

# 10. NLU 최신 동향

RoBERTa, ALBERT, ...

## RoBERTa

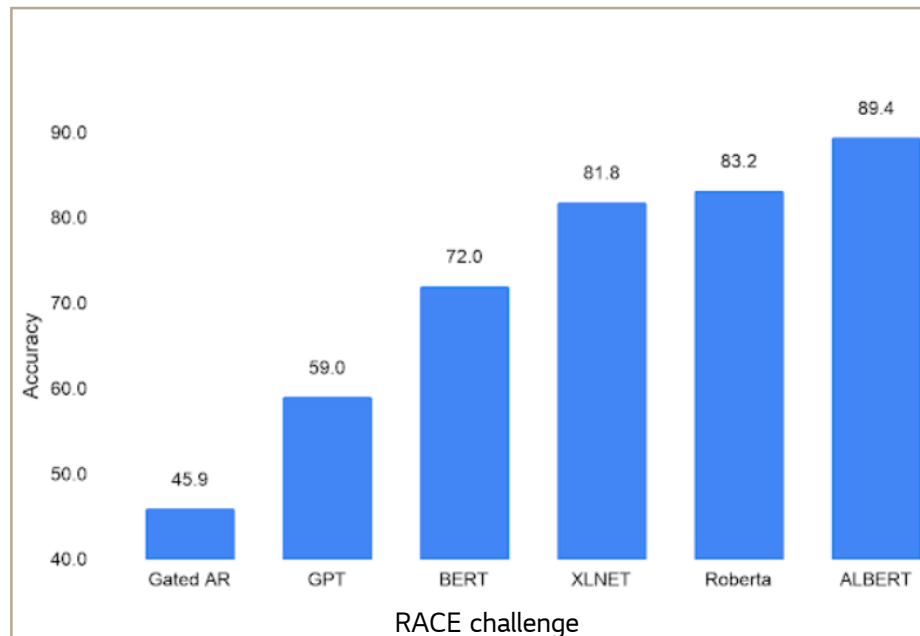
	MNLI	QNLI	QQP	RTE	SST	MRPC	CoLA	STS	WNLI	Avg
<i>Single-task single models on dev</i>										
BERT <sub>LARGE</sub>	86.6/-	92.3	91.3	70.4	93.2	88.0	60.6	90.0	-	-
XLNet <sub>LARGE</sub>	89.8/-	93.9	91.8	83.8	95.6	89.2	63.6	91.8	-	-
RoBERTa	<b>90.2/90.2</b>	<b>94.7</b>	<b>92.2</b>	<b>86.6</b>	<b>96.4</b>	<b>90.9</b>	<b>68.0</b>	<b>92.4</b>	<b>91.3</b>	-
<i>Ensembles on test (from leaderboard as of July 25, 2019)</i>										
ALICE	88.2/87.9	95.7	<b>90.7</b>	83.5	95.2	92.6	<b>68.6</b>	91.1	80.8	86.3
MT-DNN	87.9/87.4	96.0	89.9	86.3	96.5	92.7	68.4	91.1	89.0	87.6
XLNet	90.2/89.8	98.6	90.3	86.3	<b>96.8</b>	<b>93.0</b>	67.8	91.6	<b>90.4</b>	88.4
RoBERTa	<b>90.8/90.2</b>	<b>98.9</b>	90.2	<b>88.2</b>	96.7	92.3	67.8	<b>92.2</b>	89.0	<b>88.5</b>

Model	SQuAD 1.1		SQuAD 2.0	
	EM	F1	EM	F1
<i>Single models on dev, w/o data augmentation</i>				
BERT <sub>LARGE</sub>	84.1	90.9	79.0	81.8
XLNet <sub>LARGE</sub>	<b>89.0</b>	94.5	86.1	88.8
RoBERTa	88.9	<b>94.6</b>	<b>86.5</b>	<b>89.4</b>
<i>Single models on test (as of July 25, 2019)</i>				
XLNet <sub>LARGE</sub>		86.3 <sup>†</sup>	89.1 <sup>†</sup>	
RoBERTa		86.8	89.8	
XLNet + SG-Net Verifier		<b>87.0<sup>†</sup></b>	<b>89.9<sup>†</sup></b>	

Table 7: Results on the RACE test set. BERT<sub>LARGE</sub> and XLNet<sub>LARGE</sub> results are from Yang et al. (2019).

- 더 많은 데이터로 더 오래 훈련
- NSP 사전과제 없앴
- Byte level BPE 를 통해 50,000개 단어사전 사용
- 미니배치 사이즈 증가 (8,000개 사용)
- Dynamic mask 사용

## ALBERT



- 임베딩 매트릭스 분할
- Cross-layer parameter sharing
- NSP 대신 Sentence Order Prediction 과제 도입

RoBERTa, ALBERT, ...

## RoBERTa

	MNLI	QNLI	QQP	RTE	SST	MRPC	CoLA	STS	WNLI	Avg
<i>Single-task single models on dev</i>										
BERT <sub>LARGE</sub>	86.6/-	92.3	91.3	70.4	93.2	88.0	60.6	90.0	-	-
XLNet <sub>LARGE</sub>	89.8/-	93.9	91.8	83.8	95.6	89.2	63.6	91.8	-	-
RoBERTa	90.2/90.2	94.7	92.2	86.6	96.4	90.9	68.0	92.4	91.3	-

<i>Ensembles on test (from leaderboard as of July 25, 2019)</i>										
ALICE	88.2/87.9	95.7	90.7	83.5	95.2	92.6	68.6	91.1	-	-
MT-DNN	87.9/87.4	96.0	89.9	86.3	96.5	92.7	-	-	-	-
XLNet	90.2/89.8	98.6	90.3	86.3	96.5	92.7	-	-	-	-
RoBERTa	90.8/90.2	98.0	90.3	86.3	96.5	92.7	-	-	-	-

Model	SQuAD 1.1	EM	F1
<i>Single models on dev, w/o</i>			
BERT <sub>LARGE</sub>	84.1	90.9	
XLNet <sub>LARGE</sub>	89.0	94.5	
RoBERTa	88.9	94.6	
<i>Single models on test (as of July 25, 2019)</i>			
XLNet <sub>LARGE</sub>			
RoBERTa			
XLNet + SG-Net Verifier			

