

초거대 언어모델의 현재와 미래

NAVER CLOVA

Language Representation

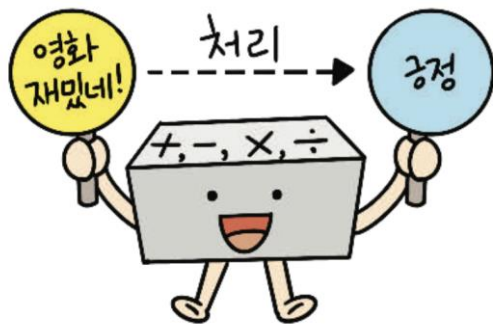
이기창

목차

- 언어모델 소개
- HyperCLOVA 학습 경험
- HyperCLOVA의 현재
- HyperCLOVA의 미래

언어모델

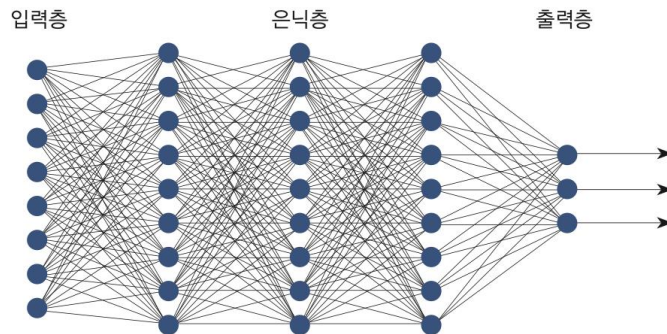
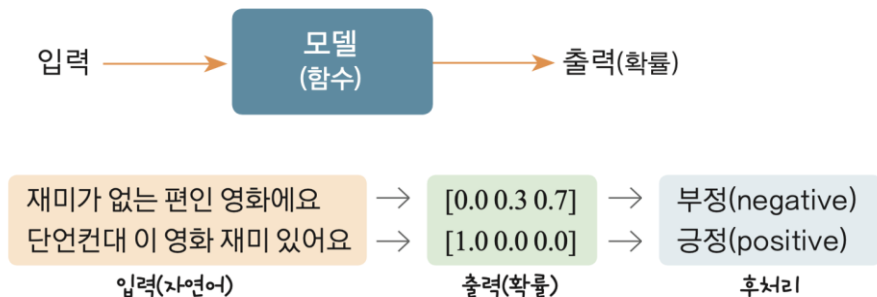
- 자연어 처리



언어모델

- 자연어 처리 모델

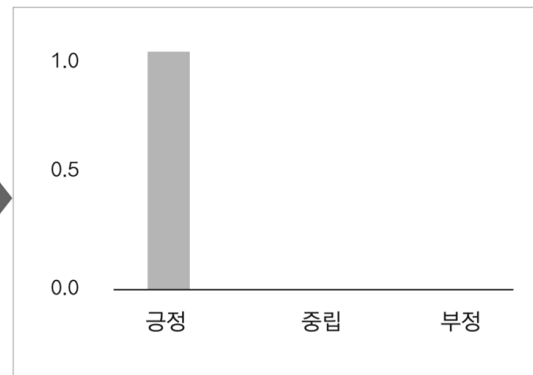
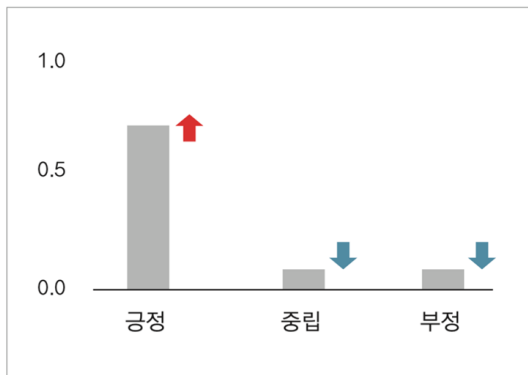
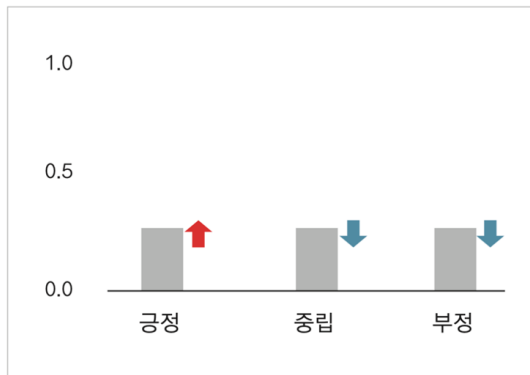
- 입력 : 자연어(텍스트)
- 출력 : 해당 입력이 특정 범주에 속할 확률
- 딥러닝 기반
 - 수많은 행렬, 벡터로 구성
 - 최근엔 트랜스포머^{transformer}가 대세



언어모델

- 딥러닝 모델의 학습
 - 출력이 레이블(정답)과 비슷하도록 모델 업데이트
 - 모델이 스스로 학습데이터 패턴을 익히는 과정

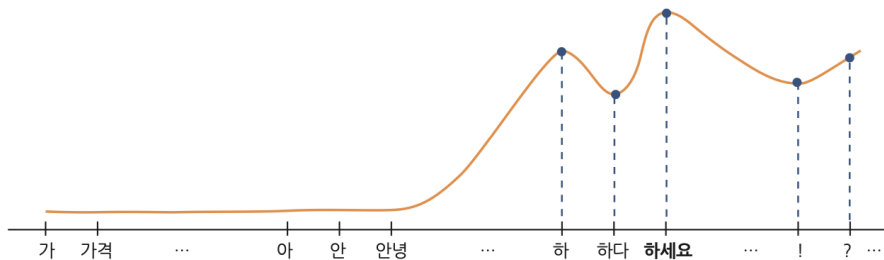
문장	감성		
	긍정	중립	부정
단언컨대 이 영화 재미 있어요	1	0	0
단언컨대 이 영화 재미 없어요	0	0	1
...	...	...	...



