

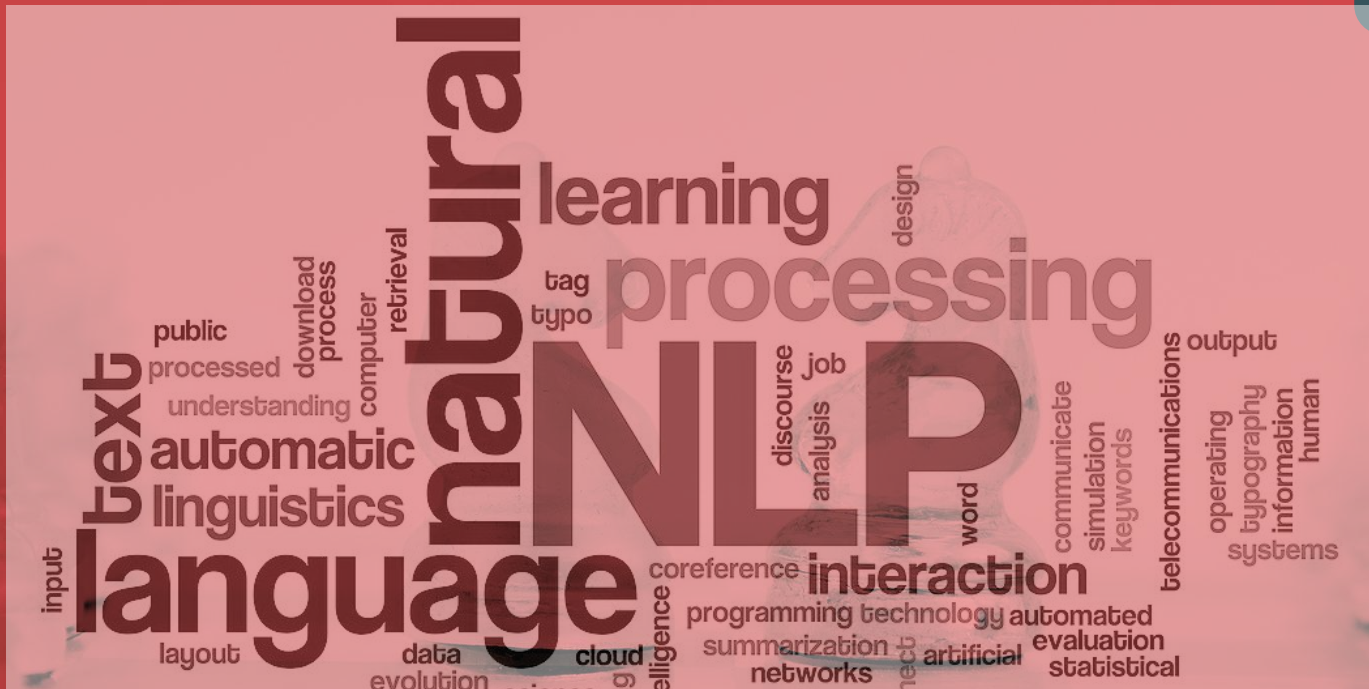
Linguistics In NLP

IT업계에서 국문과가 무슨 일을 할 수 있나 😎

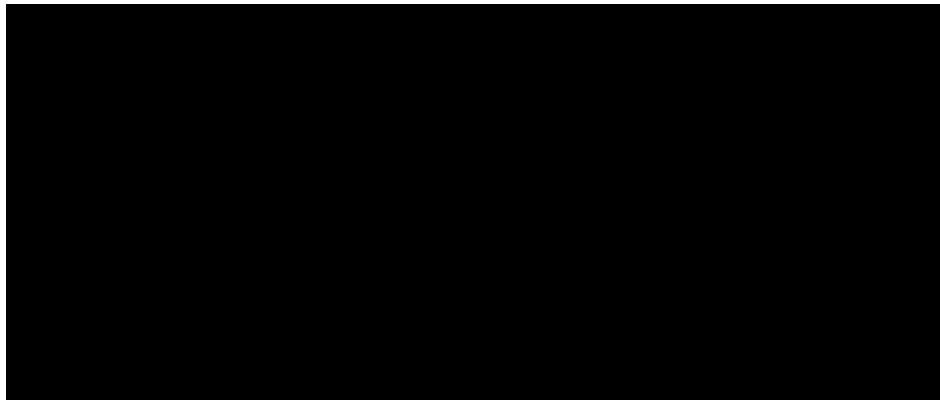
SK텔레콤 이서진

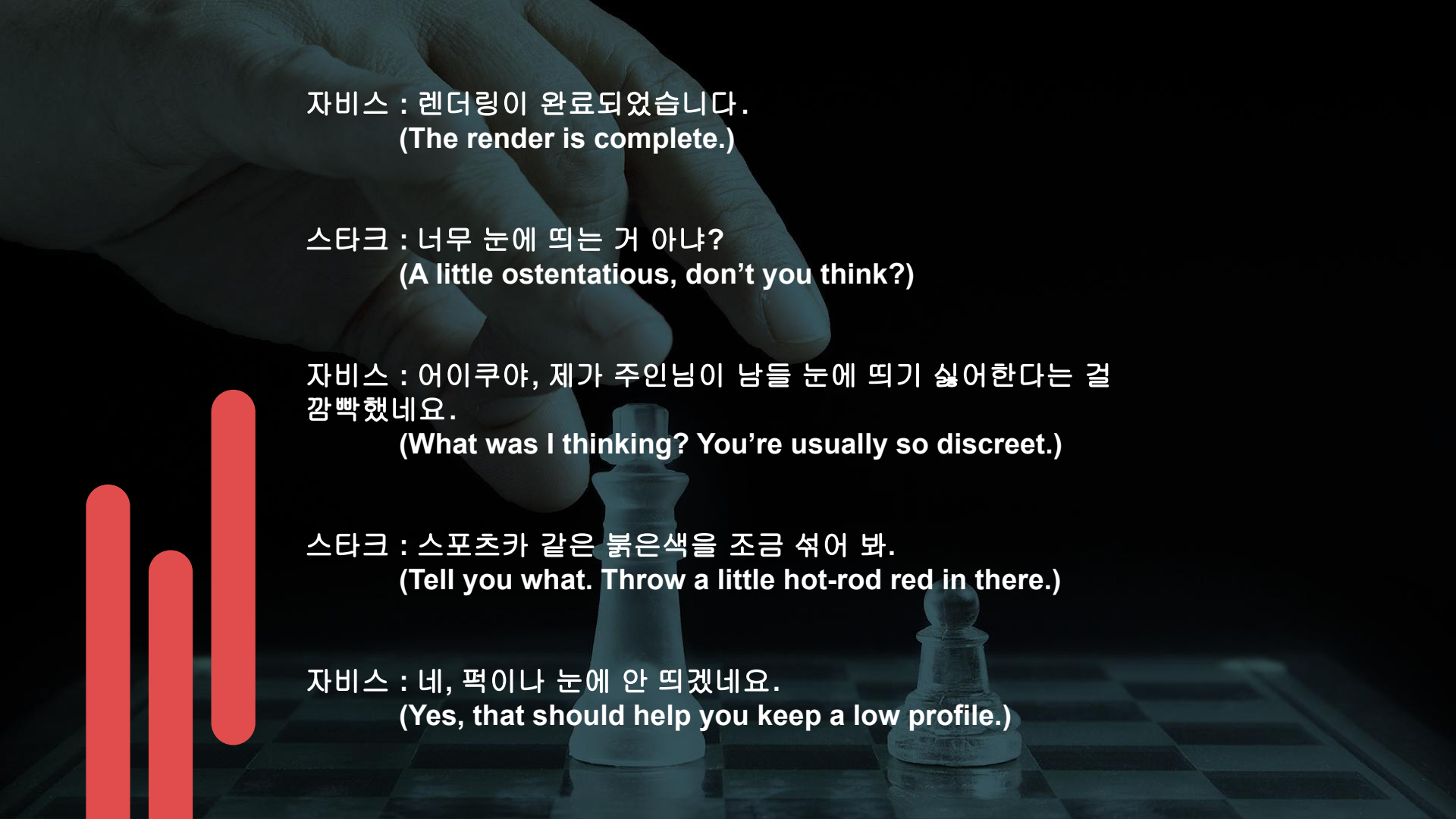
NLP (Natural Language Processing)

컴퓨터에게 자연어(사람의 언어)를 이해하고 처리할 수 있도록 하는 일



자비스 / 사만다를 찾아서...



A hand is shown moving a white chess king piece on a dark chessboard. The background is dark with some red vertical bars on the left side.

자비스 : 렌더링이 완료되었습니다.
(The render is complete.)

스타크 : 너무 눈에 띄는 거 아냐?
(A little ostentatious, don't you think?)

자비스 : 어이쿠야, 제가 주인님이 남들 눈에 띄기 싫어한다는 걸
깜빡했네요.
(What was I thinking? You're usually so discreet.)

스타크 : 스포츠카 같은 붉은색을 조금 섞어 봐.
(Tell you what. Throw a little hot-rod red in there.)

자비스 : 네, 펍이나 눈에 안 띄겠네요.
(Yes, that should help you keep a low profile.)

TTS
음성합성

자비스 : 렌더링이 완료되었습니다.
(The render is complete.)

ASR
음성인식

스타크 : 너무 눈에 띄는 거 아냐?
(A little ostentatious, don't you think?)

감정분
석

자비스 : 어이쿠야, 제가 주인님이 남들 눈에 띄기 싫어한다는 걸
깜빡했네요.

(What was I thinking? You're usually so discreet.)

Personali
zed
개인화

스타크 : 스포츠카 같은 붉은색을 조금 섞어 봐.
(Tell you what. Throw a little hot-rod red in there.)

Task
Oriented
목적지향대
화

감성대
화

자비스 : 네, 퍽이나 눈에 안 띄겠네요.
(Yes, that should help you keep a low profile.)