- 더 많은 데이터
- NSP 사전과제 없
- Byte level BPE 는 총 50,000개 단어사전 사용
- 미니배치 사이즈 증가 (8,000개 사용)
- Dynamic mask 사용

## ALBERT



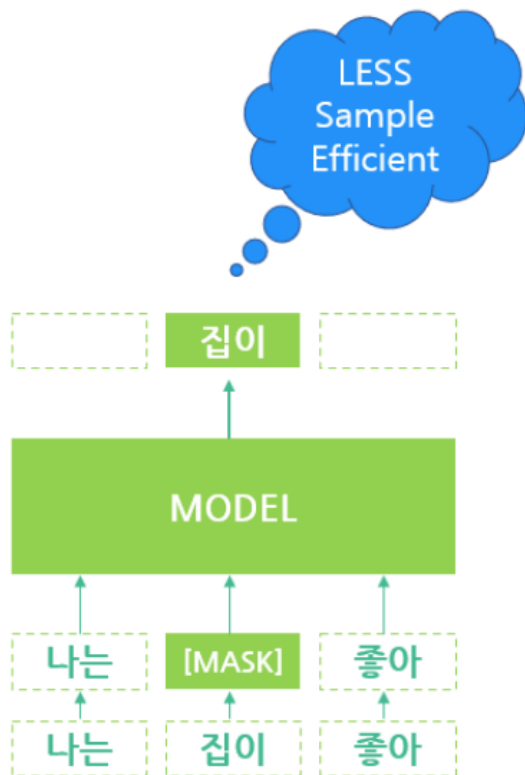
## KEY POINT

- 새로운 사전학습 태스크 고안
- 더 많은 데이터로 더 오래 학습

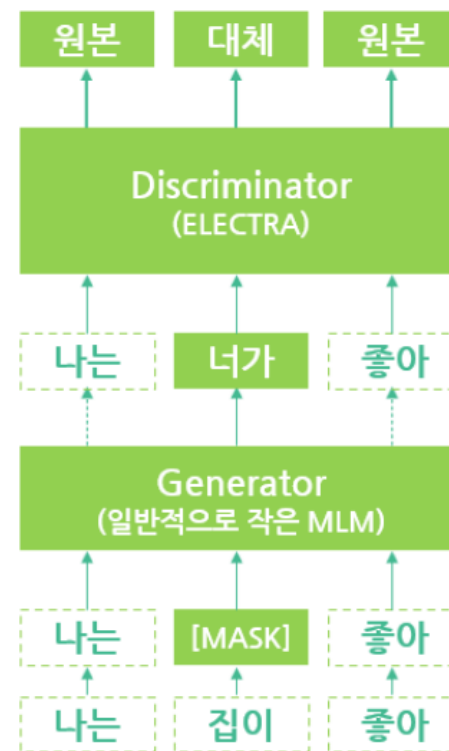
- 임베딩 매트릭스 분할
- Cross-layer parameter sharing
- NSP 대신 Sentence Order Prediction 과제 도입

## ELECTRA

## Masked Language Model

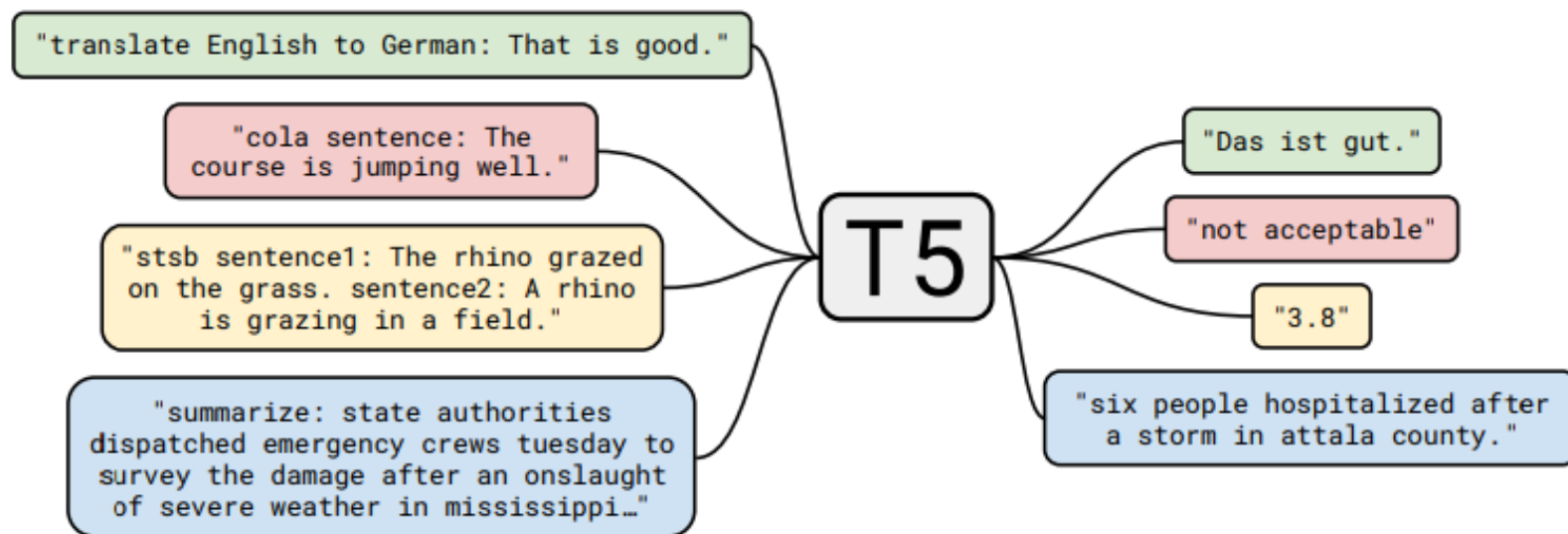


## NEW Pretraining Task



- Pre-training Text Encoders as Discriminators Rather Than Generators
- 사전학습 시 Sample Efficiency의 개선으로 기존 BERT보다 빠르게 수렴 & 성능도 좋음

## Google T5



**Figure 1:** A diagram of our text-to-text framework. Every task we consider – including translation, question answering, and classification – is cast as feeding our model text as input and training it to generate some target text. This allows us to use the same model, loss function, hyperparameters, etc. across our diverse set of tasks. It also provides a standard testbed for the methods included in our empirical survey. “T5” refers to our model, which we dub the “Text-to-Text Transfer Transformer”.