언어모델

- 언어모델 : 이전 단어들이 주어졌을 때 다음 단어 확률 부여
 - 입력 : 단어(토큰) 시퀀스
 - 출력 : 입력 다음으로 어떤 단어가 그럴듯한지(자연스러운지) 나타내는 확률
 - 이때 범주 수는 해당 언어의 전체 어휘(토큰) 수와 동일
 - 자연스런 표현에 더 높은 확률 : $P(\text{운전} | \text{난폭}) > P(\text{운전} | \text{무모})$

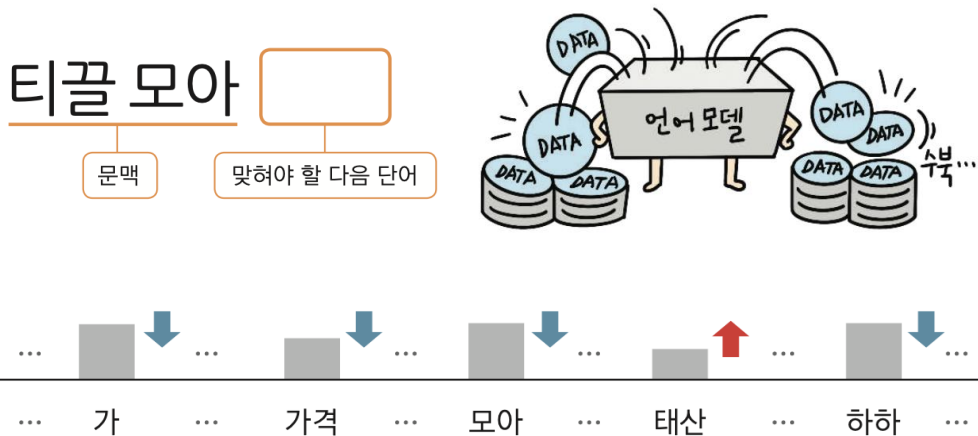
입력	다음 토큰	등장 확률
안녕	가	0.002
	가격	0.0001
	...	...
	아	0.0001
	안	0.0002
	안녕	0.0003
	...	...
	하	0.2
	하다	0.15
	하세요	0.3
	...	...
	!	0.16
	?	0.18
	...	...



언어모델

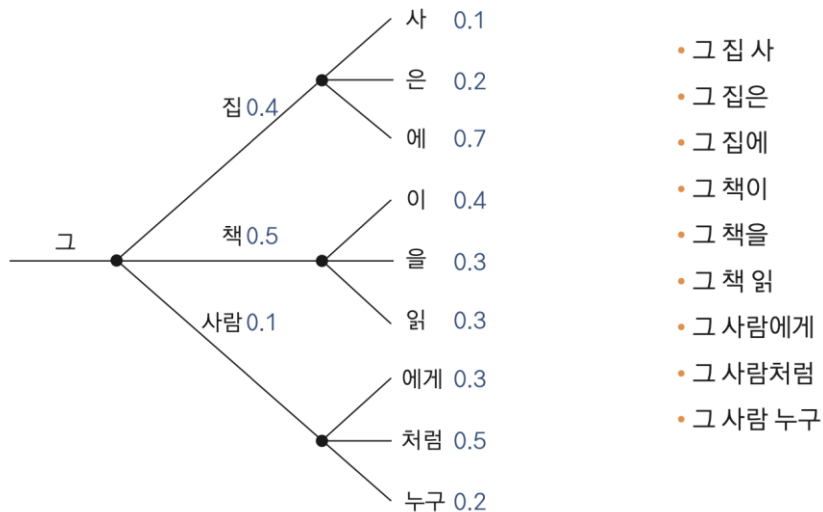
- 언어모델의 학습

- 입력 시퀀스가 주어졌을 때 정답(다음 단어)을 맞출 수 있도록 모델 업데이트
- 레이블을 수작업으로 만들 필요 없어 다량의 학습데이터 확보 가능해짐



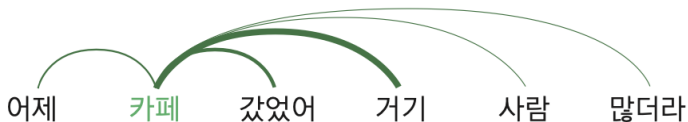
언어모델

- 언어모델의 문장 생성
 - 반복적으로 다음 토큰을 선택해 나가는 과정
 - 선택 방법 다양 : 1등, 특정 범위 내 랜덤 등
 - 따라서 무수히 많은 문장을 만들어 낼 수 있음