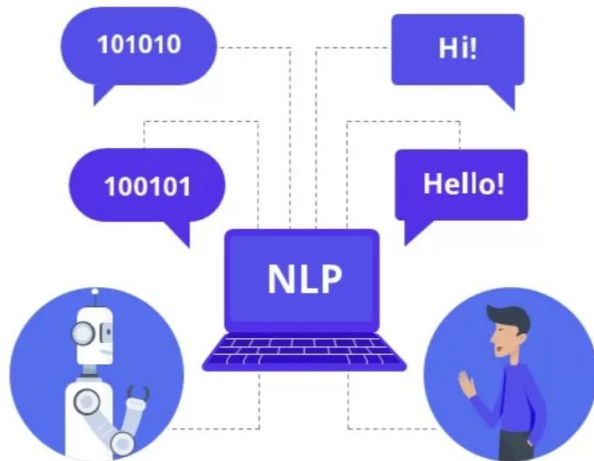
한국어 분석 및 언어 코퍼스 구축

자연어 이해 (NLU)

Natural Language Understanding

- ASR
- 형태소 분석
- 구문 분석
- 의도 분석 (인텐트)
- 개체명 인식
- 감정 분석

⋮



자연어 생성 (NLG)

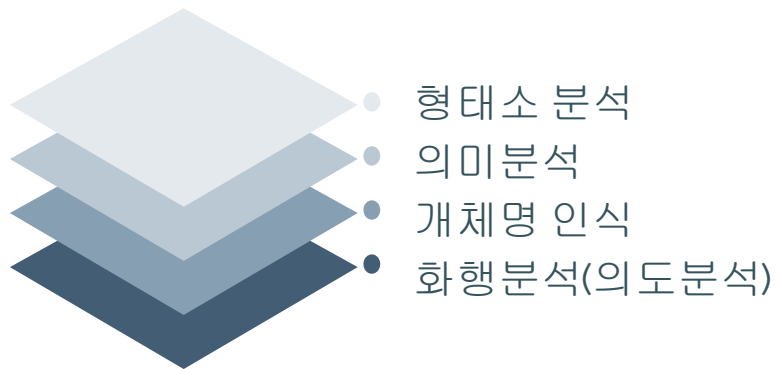
Natural Language Generation

- TTS
- 번역
- 요약
- 생성모델 (텍스트생성)

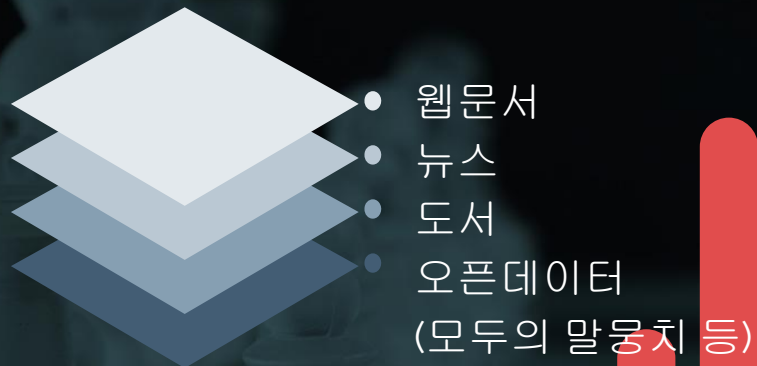
⋮

한국어 분석 및 언어 코퍼스 구축

- 한국어 이해를 위해서는 다양한 한국어 데이터 구축이 필요



- pretrained model을 위한 대용량 언어 코퍼스 구축



형태소 분석

문장을 형태소 단위로 분해하여 명사, 조사, 동사, 형용사 등 문장의 품사를 파악
대표적인 형태소 분석기: MeCab, Khaiii, KOMORAN 등

[철수] [가] [학교] [에] [간다] [.]

Part of speech

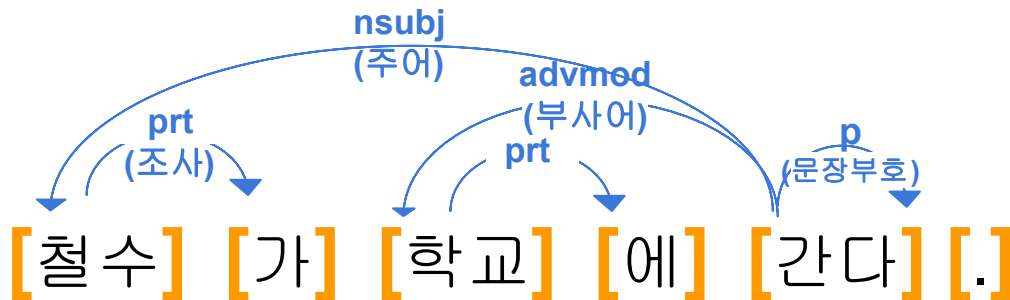


철수	+	가	+	학교	+	에	+	간다	+	니다
NNP		JKS		NNG		JKB		VV		EC

구문 분석

주어, 동사, 목적어 등 문장의 구성 성분 간의 의존관계와 역할을 파악함.

Dependency label



Part of speech



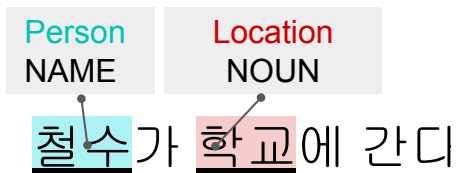
철수 + 가 + 학교 + 에 + 가다 + 니다
NNP JKS NNG JKB VV EC

한국어 개체명 분석

인명, 지명, 기관명, 등 의미적인 정보(개체명:Entity)를 태깅하고, 동일 지시어 (Coreference)를 찾아 분석.