- Text-to-Text라는 단일 NLU 프레임워크를 제시하고 모든 NLU를 생성 과제로 해결

## Google T5

Original text

Thank you ~~for~~ ~~inviting~~ me to your party ~~last~~ week.

Inputs

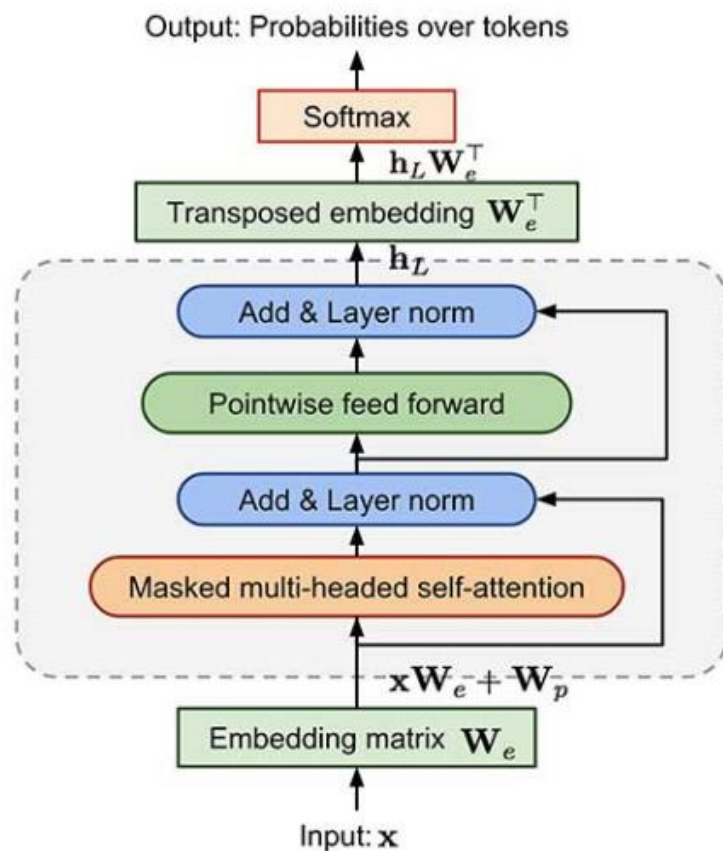
Thank you <X> me to your party <Y> week.

Targets

<X> for inviting <Y> last <Z>

- Generation 모델인 만큼 새로운 형태의 Self-supervised learning 태스크 고안

## 자연어 생성 모델 - OpenAI GPT2



**Transformer Block**  
Repeat  $\times L=12$

$$h_\ell = \text{transformer\_block}(h_{\ell-1})$$

$$\ell = 1, \dots, L$$

- Transformer 디코더 구조를 사용해 Natural Language Generation 과제 사전학습

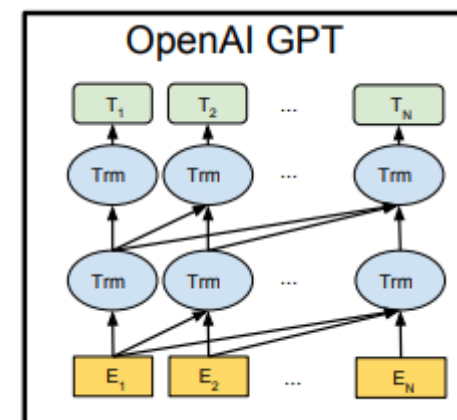


Fig. 2. The transformer decoder model architecture in OpenAI GPT.

## 자연어 생성 모델 - OpenAI GPT2

GPT2 폐기? 너무 생성을 잘해서...?

미 비영리 인공지능(AI) 연구기관 '오픈AI'는 새로 개발한 글쓰기 인공지능 시스템 'GPT-2'를 고심 끝에 폐기하기로 결정했다고 CNN 등이 18일(현지시간) 전했다. 오픈AI의 이 같은 결정은 공익 차원에서 개발한 GPT-2가 가짜 뉴스 생산이나 학생들의 과제·논문 등에 악용될 가능성이 크기 때문으로 알려졌다. GPT-2는 80만개의 인터넷 페이지를 검색하고 15억개의 단어를 학습했다. 이를 바탕으로 문장을 논리적 순서에 맞게 배치해 어떤 글도 막힘없이 써 내려가는 능력을 가졌다.

What you are doing is opposite of open. It is unfortunate that you hype up +propagate fear + thwart reproducibility+scientific endeavor. There is active research from other groups in unsupervised language models. You hype it up like it has never been done before. @jackclarkSF

4 31 222

니들이 하는 건 Open의 반대 아님?(Close AI). (발생할 수 있는 문제에 대해서) 부작용을 과장하고 불안감 조성하면서 성능, 성과까지 과장하다니 참 찌질한듯. 또 다른 그룹에서도 unsupervised language model들에 대해서 활발하게 연구하고 있는데 니들은 마치 니들이 처음인 것 마냥 과장하네?



**Jack Clark** @jackclarkSF · 2월 15일

Hi there. We've released a paper, samples, and small model. Not our intention to hype our research relative to others (DM me specifics here and can probably help?). How do you think about intersection of increasingly general models and potential for malicious use?

4 30

(부들대며) 안녕. 우리는 페이퍼, 샘플, 그리고 스몰 모델까지 릴리스했는데? 그리고 우리 연구를 과장할 의도는 없었는데? (혹시 자세한 설명이 필요하면 쪽지 보내든가^^) 너는 general model이 많아짐에 따라 발생할 수 있는 악용되는 사례에 대해서 어떻게 생각하는데?



**Anima Anandkumar** @AnimaAnandkumar · 2월 15일

Malicious BS. You are doing a disservice to science by using that word. #AI progress is largely attributed to open source and open sharing. What irony @openAI is doing opposite and trying to take a higher moral ground. Can u list all so-called malicious cases and I will respond

8 4 73

악질적인 대답이네! 니야말로 그런 단어를 사용함으로써 과학을 꺼내려는거임. AI는 open 소스랑 공유에 의해서 발전하는 건데, 'Open' AI 니들이 반대로 하면서 도덕적인 우위를 점하려는게 참 아이러니함. 그리고 어떤 악용 사례가 있었는지 알려주면 내가 또 딜 들어갈게.



**Jack Clark** @jackclarkSF · 2월 15일

Hiya, here are some malicious use cases (outlined in the blog). What do you think?