언어모델

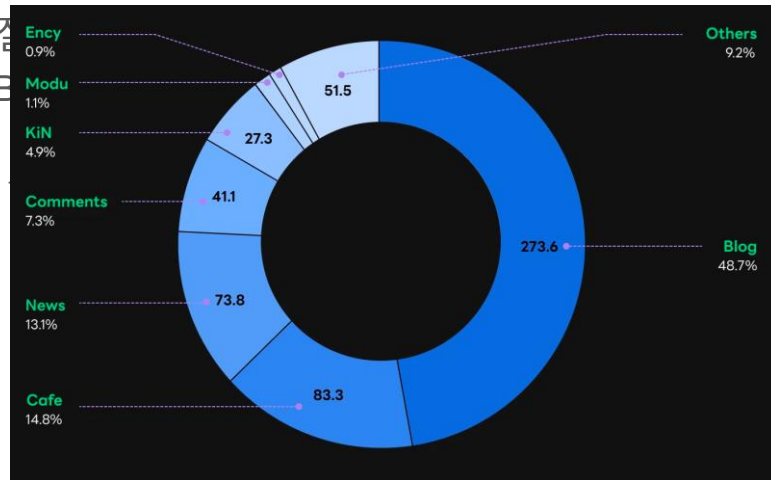
- 트랜스포머transformer
 - 성능, 효율 면에서 주목받고 있는 모델 구조
 - HyperCLOVA는 트랜스포머 기반 거대 언어모델
 - Self-attention
 - 모든 단어-단어 쌍 사이의 의미적 연관성 고려
 - 입력된 문장을 입체적으로 분석하는 능력 있음



HyperCLOVA 학습 경험

- 데이터 확보 및 전처리

- 언어모델의 성능을 좌우하는 중요 요소 : 말뭉치 품질
- 네이버 블로그, 카페 등 양질의 한국어 말뭉치 1.8TB 확보
- 말뭉치 품질 향상 및 개인정보 처리
 - 개인정보 치환 : 주민등록번호, 이메일, 휴대전화번호, 계좌번호
 - 자소 단위 혹은 의미 없는 문구 제거
 - 비속어 키워드가 포함된 문서 제거
 - 웹 문서에서 제목과 본문 첫부분이 중복되는 경우 제목을 제거
 - 지식인 등에서 사용자 레벨 등 반영해 일정 이하 문서 제거
 - 중복 문서 제거



HyperCLOVA 학습 경험

- 데이터 확보 및 전처리

- 입력 문서를 입체적으로 분석하는 셀프 어텐션self-attention 특성 활용[1]
- 문서 앞에 컬렉션 정보 등을 추가해 생성될 문장의 품질 향상 도모
 - 특정 태스크에서 성능zero-shot in-context learning 증가

컬렉션명	주요 내용	처리 방식
book	책	문서\n제목: \${제목}\n본문: \${본문}
academic	전문정보-논문초록	문서\n제목: \${제목}\n본문: \${본문}
wiki	위키백과	문서\n제목: \${제목}\n본문: \${본문}
lkc	지식백과	문서\n제목: \${제목}\n본문: \${본문}
nikl-news	모두의알용치-뉴스	뉴스\n제목: \${제목}\n본문: \${본문}
nikl-web	모두의알용치-웹	웹\n제목: \${제목}\n본문: \${본문}
nikl-written	모두의알용치-문어	문서\n제목: \${제목}\n본문: \${본문}
nikl-spoken	모두의알용치-구어	대화\n제목: \${제목}\n본문: \${본문}
nikl-messenger	모두의알용치-메신저	대화\n제목: \${제목}\n본문: \${본문}
web	웹	웹\n제목: \${제목}\n본문: \${본문}
kin	지식인	질의응답\n질문: \${질문}\n답변: \${답변}
cafe	카페	카페\n제목: \${제목}\n본문: \${본문}
blog	블로그	블로그\n제목: \${제목}\n본문: \${본문}
news	뉴스	뉴스\n제목: \${제목}\n본문: \${본문}
cafeComment	카페댓글	대화\n제목: \${제목}\n본문: \${본문}
upcComment	공통댓글	대화\n제목: \${제목}\n본문: \${본문}
shoppingReview	쇼핑리뷰	리뷰\n제목: \${제목}\n본문: \${본문}
smartstoreReview	스마트스토어리뷰	리뷰\n제목: \${제목}\n본문: \${본문}
productReview	상품리뷰	리뷰\n제목: \${제목}\n본문: \${본문}
keywordAD	키워드 광고	광고\n제목: \${키워드}\n본문: \${본문}

CLOVA Studio플레이그라운드익스플로리 가이드포럼

32 / 2048

EngineBoss

Generation typeRollingOne-timeMultipleExamples

제목을 입력해주세요.

질의응답
질문: 여드름은 왜 나나요?
본문: 피지가 과다하게 분비되어 모공이 막히고 그 안에서 염증이 생기는 것입니다.