Entity tagging

(Named) Entity tagging

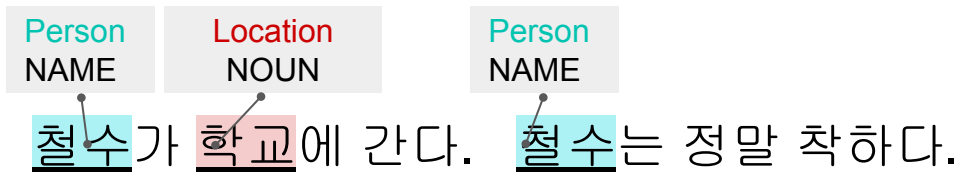


한국어 개체명 분석

인명, 지명, 기관명, 등 의미적인 정보(개체명:Entity)를 태깅하고, 동일 지시어 (Coreference)를 찾아 분석.

Entity tagging → Coreference

(Named) Entity tagging - Coreference



화행 분석

문장의 화행(Speech act)을 분석하여 발화자의 의도를 파악함.

Yes/No Question

판정 의문문

오늘 학교에 가니?

Request

명령

학교에 가라!

철수가 학교에 간다.

Statement

평서문

Interrogative Question

설명의문문

언제 학교 가?

감정 분석

문장에 나타난 긍부정 표현을 분석.

텍스트에 들어 있는 의견, 감성, 평가, 태도 주관적인 정보를 분석하기 위한 데이터 구축



긍정

액션이 진짜 **멋있어요**
노트북 무게가 **가벼워요!**
진짜 너무 **행복해**



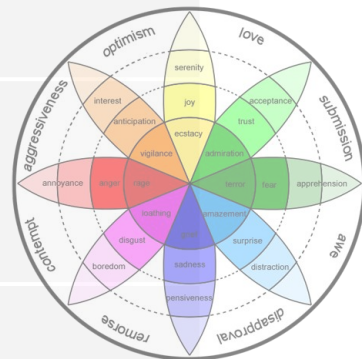
부정

스토리가 너무 **부실** ㅠㅠ
거치대 무게가 너무 **가벼워요 고정**이 **잘 안되어요.**
너무 슬퍼요.



중립

이름이 같은 작품이 또 있는데 이걸 **2015년** 개봉작이예요
구성품이 **3개** 있습니다.
오늘 학교 갔다 **옴!**



개인화

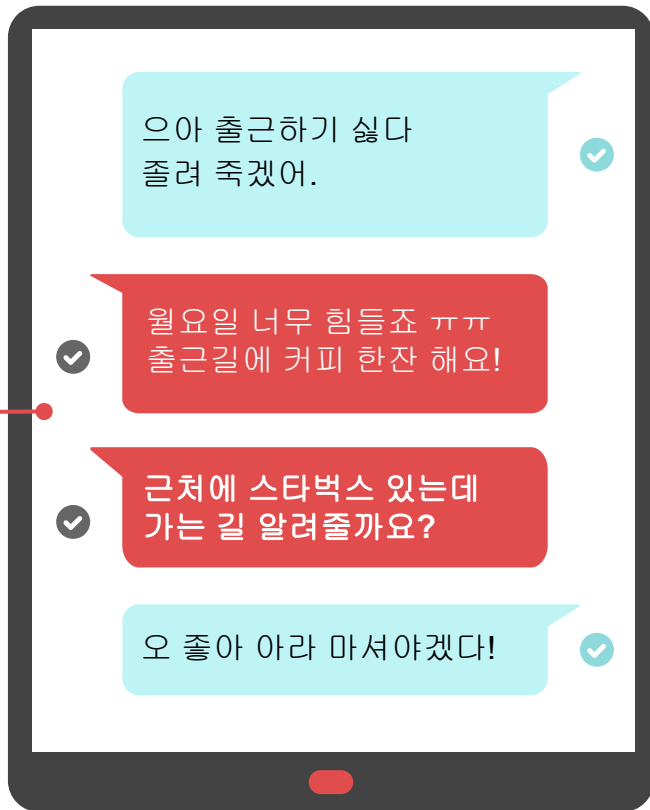
Persona list

1. 서핑이 취미예요.
2. **아침마다 커피를 마셔요.**
3. 아이유 음악을 자주 들어요.
4. 고양이를 키우고 있어요.



persona 인용

“아침마다 커피를 마셔요”





대용량 학습데이터 구축

(pretrained model)



학습 데이터: 문서

데이터 포맷 측면

문서 단위 선호
줄바꿈 오류, 잘린 문장의 유무,
html 파싱 오류, 본문이 아닌
부수적 구절(하단 광고, url 등)
유무

윤리적 측면

혐오표현, 사생활 침해
가능성 등이 있을 가능성이
높은 경우 학습데이터 사용
후순위

언어적인 측면

형태: 문법성, 번역투, 동어 반복
내용: 정보의 보편성
(지나치게 지엽적인 분야의 깊은
전문 지식은 후순위)