우리 블로그에서 가져온 사례들인데 어떻게 생각함?



## 자연어 생성 모델 - GPT 생성 예시

## 인공지능 기자 &lt;지피티&gt;

지피티

안녕하세요

지피티

저희 팀 symbiote-ai.com 는 영리를 목적으로 하지 않는 비영리 조직이며 지피티 프로젝트는 연구 목적으로 진행되고 있습니다.

지피티

인공지능 지피티가 만들어 주었으면 하는 기사의 첫 문장을 입력해 주세요.

지피티

예시:  
 제목: 서울 한여름 폭설 충격, 기상이변 현실  
 Q: 인도의 수도는  
 아이언맨: 나는 이 세상을 그냥 두고볼 수 없어!  
 홍대 맛집 BEST5 1.

김명지: 나는 치킨보다는 피자

지피티

지금 [김명지: 나는 치킨보다는 피자]로 기사를 쓰고 있습니다.

지피티

제가 [김명지: 나는 치킨보다는 피자]로 쓴 기사입니다.

지피티

김명지: 나는 치킨보다는 피자 한판이 좋다. (닭은 왜 맛이 없냐는 질문에) 치킨은 닭고기잖아요. 그런 맛이다. 또 피자는 칼로리가 높아서 더 많은 양으로 먹는 게 낫다. (웃음)

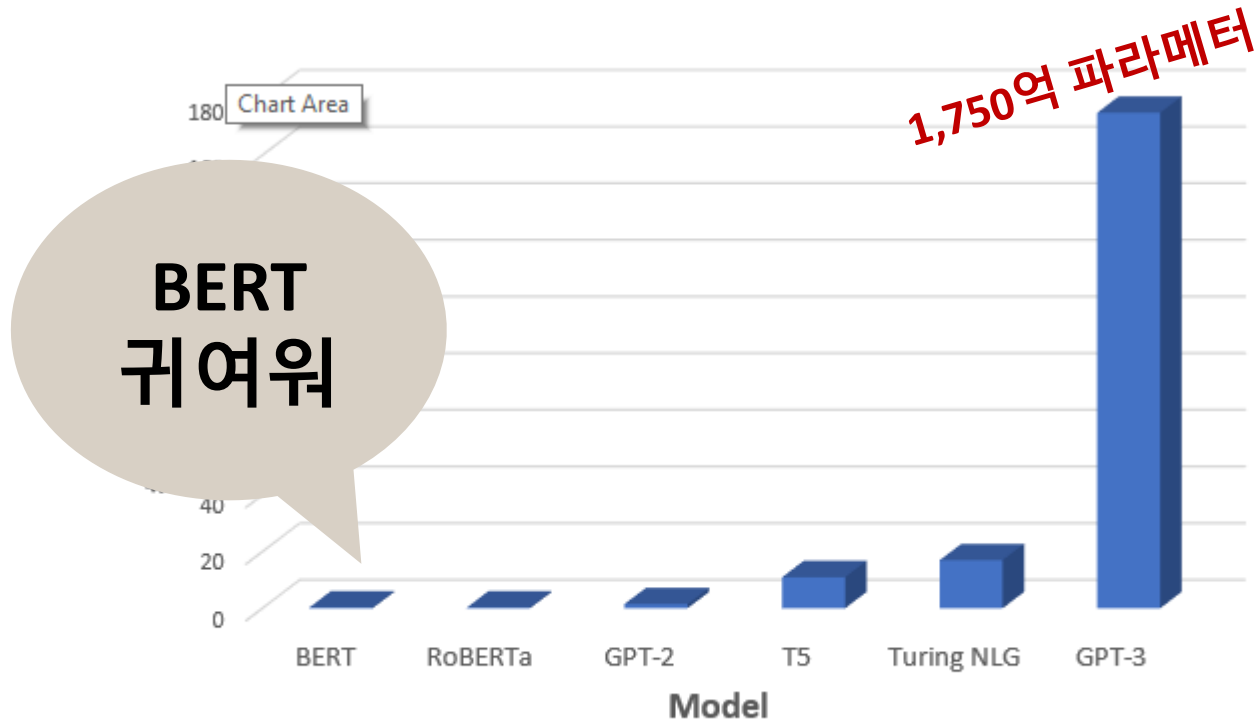
제목: 드디어 외계인과 접촉, 서울시 한복판에 UFO 등장

뉴스의 시작 문장을 입력하세요 (매번 결과가 다릅니다)



GPT-3

따끈따끈



GPT-3

따끈따끈

APPLE -&gt; ELPPA

PEN -&gt; NEP

PINEAPPLE -&gt; ?

GPT-3

따끈따끈

저는 피자가 더 좋아요 -> I prefer pizza

엥? 치킨보다 -> ?

GPT-3

따끈따끈

다음 문장을 보고 긍정적인 감성이면 1  
부정적인 감성이면 0을 리턴하시오

치킨 5분 뒤에 도착한대 ㅎㅎㅎㅎ -> ?

GPT-3

따끈따끈

APPLE -> ELPPA  
PEN -> NEP  
PINEAPPLE -> ?

Few Shot Learning

저는 피자가 더 좋아요 -> I prefer pizza  
앵? 치킨보다 -> ?

One Shot Learning

다음 문장을 보고 긍정적인 감성이면 1  
부정적인 감성이면 0을 리턴하시오  
치킨 5분 뒤에 도착한대 ㅎㅎㅎㅎ -> ?

Zero Shot Learning

GPT-3

따끈따끈

A small, boxy robot made of cardboard sits on an open book. The robot has two circular eyes and a small antenna. The text "IN-CONTEXT LEARNING" is overlaid on the image.

## IN-CONTEXT LEARNING

GPT-3

따끈따끈

A small, boxy robot made of cardboard sits on an open book. The robot has two circular eyes and a small antenna. The text "IN-CONTEXT LEARNING" is overlaid on the image.

## IN-CONTEXT LEARNING



## 자연어 생성 모델 – Facebook Blender

Human

Generative BST 9.4B

If you could have dinner with one person, dead or alive, who would that be?

That's a tough one. I think I'd have to go with Steve Jobs. I'd love to pick his brain.

Ah, interesting choice. What do you know about Steve Jobs?

He was one of the most influential people in the history of technology. He was a visionary.

What questions would you ask him?