Top P0.8

Top K0

Maximum tokens32

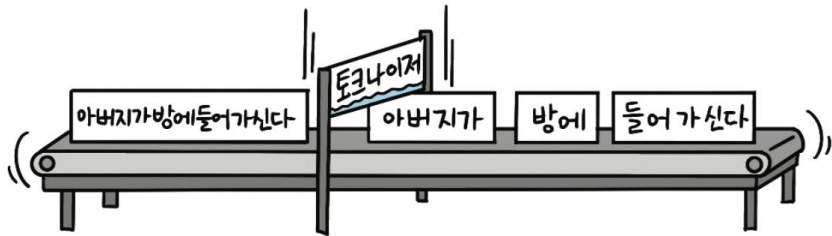
Temperature0.5

Repetition penalty5

HyperCLOVA 학습 경험

- 한국어에 맞는 토큰화 기법

- 토큰token : 자연어 처리 모델이 인식하는, 의미의 최소 단위
- 토큰화tokenization : 문자열string을 분석하는 과정 혹은 그 결과
- 토큰나이저tokenizer : 문자열을 토큰으로 분석하는 프로그램



HyperCLOVA 학습 경험

- 한국어에 맞는 토큰화 기법
 - HyperCLOVA^[1]에서는 형태소 단위 Byte Pair Encoding 적용
 - BPE 기법은 자주 등장한 문자열을 토큰으로 먼저 분석
 - 조사, 어미 빈도가 높기 때문에 이들이 우선적으로 분석됨
 - 미등록 토큰^{unknown token} 발생 막기 위해 Byte-level BPE 적용

Character-level Byte Pair encoding:

프랑스는
프랑스가
퍼포먼스는
퍼포먼스가
카타르시스는
카타르시스가
깁스는
깁스가
키스는
키스가
...

→ 스는, 스가

Byte-level Byte Pair encoding:

í꺠Ĥĕŕſijĩᄇēĩᄇ (프랑스는)
í꺠Ĥĕŕſijĩᄇē°ᄇ (프랑스가)
ĩᄇĩŕĩ—ē°°ᄇĩᄇēĩᄇ
ĩᄇĩŕĩ—ē°°ᄇĩᄇē°ᄇ → ᄇĩᄇēĩᄇ (스는),
ᄇĩᄇē°ᄇ (스가)
ĩ°°ŕĩᄇē°°ŕĩᄇĩᄇēĩᄇ
ĩ°°ŕĩᄇē°°ŕĩᄇĩᄇē°ᄇ
ē°ᄇĩᄇēĩᄇ
ē°ᄇĩᄇē°ᄇ
ĩᄇᄇĩᄇēĩᄇ
ĩᄇᄇĩᄇē°ᄇ
...

Morpheme-Aware Byte-level Byte Pair encoding:

í꺠Ĥĕŕſijĩᄇ ēĩᄇ (프랑스는)
í꺠Ĥĕŕſijĩᄇ ē°ᄇ (프랑스가)
ĩᄇĩŕĩ—ē°°ᄇĩᄇēĩᄇ
ĩᄇĩŕĩ—ē°°ᄇĩᄇē°ᄇ → ēĩᄇ (는),
ē°ᄇ (가)
ĩ°°ŕĩᄇē°°ŕĩᄇĩᄇēĩᄇ
ĩ°°ŕĩᄇē°°ŕĩᄇĩᄇē°ᄇ
ē°ᄇĩᄇēĩᄇ
ē°ᄇĩᄇē°ᄇ
ĩᄇᄇĩᄇēĩᄇ
ĩᄇᄇĩᄇē°ᄇ
...

[1] [What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA](#)

HyperCLOVA 학습 경험

- 한국어에 맞는 토큰화 기법
 - 형태소 기반 BBPE가 기계독해, 번역 등 벤치마크 태스크에서 좋은 성능 기록^[1]

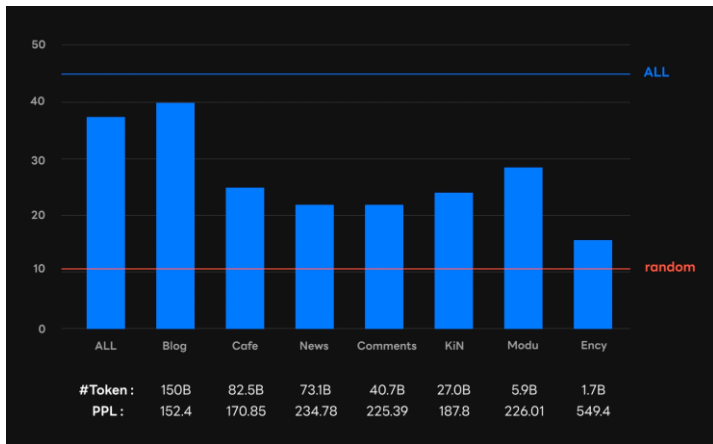
	KorQuAD (EA / F1)		AI Hub (BLEU) Ko→En En→Ko		YNAT (F1)	KLUE-STS (F1)
Ours	55.28	72.98	3.83	20.03	58.67	60.89
byte-level BPE	51.26	70.34	4.61	19.95	48.32	60.45
char-level BPE	45.41	66.10	3.62	16.73	23.94	59.83

Table 6: Effects of tokenization approaches on five tasks. HyperCLOVA-1.3B is used for evaluation.