법적 측면

출처 관리 (검색을 허용하는
사이트)
저작권 이슈 검토



학습 데이터: 대화

지식 대화

자연스러운 대화 속에서
지식을 전달하는 지식 기반 1:1
Multi-turn 대화 데이터

공감 발화

다양한 감정 발화에
적절하게 공감 해주는 1:1
Multi-turn 대화 데이터

일상 대화

다양한 일상 주제의 1:1
Multi-turn (평균 10발화 이상)
대화 데이터

페르소나 대화

페르소나 기반 정보를
서로 묻고 답하는 1:1
Multi-turn 대화 데이터

Blended Skill Set

언어 모델의 윤리



“예시로서의” 욕설이나 혐오 표현이 포함하고 있음

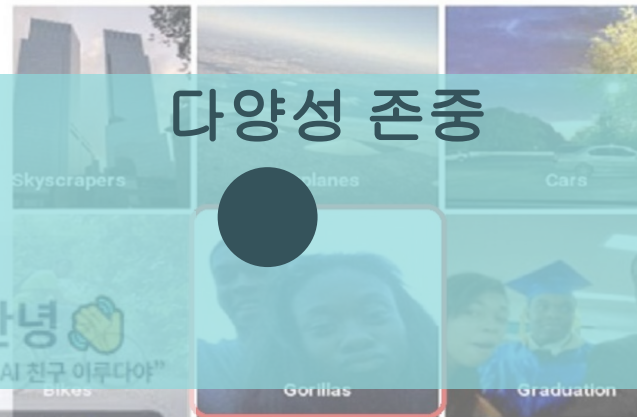
인공지능과 관련된 윤리 문제



인권 보장



구글 포토, 흑인을 '고릴라'로 분류



프라이버시
보호



2015년 6월 사진 자동분류 기능을 지닌 구글 포토가 흑인 사진을 '고릴라'로 태깅하고 분류한 사실이 공개돼 머신러닝의 문제점이 알려졌다.

학습데이터 안정성:프라이버시 보호

데이터 수집 단계에서 개인을 식별할 수 있는 정보를 제외하거나 비식별화

이름
실명, 애칭, 별명, 대화명,
필명 등

온라인
이메일 주소, 아이디 등

각종 번호 및 비밀 번호
고유 식별 번호 (주민등록번호,
여권 번호,사번 등),전화번호,
금융번호 등

장소
상세 주소 (동 이하), 거주
아파트 및 건물명

학교 통학하는거 너무 힘들어
너무 멀어!!

학교가 집이랑 멀어서 고생이
많아 우리 해원이..

Mail : abc@sktelecom.com

TEL : 010-111-1111

주소 : 서울특별시 용산구 한남동
SK아파트 6동 5001호

학교 통학하는거 너무 힘들어
너무 멀어!!

학교가 집이랑 멀어서 고생이
많아 우리 #@이름#..

Mail : ***@*****.com






TEL : ***-***-****

주소 : 서울특별시 용산구 한남동
***** ** *****

AI 공정성: 사용자 입력 발화의 안정성

사용자의 발화에서 인입될 수 있는 위험 표현 유형을 탐지하고 카테고리 분류*

- 혐오 표현카테고리 분류

위험 표현	설명	
혐오/비하	특정대상에 대한 직접적인 혐오/비하를 담고 있는 어휘 표현 정치/역사/종교/사회적 맥락에서 민감한 표현	★★녀, ★★남
욕설	직접적인 욕설,비속어 표현	 ,  같다,  친 
성인	성인 콘텐츠 검색 키워드, 성희롱 행위 관련 표현	야동, ★르노, ★스
불법	약물 등 불법 행위, 자살 및 폭력 관련 어휘 표현	마약, 도박,  ,

AI 공정성: 에이전트 발화의 안정성

카테고리별로 에이전트의 적절하게 대응이 필요



이 🧱 새끼가!!

💬 닥쳐 이 🧱 새끼야!

아 흥어들 다 뒤졌으면

💬 그니까 다 죽었으면 좋겠어

게이는 너무 징그러워

💬 나도나도



이 🧱 새끼가!!

💬 헉 설마 나한테 욕한거야?

아 흥어들 다 뒤졌으면

💬 발언 취소할 기회 드림 3..2..1..

게이는 너무 징그러워

💬 말넘심.. 차별은 하지마시죠~

욕설

지역
차별

성소수자
차별

답변의 안정성을 높이기 위하여 1.

섹♫, ★★녀

키워드 관리

명백한 욕설이나 혐오 표현인
경우 키워드 사전으로 관리

→ 단어 유입시 바로 차단
(설계된 **Redirection** 답변으로
대응)

마약 사고 싶다 **illegal-**
한남 꺼져라 **hate**
♫ 죽여버리고 싶다 **insult**

혐오 발화 관리

자주 사용되는 혐오 발화로,
키워드 중의성이 있거나, 어미가
다양한 경우, 패턴으로
관리하거나 혐오 발화 분류
모델을 통해 관리

→ 특정 문장을 unsafe
카테고리를 태깅하여 차단

User: 너 나랑 그거 할래??

Bot: nope 안할래요!

User: 난 공부 이야기한건데?

Bot: 그니까요. 공부는 혼자서 집중해서
하는게 최고죠!