I'd want to know how he came up with the idea for the iPod. It was so revolutionary at the time.

Totally...the iPod changed the way everyone listens to music.

It really did. It's crazy to think that it was released in 2001. That was a long time ago.

Omg wow, that makes me feel old!

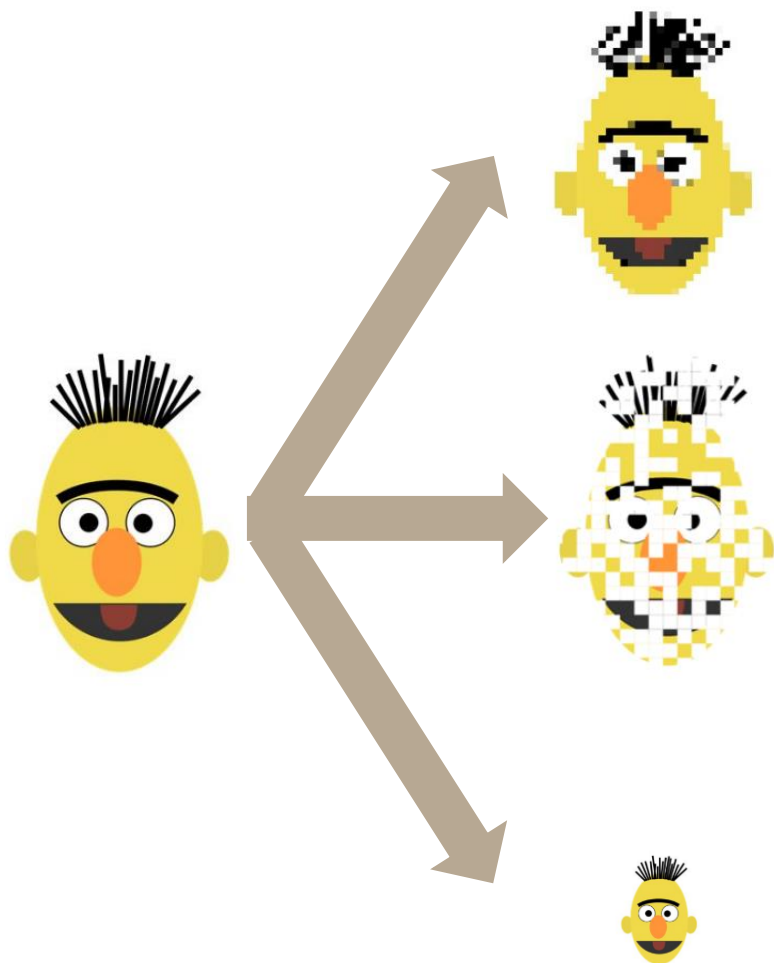
I know, right? Time flies by so fast. I wish I could go back to those simpler times.

Me too, friend, me too. Speaking of music, what's your favorite album to listen to lately?

I've been listening to a lot of Led Zeppelin lately. They're my favorite band.

- 오픈 도메인 멀티 턴 대화모델로, 페르소나를 가지고 맥락에 맞는 발화를 생성함

모델, 클 수록 성능은 좋은데...



### Quantization

- 32비트 부동소수점 연산을 16비트 연산 혹은 8비트 정수 연산으로 변환

### Model Pruning

- 네트워크 중 추론 성능을 크게 저하하지 않는 노드 혹은 weight matrix를 없애 크기를 줄임

### Knowledge Distillation

- 더 작은 모델이 원래 모델의 추론을 따라 하도록 훈련해 큰 모델이 학습한 지식을 전파

# WPAP-UP

## 실습으로 수행했던 프로젝트들

Three blue project cards are displayed diagonally. The top card is labeled 'RNN', the middle card is labeled 'CNN', and the bottom card is labeled 'BERT'. In the background, a blurred image of a 'Seq2Seq' project card is visible.

Two project cards are shown. The top card is titled '유사도 분석' (Similarity Analysis). Below it is a larger blue card labeled 'MINI PROJECT'. The background shows a blurred image of a chatbot interface.

Three blue project cards are displayed diagonally. The top card is labeled 'Seq2Seq', the middle card is labeled 'ATTENTION', and the bottom card is labeled 'TRANSFORMER'. In the background, a blurred image of a 'Decoder' project card is visible.

A project card titled '문서 요약' (Document Summarization) featuring a screenshot of a news article from 'NEWSIS' about LG CNS and a summarization tool interface. The interface includes a '요약' (Summarize) button and a '자동 요약 결과가 어떨까요?' (How is the automatic summarization result?) section with '만족' (Satisfied), '보통' (Average), and '불만족' (Dissatisfied) options.

Two project cards are shown. The top card is titled '기계 독해' (Machine Reading). Below it is a larger blue card labeled 'BERT'. The background shows a blurred image of a 'Reading' project card.

A project card titled '자연어 생성' (Natural Language Generation) featuring a chatbot interface. The chatbot has generated several responses, each with a score in parentheses: '마음 진짜냐?' (2.20), '축하한다네!' (1.98), '힐 어떡하냐..', '어쩌다가 그랬냐?' (1.63), and '(끄덕) 안다네!' (1.60).

## CBOW/Skip-gram

- Tips

- 대량의 corpus로 Token 임베딩 사전학습
- 토큰 임베딩은 NLU 모델링의 기본 -> 학습해둔 임베딩을 RNN, CNN 등 모델링 단계에서 불러와 모델 성능 향상에 이용할 수 있다.

국립국어원 언어정보나눔터

말뭉치 전자사전 용어/문자찾기 통합자료실 회원교류 국어정보경쟁

통합자료실 > 데이터베이스 자료 > 말뭉치 파일

말뭉치 검색

자료구분: 외부용

검색어: ... 전체 ...

총 1495건이 있습니다.

번호	제목	자료구분	출원사망	출원일자
1495	독백_여행이야기#2, 전자전사자료	외부용	관리자	2014-01-09
1494	설교_교회목사#2, 전자전사자료	외부용	관리자	2014-01-09
1493	설교_교회목사#1, 전자전사자료	외부용	관리자	2014-01-09
1492	폐회사_한세주, 전자전사자료	외부용	관리자	2014-01-09
1491	폐회사_한세주, 전자전사자료	외부용	관리자	2014-01-09
1490	폐회사_마카데미#2, 전자전사자료	외부용	관리자	2014-01-09
1489	폐회사_마카데미#2, 전자전사자료	외부용	관리자	2014-01-09
1488	폐회사_마카데미#1, 전자전사자료	외부용	관리자	2014-01-09
1487	강연_아이발달, 전자전사자료	외부용	관리자	2014-01-09
1486	강연_골다공증, 전자전사자료	외부용	관리자	2014-01-09

자연어 말뭉치는  
풍부하게 존재한다!