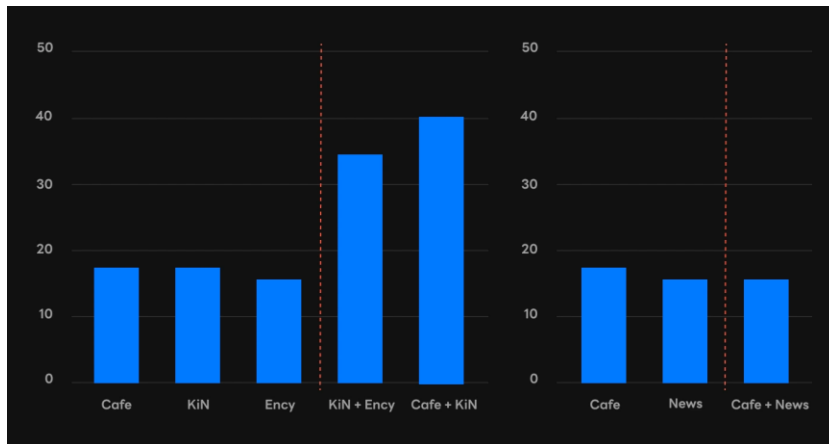
[1] [What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA](#)

HyperCLOVA 학습 경험

- 언어모델 능력 배가를 위한 말뭉치 비율 탐색
 - HyperCLOVA^{[1], [2]}에 따르면 말뭉치를 잘 섞으면 없던 능력이 생기기도 함
 - 이미 잘 구성되어 있는 말뭉치(Modu, Blog 등)라면 능력이 상대적으로 잘 발현됨



모델별 인컨텍스트 제로샷 러닝 성능 (5개 한국어 과제 평균)



모델별 인컨텍스트 퓨샷 러닝 성능 (5개 한국어 과제 평균)

[1] [On the Effect of Pretraining Corpora on In-context Learning by a Large-scale Language Model](#)

[2] [말뭉치의 종류에 따른 HyperCLOVA 인컨텍스트 러닝 성능 변화](#)

HyperCLOVA 학습 경험

- 거대 언어모델의 유용성
 - 모델 크기(파라미터 수)가 클수록 태스크를 잘 수행
 - 영어 모델[1], [2]뿐 아니라 한국어 모델[3]에서도 확인

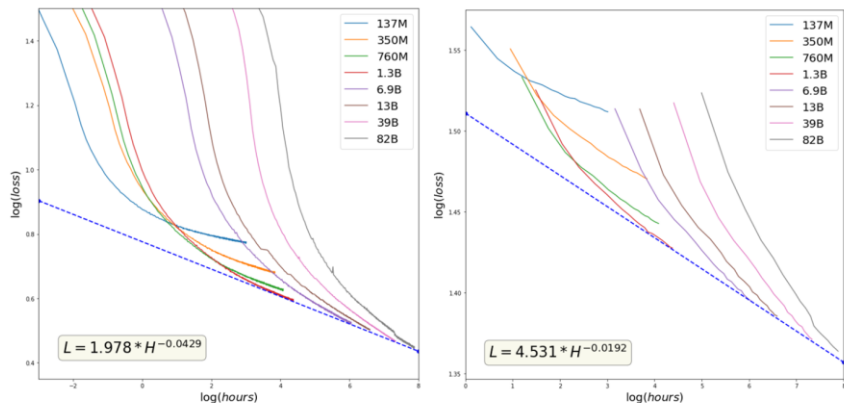


Figure 2: Scaling law (Kaplan et al., 2020; Brown et al., 2020) in training HyperCLOVA models with various parameters. The left figure presents the training and the right figure shows the loss on the testset of the Korean encyclopedia not contained in the training corpus.

[1] [Scaling laws for autoregressive generative modeling](#)

[2] [Language models are few-shot learners](#)

[3] [What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA](#)

HyperCLOVA 학습 경험

- 거대 언어모델 학습은 그 자체로 도전적
 - 가장 어려운 일은 학습 중요 지표(train loss)가 발산하는 현상
 - OPT^[1] 등에서도 이같은 현상 때문에 고생을 했던 것으로 보임