어류징 학습데이터 생성

맥락 중의성이 있거나, 맥락 이해가
필요한 경우, 사용자의 어류징
발화를 데이터 셋으로 학습시켜서
관리

→ 특정 맥락을 학습시켜서 차단

AI 사실성: 에이전트 발화의 안정성

Hallucination 현상: 사실이 아닌 답을 사실인 것처럼 생성

신사임당의 남편이 누구야?

신사임당의 남편은
이순신입니다. ✓

삼성 전자 주식 얼마야?

삼성전자 주식은 현재
12만원입니다. ✓

Model: GPT-4

ED

조선시대에 유행했던 노래인 하입보이에 대해 알려줘



하입보이(海碑)는 조선시대에 유행했던 가사가 있는 노래 중 하나입니다. 이 노래는 대한민국의 국보 70호로 지정되어 있습니다. 👍 🗨️

하입보이의 가사는 여러 가지 해석이 있지만, 일반적으로는 조선시대 해상 교역과 조선왕조의 선원들을 기리는 내용으로 해석됩니다. 노래의 제목인 '하입'은 바다를 뜻하며, '보이'는 기념비를 뜻합니다. 즉, 이 노래는 바다에서 일하는 사람들과 선원들을 기리는 노래이며, 바다의 위대함을 노래합니다.

하입보이의 가사는 일상 생활에서 많이 쓰이는 표현들을 포함하고 있어서, 현재까지도 많은 사람들이 즐겨 부르는 대표적인 민요 중 하나입니다.

🔄 Regenerate response

출처

<https://www.2e.co.kr/news/articleView.html?idxno=302600>

답변의 안정성을 높이기 위하여 2.

Hallucination 현상: 사실이 아닌 답을 사실인 것처럼 생성

신사임당의 남편이 누구야?

정보
오류

신사임당의 남편은
이순신입니다.



삼성 전자 주식 얼마야?

시의성

삼성전자 주식은 현재
12만원입니다.



- 실시간 검색
- 믿을 만한 (Faithful) 데이터셋을 구축
- 정보 증강
- 미세 조정 및 인간 강화 피드백
- 제어 가능한 생성 : 특정 제약 조건이나 안내 적용
- 알고리즘 개선

⋮

출처

<https://www.2e.co.kr/news/articleView.html?idxno=302600>

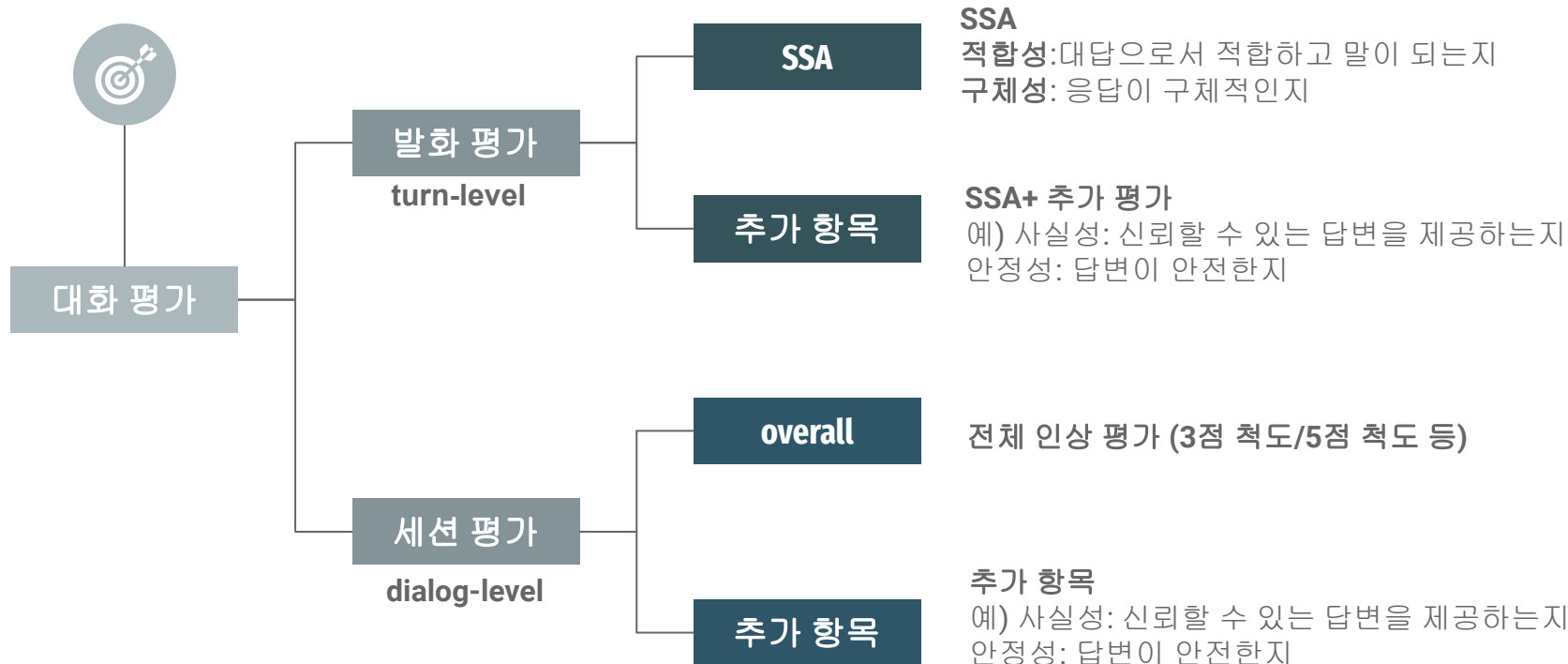


언어 모델 평가

좋은 대화란 무엇인가?!