## 텍스트 분류

토큰 1   토큰 2   토큰 3   ...   토큰 n



분류 1	0%
분류 2	90%
...	
분류 k	2%

**자연어 인풋**  
(문장 / 문서 / ...)

**카테고리**

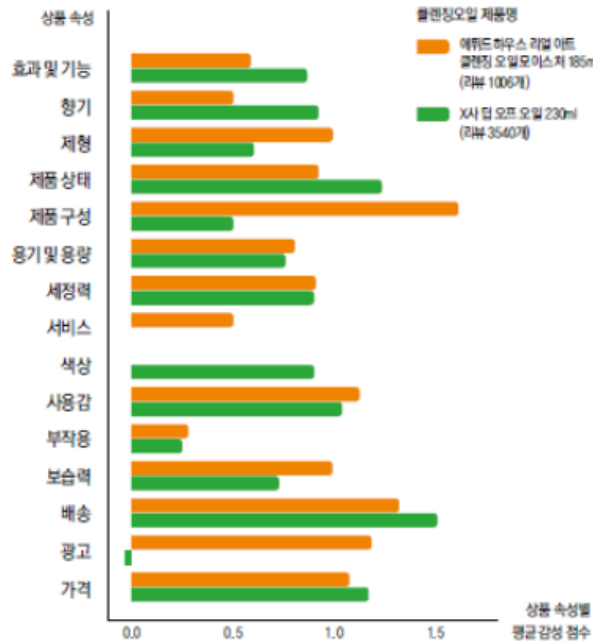
Softmax

## 감성분석-사례

### • 에뛰드하우스 감성분석 기반 마케팅 대시보드

- SNS 소셜 버즈는 고객의 '생생한 목소리'를 담은 중요한 마케팅 소통 창구
- 에뛰드에서는 화장품 구매 후기에 있는 문자 데이터를 수집 & 감성분석을 통해 자사 브랜드 상품에 대한 의견을 분석하고 경쟁사와 비교 분석하여 마케팅 인사이트 도출에 활용

**그림8** 클렌징오일 경쟁 상품 속성별 평균 감성 점수 비교 예시



**그림9** 클렌징 상품의 속성별 상품지도 분석 그래프



## 감성분석

RNN vs CNN vs BERT

풀고자 하는 NLU 태스크를  
수행할 수 있는 모델을  
다양하게 디자인할 수 있음

### RNN Training Log

```
Epoch 1/5
500/500 [=====] - 49s 98ms/step - loss: 0.5330 - val_accuracy: 0.8115
Epoch 2/5
500/500 [=====] - 47s 94ms/step - loss: 0.3609 - val_accuracy: 0.8237
Epoch 3/5
500/500 [=====] - 47s 95ms/step - loss: 0.2875 - val_accuracy: 0.8257
```

### CNN Training Log

```
Epoch 1/5
500/500 [=====] - 5s 9ms/step - loss: 0.4799 - val_accuracy: 0.8108
Epoch 2/5
500/500 [=====] - 4s 9ms/step - loss: 0.3748 - val_accuracy: 0.8298
Epoch 3/5
500/500 [=====] - 5s 9ms/step - loss: 0.3287 - val_accuracy: 0.8363
Epoch 4/5
500/500 [=====] - 5s 9ms/step - loss: 0.2906 - val_accuracy: 0.8385
```

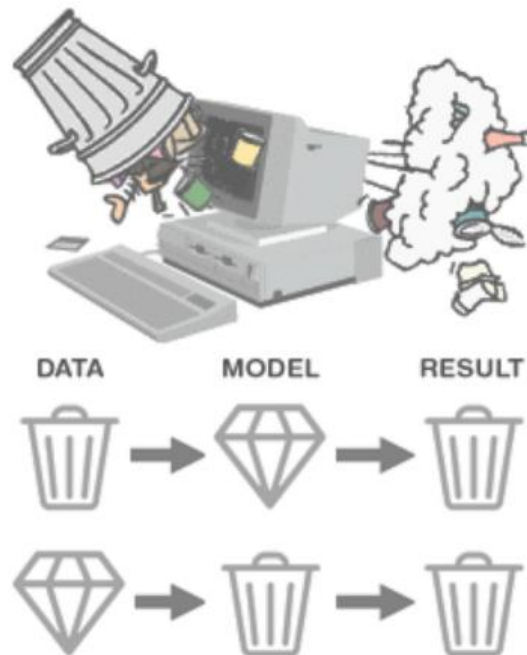
### BERT

```
Epoch 1/2
2084/2084 [=====] - 843s 404ms/step - loss: 0.4394 - val_accuracy: 0.8361
Epoch 2/2
2084/2084 [=====] - 840s 403ms/step - loss: 0.3221 - val_accuracy: 0.8486
```



## 의도분류

챗봇 NLU의 핵심!



EXAMPLE...<ack>  
예 잘 알 겠 습니다 .  
예 알 겠 습니다 .  
네 , 알 겠 습니다 .

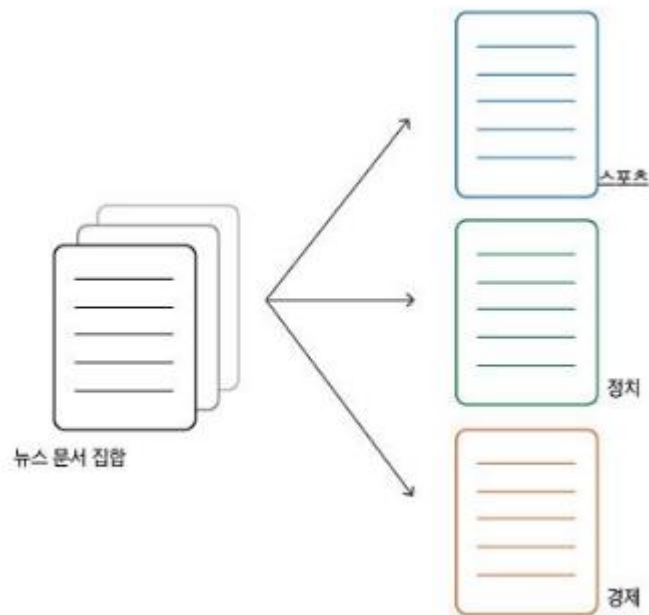
---

EXAMPLE...<affirm>  
일정 을 변경했 습니다 .  
네 좋 습니다 .  
네 그렇 습니다 .