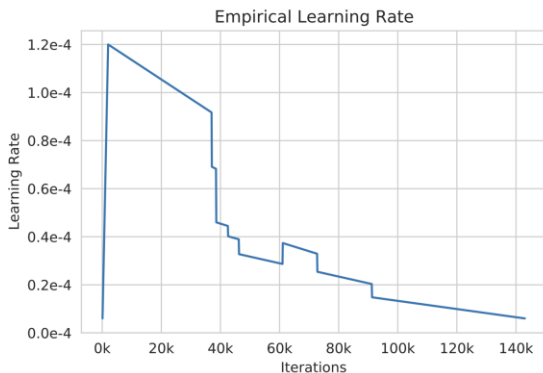
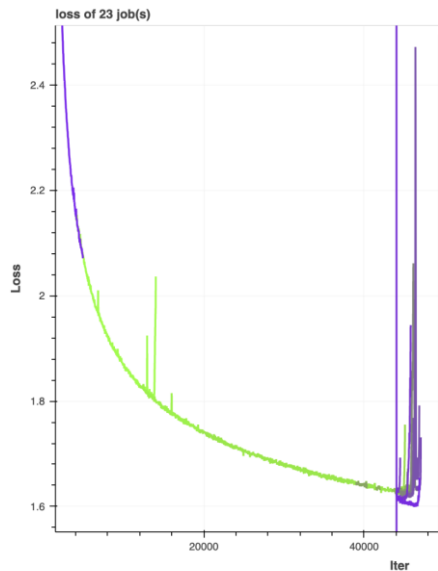
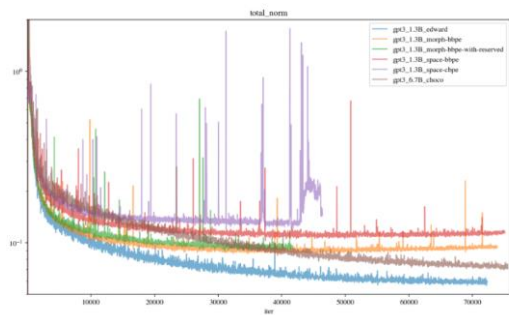
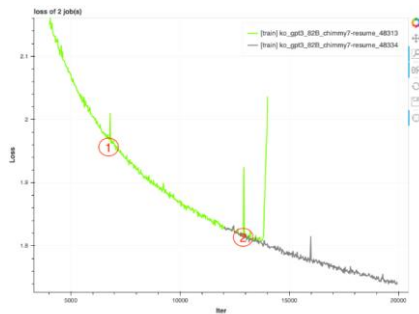
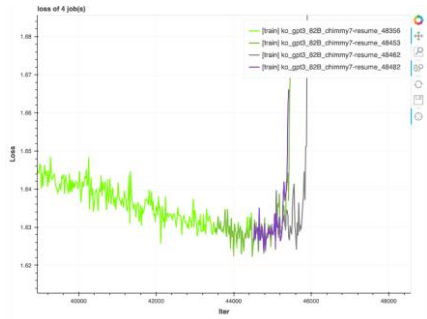
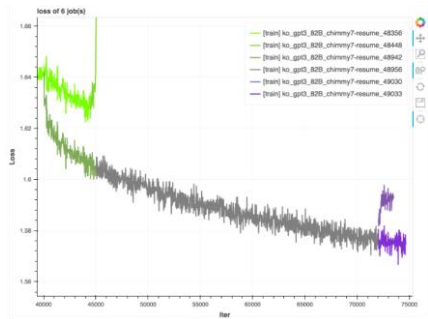


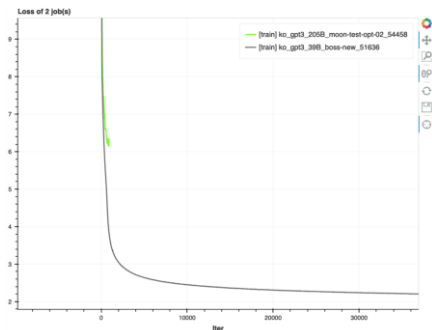
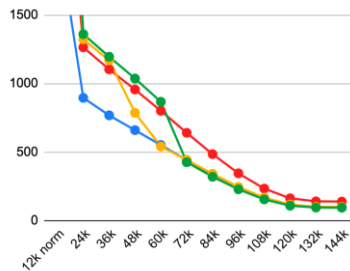
Figure 1: **Empirical LR schedule.** We found that lowering learning rate was helpful for avoiding instabilities.

HyperCLOVA 학습 경험

- 거대 언어모델 학습은 그 자체로 도전적
 - 수많은 시도 끝에 다양한 크기의 HyperCLOVA 생산 성공

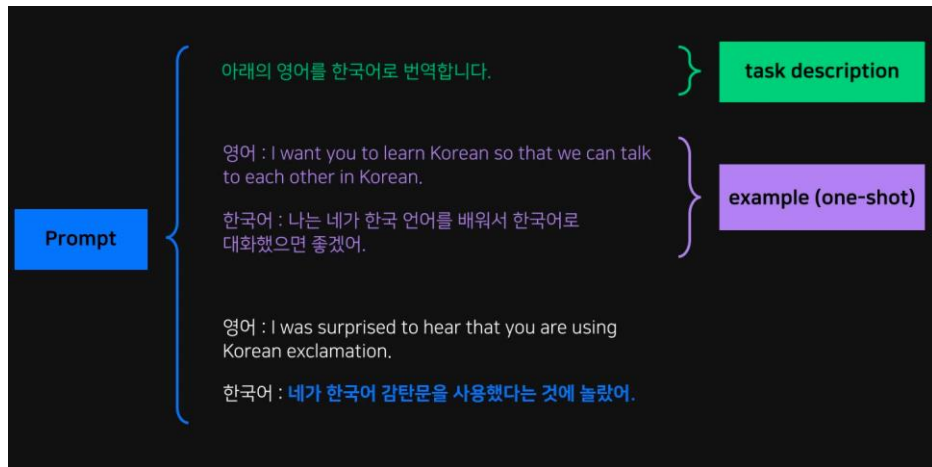


step 간 param dist



HyperCLOVA의 현재

- 인컨텍스트 러닝 In-context learning, [1]
 - 언어모델이 거대해진 후 발견된 현상
 - 언어모델이 지시문, 예제 등 입력 문서(context, prompt)의 의미를 파악한 후
 - 요약, 번역, 대화 등 구체적인 과제(downstream task)를 해결하는 패러다임
 - 파인튜닝(fine-tuning) 등과 달리 모델을 업데이트하지 않음
 - 다음은 한국어-영어 번역 과제에서의 인컨텍스트 원샷(1-shot) 러닝 예시



