부 확률로 표현할 수 있음
- gpt 모델은 입력된 텍스트에 가장 어울리는 단어들을 보고 확률이 높은 단어를 골라 다음에 이어질 만한 단어들에 대한 확률을 또 구할 수 있음



### 제목 생성 결과

언어 모델의 개념

GPT 모델의  
조건부 확률

# 클로바 노트의 소제목 생성 과정

## 딥러닝 스터디 7주차: BERT와 GPT

전체 노트 · 5.18 목 오후 7:14(5.26 금 수정) · 37분 35초

이성진

음성 기록 + 메모 보기 ▾

### 음성 기록

점유율 편집

그래서 이 센텐스 임베딩이라는 거는 임베딩이라는 개념 자체를 많이 접하지 않으신 분들이면 이해가 잘 안 될 수 있어요.  
그래서 여기 링크에 굉장히 좀 이제 좋은 소스가 있는데 이걸 잘 들어가 보시면 설명이 굉장히 잘 돼 있어요.  
이제 위클리 lp라고 하는 칼럼을 쓰시는 분이 설명을 해둔 건데 이게 이미 제가 쓴 거 아니에요

### 이성진 36:07

그래서 이제 그래서 거리가 두 개에서 벡터 사이의 거리를 어떻게 구하고 그래서 이 벡터를 이용하면 비슷한 단어들이 얼마나 가까이 있는지에 대해서 계산할 수 있고 그런 거에 대한 얘기가 있으니까 궁금하신 분들은 찾아보시면 좋을 것 같고 이거 말고도 되게 좋은 내용들이 많아요.  
그래서 자연어 처리에 관심이 있다 하시는 분들은 그 글이 그렇게 어렵지 않게 써져 있으니까 한번 쭉 읽어보시면은 전체적으로 좀 잘 이해를 하시는데

### 이성진 36:32

도움이 많이 될 것 같습니다. 그러면 그래서 이 코드는 조만간 제가 좀 올려드리도록 하겠고요.  
오늘은 여기까지 하고 공부하시느라 고생 많으셨고 다음 주에 축제니까 또 잘 즐기시고 아마 축제 끝나면 그다음에 아마 마지막인가 그다음 다음이 마지막인가 그럴 거예요.  
또 우리 기말고사 기간이 있어서 그리고 또 이번 주 토요일에 저희 고려대학교 아이쿠랑 같이 연합 컨퍼런스를 진행을 하는데 그냥 뭐 엄청 대단한 그런 걸 하는 게 아니라 두 동아리 사이에서 서로 침묵

### 이성진 37:07

교류를 하는 목적에서 뭔가 서로 우리는 어떤 거 하고 있나 그런 것들을 좀 발표하고 또 평소에 내가 공부하기 어려웠던 다른 분야의 지식들을 또 볼 수 있는 그런 기회니까 잘 참여하시는 분들은 오셔서 또 재밌게 잘 보시고 또 가면 좋을 것 같아요.  
고생하셨습니다. 듀프리 가시는 분들 말고는 다 이제 가서도 돼요.

### 메모

더 똑똑한 AI 요약으로 성장할게요 ★★★★★

00:00

#### 자연어 처리의 목표

- 자연어 처리의 가장 중요한 목표는 자연어 텍스트로 된 데이터를 컴퓨터가 이해할 수 있는 표현 벡터 형태로 변환하는 것임
- 자연어 처리의 가장 핵심적인 목표 중에 하나는 단어들 사이의 관계를 벡터 계산으로 표현하는 것임
- 시퀀스 2 시퀀스 모델은 인풋으로 어떤 텍스트를 받아서 컴퓨터가 이해할 수 있는 유용한 정보로 표현을 한 다음 그걸 이용해서 다시 출력을 하는 모델임
- 버트와 gpt는 트랜스포머에서 인코더 구조만 따온 게 버트고 디코더 구조만 따온 게 gpt임

03:21

#### 언어 모델의 개념

- 언어 모델은 주어진 문장에 대한 확률을 계산할 수 있는 함수라고 생각을 할 수 있음
- 언어 모델이 얼마나 자연스럽게 적절한가에 따라서 확률을 계산할 수도 있음
- 언어 모델은 문장 뒤에 이어질 가장 자연스러운 텍스트에 대한 확률을 구하는 모델이라고 생각을 하면 됨

07:07

#### 마스크드 랭귀지 모델링

- 마스크드 랭귀지 모델링은 주어진 입력에서 특정 부분을 마스킹을 해서 지워서 안 보이게 했을 때 그 부분에 들어갈 단어들의 확률 분포를 리턴하도록 학습된 모델임
- 마스크드 랭귀지 모델링은 이미 학습된 모델의 능력을 재사용하는 과정임
- 마스크드 랭귀지 모델링을 엄청나게 많은 텍스트에 대해서 사전 학습을 시킴
- 프리체이닝이 완료된 모델을 내가 원하는 태스크에 파인튜닝을 하는 과정도 거치게 됨



# 1학기 활동 마무리 & 추후 계획

## 종강 기념 파티 - 6/24

- 파티룸에서 1박 2일 진행 또는 당일 저녁 식사(예정)

## 방학 프로젝트

- 1학기 활동 정리 및 2학기 프로젝트 활동 대비
- 논문 구현 스터디 및 선형대수 스터디 운영 예정 (희망자)
- AI 앱 개발 프로젝트(필수 참여)

## 2학기 활동

- 팀별 자유 주제 프로젝트 진행
- 프로젝트 진행 사항 공유 및 발표
- 각종 세미나 및 연합 컨퍼런스



# With HAI, Fly High

---

Hanyang Artificial  
Intelligence