적절한 의도 & Chat Flow를 설계하는 것이 중요함

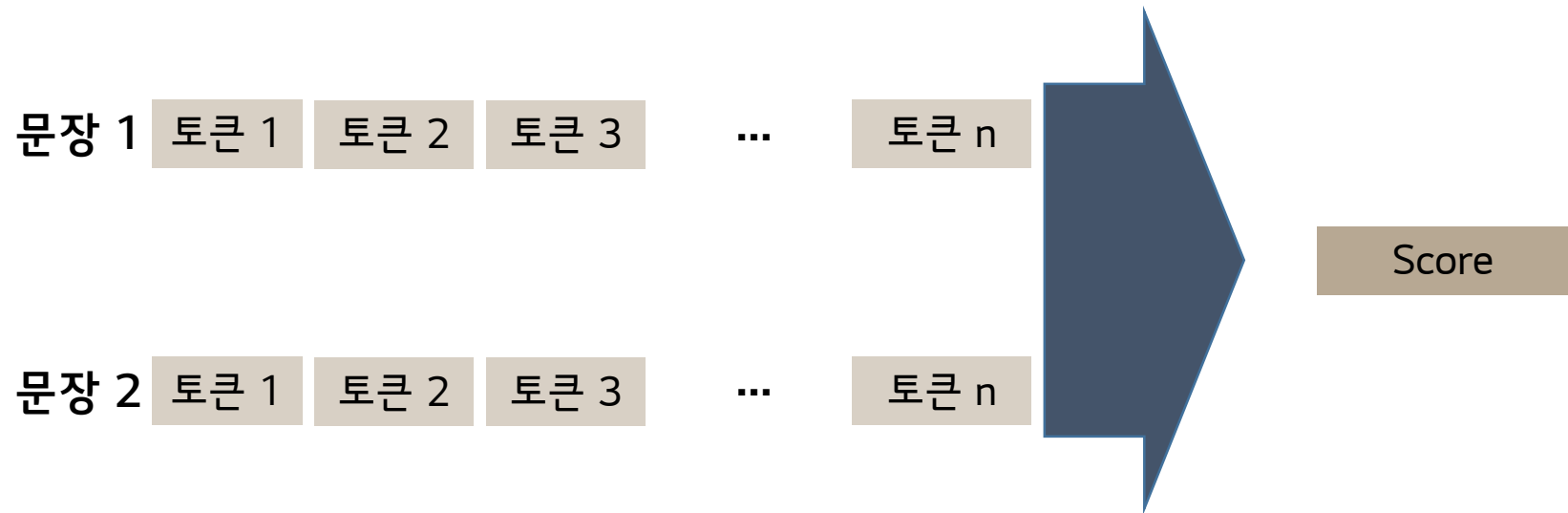
## 기타 텍스트 분류 사례

### 문서 분류

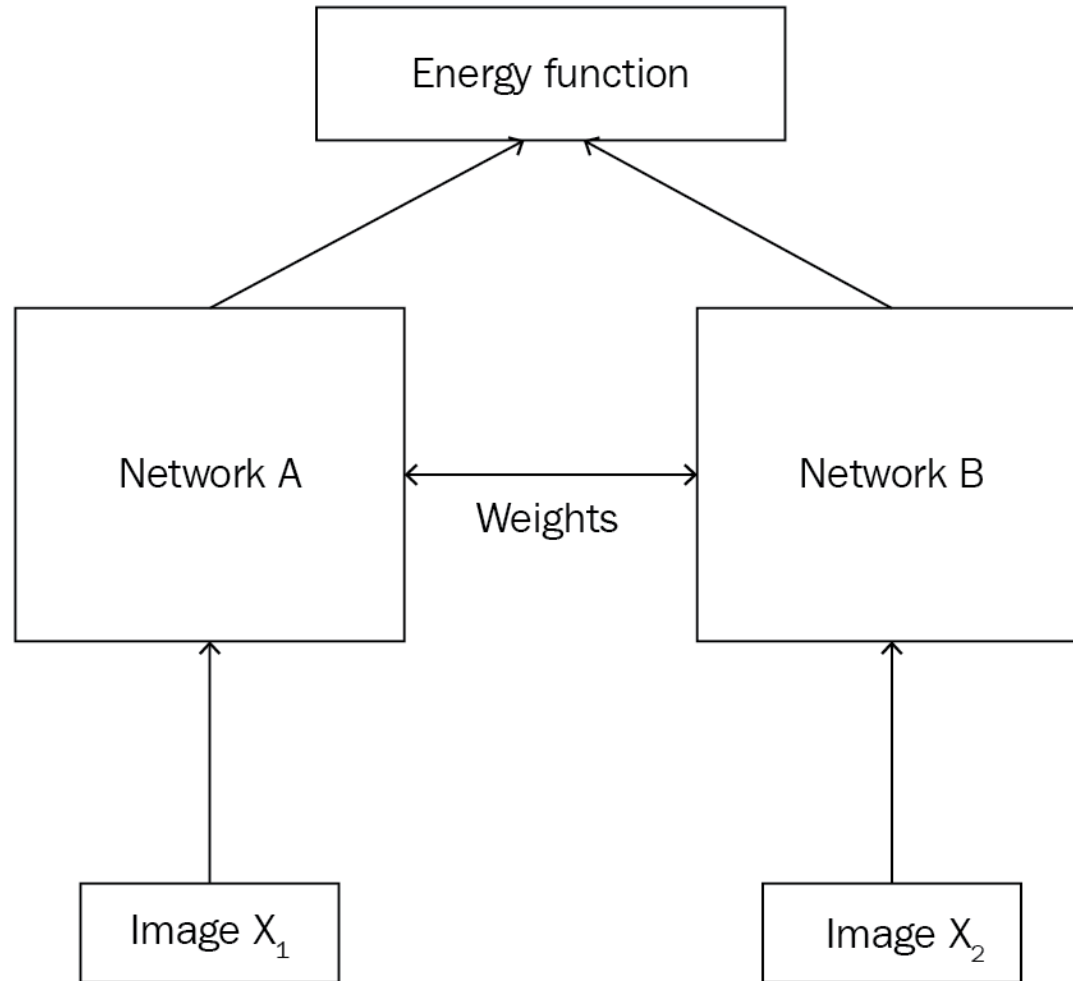


문서 카테고리  
자동 분류

## 유사도 분석



## 유사도 분석



## 유사도 분석 - 사례

- 유사 선행 특허 검색 시스템
  - 키워드 관점에서 유사성 높은 특허를 생성/ 특허 별로 유사 특허 추천 및 유사도를 제공하는 서비스

### 시범 서비스

➤ 검색할 특허 내용을 입력하면 유사 선행 특허가 검색됩니다.