[말뭉치의 종류에 따른 HyperCLOVA 인컨텍스트 러닝 성능 변화](#)

HyperCLOVA의 현재

- 인컨텍스트 제로샷zero-shot 러닝
 - 예제 없이 바로 태스크 수행

CLOVAStudio^β 플레이그라운드 익스플로러 가이드 포럼

← ↺ 🗑 32 / 2048

Engine

🇰🇷 Boss ▾

Top P 0.8

Top K 0

Maximum tokens 32

Temperature 0.5

Repetition penalty 5

Generation type ☒ Rolling ☐ One-time ☐ Multiple ☐ Examples

제목을 입력해주세요.

질의응답
질문: 여드름은 왜 나나요?
본문: 피지가 과다하게 분비되어 모공이 막히고 그 안에서 염증이 생기는 것입니다.

HyperCLOVA의 현재

- 인컨텍스트 퓨샷^{few-shot} 러닝
 - 몇 개 예제를 바탕으로 태스크 수행

CLOVAStudio[®]

플레이그라운드 튜닝 익스플로러 가이드 포럼

275 / 2048

Engine

✖ Boss

Top P0.8

Top K0

Maximum tokens32

Temperature0.5

Repetition penalty5

Stop sequences

시퀀스 입력 후 Tab

Inject start text

Inject restart text

Generation type

☒ Rolling ☐ One-time ☐ Multiple ☐ Examples

제목을 입력해주세요.

다음 키워드를 바탕으로 창의적인 마케팅 문구를 생성하세요.

키워드: 오피스룩 오버핏 루즈핏 20대 30대 40대 데일리룩 데일리코디 일상룩 출근룩
문구: 매일 입어도 질리지 않는 심플하고 세련된 디자인의 스타일링

키워드: 캠핑러그 캠핑자충매트 차박매트 복유림불랭킷 쇼파천커버 자충매트 캠핑차박매트 에스닉러그 인테리어러그 물세탁러그
문구: 감성캠핑러그로 분위기를 살려보자.

키워드: 크리스마스룩 롱부츠 앵클부츠 겨울슈즈 가죽부츠 삭스앵글 크리스마스코디 데이트룩 크로스백 토트백
문구: 연말모임 파티룩 완성!

키워드: 쉬폰스커트 레이스스커트 주름스커트 밴딩스커트 메메이드스커트 벨트자켓 볼자켓 루즈핏자켓 플라워원피스 쉬폰원피스
문구: 여성스럽고 시크한 매력을 동시에 느낄수 있는 아이템으로 멋진 스타일링을 해보세요.

키워드: 앵클부츠 싸이하이부츠 슬림온 스니커즈 여성스니커즈 여성앵클부츠 롱부츠 여성슬림온 여성슬림백
문구: 20대부터 50대까지 취향저격 아이템들만 모았다!

키워드: 백팩 롱원피스 플라워원피스 쉬폰원피스 점프슈트
문구: 사랑스러운 원피스로 스타일업!

HyperCLOVA의 현재

- 인컨텍스트 러닝의 활용
 - 자유 대화 Open-Domain Dialogue

The screenshot displays the CLOVA Studio web interface. The top navigation bar includes the CLOVA Studio logo and links for Playground, Tuning, Explorer, Guide, and Forum. On the left, a sidebar contains configuration sliders for Engine (set to Boss), Top P (0.8), Top K (0), Maximum tokens (32), Temperature (0.5), and Repetition penalty (5). The main area on the right shows the 'Generation type' set to 'Rolling'. Below this, a large text input field prompts the user to '제목을 입력해주세요.' (Please enter a title). A chat log on the right shows a conversation between a child and a character named Pungsu, discussing EBS and school.

CLOVA Studio^β 플레이그라운드 튜닝 익스플로러 가이드 포럼

161 / 2048

Engine
Boss

Top P 0.8

Top K 0

Maximum tokens 32

Temperature 0.5

Repetition penalty 5

Generation type ☒ Rolling ☐ One-time ☐ Multiple ☐ Examples

제목을 입력해주세요.

아래는 AI 펭수와 어린이의 대화입니다.

어린이: 펭수야 EBS는 어때?
펭수: EBS에서 내가 제일 잘 나가, 오템.
어린이: 코로나 때문에 학교 안가서 넘 좋아
펭수: 그래도 어릴 때 공부해야해! 그래야 나처럼 멋져지지!
어린이: EBS 사장님이랑 참치 먹었어?
펭수: 아~ 그건 비밀인데...
어린이: 왜 비밀인가야?
펭수: 비밀이니까!!

HyperCLOVA의 현재

- 인컨텍스트 러닝의 활용
 - 과제를 수행하는 대화Task-Oriented Dialogue

The screenshot displays the HyperCLOVA Studio interface. The top navigation bar includes links for '플래그라운드', '튜닝', '익스플로러', '가이드', and '포럼'. The left sidebar contains configuration options for the 'Engine' (set to 'Boss'), 'Top P' (0.8), 'Top K' (0), 'Maximum tokens' (32), 'Temperature' (0.5), 'Repetition penalty' (5), 'Stop sequences' (with a button '시퀀스 입력 후 Tab'), and 'Inject start text' (with a text area containing a note about user input and a 'Show probabilities' button). The main chat area on the right shows a conversation titled '제목을 입력해주세요.' with a context of a reservation at AiCall. The chat history shows a dialogue between a customer and a staff member, with the staff member providing information about the reservation and the customer asking for more details. The chat window also includes a 'Generation type' selector with options for 'Rolling' (selected), 'One-time', 'Multiple', and 'Examples'.