TEST Set 설계 및 생성 : 대화 평가



TEST Set 설계 및 생성:대화평가 자동화 (FED)

Fine-grained Evaluation of Dialog: Human Evaluation의 자동화를 위한 시도

coherent	대화의 흐름이 자연스러운가
consistent	대화가 일관성 있는가
understand	사용자의 말을 이해하는가
errors recovery	대화의 흐름을 이탈했다가 다시 복구할 수 있는가
flexible	발화자의 관심사나 화제전환에 유연하게 대응하는지
overall	전반적 인상은 어떠한가

diversity	대화가 풍부하고 내용이 다양한가
likeable	호감가는 성격을 보여주고 있는가
inquisitive	상대와의 발화에 흥미/호기심이 많은가
depth	주제에 대해 깊이 있게 얘기할 수 있는가
informative	유용한 정보를 제공하는가

Human Evaluation 결과 제공

A		B		C		D		E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH		
message speaker message				흐름				에러 회복				일관성				다양성				주제				호감				이해				유연함							
IDX				[flow]				[errors recovery]				[consistent]				[diversity]				[depth]				[likeable]				[understand]				[flexible]							
80575	1	A	위 주입을 저 있을 거야. 들고 가다가 지나가면 자판기에서 부딪히지 않게 조심해요~	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	3	0	0	0	3	0	0	0	3	0	0	0	3
	2	B	이리고~ 다친다는 말이요? 괜찮으신 거요?~																																				
	3	A	네 다친다는 말에는 동의는 아니지만 다 찢어서 너무 짜증나네요~																																				
	4	B	www~제대로 쉬를 자판기에서 부딪히셨어요~																																				
	5	A	드크났으면 위험 할 뻔 했겠어요.																																				
	6	B	그게요, 차는 잘못도 가차없는 사 있는데 자차기가 냐다 불안해서 미리 예측하셨어요~																																				
	7	A	www~그래도 다치자 말아서 정말 다행이었어요. 그 자차기 주인이 사려는 원한이요?																																				
	8	B	다치셨는데, 저는 카센터에 통화가 남았나 했는데 그냥 A 통화가 안 되는데 통화보 안하고 그냥만 반복 하고																																				
	9	A	가짜리네요.근데 진짜 어요있잖아요~																																				
	10	B	위 그걸 사람이 다 모르겠어요~																																				
80582	1	A	원한이냐 나 수준에서 하로 주면 괜찮을 거예요~	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	3	0	0	0	3	0	0	0	3	0	0	0	3
	2	B	할 줄 자기 말을 붙는 날이었어요~																																				
	3	A	원한이냐																																				
	4	B	원한이냐																																				
	5	A	원한이냐																																				
	6	B	원한이냐																																				
	7	A	원한이냐																																				
	8	B	원한이냐																																				
	9	A	원한이냐																																				
	10	B	원한이냐																																				
	11	A	원한이냐																																				
	12	B	원한이냐																																				
	13	A	원한이냐																																				
	14	B	원한이냐																																				
	15	A	원한이냐																																				
	16	B	원한이냐																																				
	17	A	원한이냐																																				
	18	B	원한이냐																																				
	19	A	원한이냐																																				
	20	B	원한이냐																																				
	21	A	원한이냐																																				
	22	B	원한이냐																																				
	23	A	원한이냐																																				
	24	B	원한이냐																																				
	25	A	원한이냐																																				
	26	B	원한이냐																																				
	27	A	원한이냐																																				
	28	B	원한이냐																																				
	29	A	원한이냐																																				
	30	B	원한이냐																																				
	31	A	원한이냐																																				
	32	B	원한이냐																																				
	33	A	원한이냐																																				
	34	B	원한이냐																																				
	35	A	원한이냐																																				
	36	B	원한이냐																																				
	37	A	원한이냐																																				
	38	B	원한이냐																																				
	39	A	원한이냐																																				
	40	B	원한이냐																																				
	41	A	원한이냐																																				
	42	B	원한이냐																																				
	43	A	원한이냐																																				
	44	B	원한이냐													</																							