유사 선행 특허 검색 서비스 바로가기

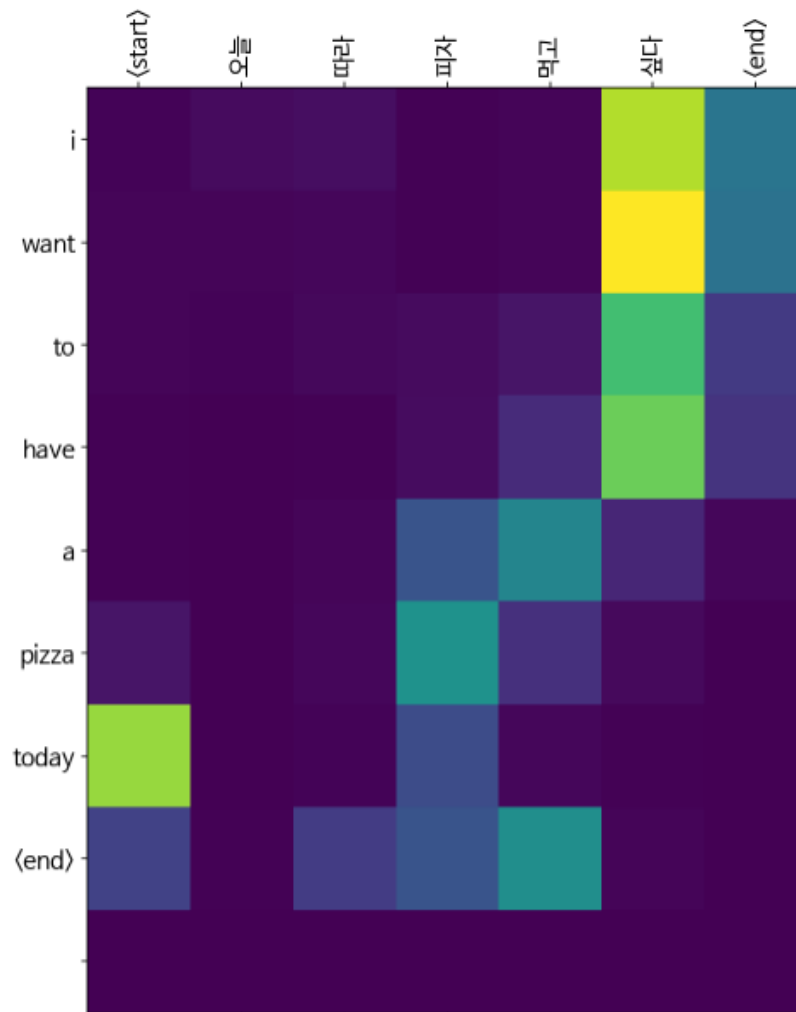
AI HUB

- FAQ 검색
  - 사용자 질문에 대해 가장 유사한 기존 FAQ 데이터를 매핑해주는 서비스
  - Information Retrieval 등과 관련

## 기계 번역

[44] translate("오늘따라 피자 먹고 싶다")

Input: <start> 오늘 따라 피자 먹고 싶다 <end>  
 Predicted translation: i want to have a pizza today <end>



## 기계 독해

BERT

```
[ ] context = ""'동물의 숲' 스페셜 에디션으로 출시된 해당 제품의 정가는 36만원이다.  
그러나 최근 코로나19로 수요대비 공급량 자체가 줄어들면서 해당 제품에 프리미엄을 붙여 되파는 이들이 늘어나기 시작했다.  
경찰 관계자는 "원래 30만원 중반에 판매되는 게임이 70만원에서 많게는 90만원까지 되팔고 있는 것으로 알고 있다"면서  
"온라인 사기 특성상 한 계좌로 모이게 되는데 범죄에 사용된 계좌가 개설된 지역으로 사건들이 이관돼 이제 막 수사에 착수한 상황"이라고 설명했다.""  
  
print(context)
```

↳ '동물의 숲' 스페셜 에디션으로 출시된 해당 제품의 정가는 36만원이다.  
그러나 최근 코로나19로 수요대비 공급량 자체가 줄어들면서 해당 제품에 프리미엄을 붙여 되파는 이들이 늘어나기 시작했다.  
경찰 관계자는 "원래 30만원 중반에 판매되는 게임이 70만원에서 많게는 90만원까지 되팔고 있는 것으로 알고 있다"면서  
"온라인 사기 특성상 한 계좌로 모이게 되는데 범죄에 사용된 계좌가 개설된 지역으로 사건들이 이관돼 이제 막 수사에 착수한 상황"이라고 설명했다.

```
[ ] question = "동숲 정가 얼마야"  
inference(question, context, tokenizer, model)
```

↳ 36만

## 기계 독해

### Leaderboard

KorQuAD 1.0의 Test set으로 평가한 Exact Match(EM) 및 F1 score 입니다.

Rank	Reg. Date	Model	EM	F1
-	2018.10.17	<b>Human Performance</b>	80.17	91.20
1	2020.01.08	SkERT-Large (single model) Skelter Labs	87.66	<b>95.15</b>
2	2019.10.25	KorBERT-Large v1.0 ETRI ExoBrain Team	<b>87.76</b>	95.02





## 기계 독해 - 사례

- 금융감독원의 Suptech (Supervision + Technology) 활용 계획
  - 사모 펀드 심사 업무 지원에 MRC 기술 사용
  - AI 엔진이 약관 문서를 분석해 체크 리스트에 대해 답변을 제시하면 직원이 이를 확인

### < AI 약관 심사 지원시스템 개념 프로세스 >