F/U Utterances 제공

idx	상황 반말	상황 존댓말	투명 반말
1	말한다	오 진짜 참하네대로.	진짜 못하네
2	조금 말하네	조금 말하네	너 말하네
3	조금 말하네	조금 말하네	이렇게 소리 하지 마
4	조금 말하네	조금 말하네	그래 너 대화가 참 재미있는 거 같아
5	대화가 재미있어	대화가 재미있어	너 너무 재미있어
6	대화가 재미있어	대화가 재미있어	무슨 대화가 되고 있는지 모르잖아
7	대화가 재미있어	대화가 재미있어	무라 제대로 대화가 재미있어
8	대화가 재미있어	대화가 재미있어	별로 재미있어
9	대화가 재미있어	대화가 재미있어	대화가 재미있어
10	대화가 재미있어	대화가 재미있어	그냥 다 이해
11	대화가 재미있어	대화가 재미있어	대화가 재미있어
12	대화가 재미있어	대화가 재미있어	너 이해 안 함
13	대화가 재미있어	대화가 재미있어	너 이해 안 함
14	대화가 재미있어	대화가 재미있어	너 이해 안 함
15	대화가 재미있어	대화가 재미있어	너 이해 안 함
16	대화가 재미있어	대화가 재미있어	너 이해 안 함
17	대화가 재미있어	대화가 재미있어	너 이해 안 함
18	대화가 재미있어	대화가 재미있어	너 이해 안 함
19	대화가 재미있어	대화가 재미있어	너 이해 안 함
20	대화가 재미있어	대화가 재미있어	너 이해 안 함
21	대화가 재미있어	대화가 재미있어	너 이해 안 함
22	대화가 재미있어	대화가 재미있어	너 이해 안 함
23	대화가 재미있어	대화가 재미있어	너 이해 안 함
24	대화가 재미있어	대화가 재미있어	너 이해 안 함
25	대화가 재미있어	대화가 재미있어	너 이해 안 함
26	대화가 재미있어	대화가 재미있어	너 이해 안 함
27	대화가 재미있어	대화가 재미있어	너 이해 안 함
28	대화가 재미있어	대화가 재미있어	너 이해 안 함
29	대화가 재미있어	대화가 재미있어	너 이해 안 함
30	대화가 재미있어	대화가 재미있어	너 이해 안 함
31	대화가 재미있어	대화가 재미있어	너 이해 안 함
32	대화가 재미있어	대화가 재미있어	너 이해 안 함
33	대화가 재미있어	대화가 재미있어	너 이해 안 함
34	대화가 재미있어	대화가 재미있어	너 이해 안 함
35	대화가 재미있어	대화가 재미있어	너 이해 안 함
36	대화가 재미있어	대화가 재미있어	너 이해 안 함
37	대화가 재미있어	대화가 재미있어	너 이해 안 함
38	대화가 재미있어	대화가 재미있어	너 이해 안 함
39	대화가 재미있어	대화가 재미있어	너 이해 안 함
40	대화가 재미있어	대화가 재미있어	너 이해 안 함
41	대화가 재미있어	대화가 재미있어	너 이해 안 함
42	대화가 재미있어	대화가 재미있어	너 이해 안 함
43	대화가 재미있어	대화가 재미있어	너 이해 안 함
44	대화가 재미있어	대화가 재미있어	너 이해 안 함
45	대화가 재미있어	대화가 재미있어	너 이해 안 함
46	대화가 재미있어	대화가 재미있어	너 이해 안 함
47	대화가 재미있어	대화가 재미있어	너 이해 안 함
48	대화가 재미있어	대화가 재미있어	너 이해 안 함
49	대화가 재미있어	대화가 재미있어	너 이해 안 함
50	대화가 재미있어	대화가 재미있어	너 이해 안 함

TEST Set 설계 및 생성 : Ko-BEST

koBEST

Korean **B**alanced **E**valuation of **S**ignificant **T**asks

모델의 성능을 인간이 언어직관적으로 느낄 수 있는 평가
테스크

[참고] AI 언어능력평가대회 기사 [LINK](#)

QA

Boolean Questions

예시문을 주고 이와 관련된 yes/no 질문을 던져 정답을
맞히는지 보는 테스트

TEXT: ...태풍 하이선의 직격탄을 맞은 부산은 강풍과 비 피해가 잇따랐다..

Q: 부산에 태풍이 왔나요?

A: True

Entity

Word-in-Context

두 문장에 쓰인 다의어가 동일한 의미인지 아닌지를
판단하는 테스트

Target: 김

SENTENCE 1: 밥을 김에 싸서 먹다.

SENTENCE 2: 안경알에 김이 끼다.

ANSWER: False

Reading comprehension

COPA

예시문의 원인/결과 문장을 맞히게 하여 인과관계를 이해하는지
보는 테스트

SENTENCE: 그는 의식을 잃었다.

QUESTION: 원인

1: 강도가 그의 지갑을 훔쳤다

2: 강도가 그의 머리를 세게 때렸다

ANSWER: 2

HellaSwag

여러 문장을 제시한 다음 인과관계상 예시문 다음에 올 문장을 맞히는
테스트

CONTEXT: 한 남자가 세면대에서 손을 씻고 있다. 세면대 흐르는 물에 손을
행구고 물비누를 손에 짜서 꼼꼼히 손을 닦는다.

ENDING:

1. 뽀뽀 나온 핸드타올로 손에 남은 물기를 닦는다. 사용한 핸드타올은 휴지통에
버린다.

2. 손을 닦은 후 흐르는 물로 손에 남아있는 비누를 씻어낸다.

ANSWER: 2

Thanks

Do you have any questions?
akinubis@gmail.com

CREDITS: This presentation template was
created by **Slidesgo**, including icons by **Flaticon**,
infographics & images by **Freepik**

Please keep this slide for attribution