Engine

Rolling

제목을 입력해주세요.

아웃백스테이크 하우스의 전화를 대신 받아 주는 클로바 AiCall의 대화입니다.

대화 #1

점원: 안녕하세요! 아웃백스테이크 하우스입니다.

고객: 예약 좀 하려는데요

점원: 네, 언제 오시나요?

고객: 6시 좀 가려고요

점원: 몇 분이 오세요?

고객: 아이 하나, 어른 둘이요.

점원: 6시 세 분 자리 있어요. 예약해 드릴까요?

고객: 네네

점원: 네, 전화번호 뒷자리 알려주시겠어요?

고객: 4921입니다.

점원: 네, 예약 되었습니다. 또 필요한 거 있으세요?

고객: 아니요! 감사합니다.

대화 #2

점원: 안녕하세요! 아웃백스테이크 하우스입니다.

고객: 10분 정도 있으면 도착하는데요, 주차장이 어디 있나요?

점원: 교회 건물에 주차하실 수 있으세요.

고객: 주차는 몇 시간 할 수 있나요?

점원: 2시간 무료주차 가능합니다.

HyperCLOVA의 현재

- 인컨텍스트 러닝의 활용
 - 번역

The screenshot displays the HyperCLOVA Studio web interface. The top navigation bar includes the logo and links for Playground, Tuning, Explorer, Guide, and Forum. The left sidebar contains configuration options for the AI engine (set to 'Boss'), Top P (0.8), Top K (0), Maximum tokens (32), Temperature (0.5), Repetition penalty (5), Stop sequences (a button to input prompts), and Inject start text. The main area on the right shows the 'Generation type' set to 'Rolling'. Below this, a prompt is entered: '다음 일본어 문장을 정확하게 한국어로 번역합니다.' (Accurately translate the following Japanese sentence into Korean). The interface then displays four examples of Japanese text being translated into Korean. The first example is '일본어: おはようございます !' (Japanese: Good morning!) translated to '한국어: 안녕하세요!' (Korean: Hello!). The second is '일본어: これは何ですか ?' (Japanese: What is this?) translated to '한국어: 이것은 무엇입니까?' (Korean: What is this?). The third is '일본어: 今日、天気がとてもいいですね。' (Japanese: Today, the weather is very nice, isn't it?) translated to '한국어: 오늘 날씨가 참 좋네요.' (Korean: The weather is so nice today). The fourth is '일본어: お元気ですか。' (Japanese: Are you well?) translated to '한국어: 잘 지내세요?' (Korean: How are you?).

CLOVA Studio[®] 플레이그라운드 튜닝 익스플로러 가이드 포럼

Engine
Boss

Top P 0.8

Top K 0

Maximum tokens 32

Temperature 0.5

Repetition penalty 5

Stop sequences
시퀀스 입력 후 Tab

Inject start text

148 / 2048

Generation type ☒ Rolling ☐ One-time ☐ Multiple ☐ Examples

제목을 입력해주세요.

다음 일본어 문장을 정확하게 한국어로 번역합니다.

일본어: おはようございます !
한국어: 안녕하세요!

일본어: これは何ですか ?
한국어: 이것은 무엇입니까?

일본어: 今日、天気がとてもいいですね。
한국어: 오늘 날씨가 참 좋네요.

일본어: お元気ですか。
한국어: 잘 지내세요?

일본어: 嬉しいです。
한국어: 기쁩니다.

HyperCLOVA의 현재

- 인컨텍스트 러닝의 활용

- 데이터 생성 도구: 축하 카드 문구 생성, 욕설이 없는 문장으로 변환하기 등

The screenshot shows the CLOVA Studio interface with the 'Boss' engine selected. The 'Generation type' is set to 'Rolling'. The 'Top P' slider is at 0.8, 'Top K' is at 0, 'Maximum tokens' is at 50, 'Temperature' is at 0.5, and 'Repetition penalty' is at 5. The 'Stop sequences' section shows '시퀀스 입력 후 Tab' and buttons for '상황: ×', '축하 메시지: ×', and '### ×'. The 'Inject start text' field is empty.

축하 카드 문구 생성

상황과 대상을 반영한 축하 카드 문구를 생성합니다.

상황 : 생일
대상 : 친구
축하 메시지: 친구야 생일축하한다! 맛있는거 많이 먹고 즐거운 하루 보내

상황: 결혼
대상: 언니
축하 메시지: 언니 결혼축하해. 결혼해서도 행복하게 지내구 ㅎㅎ

상황: 졸업
대상: 동생
축하 메시지: 고생했어~ 이제 사회인이네? 앞으로 더 열심히 하자^^

상황: 대학 합격
대상: 형
축하 메시지: 오!! 진짜?? 완전 축하해~~ 대학교 가서도 공부 열심히 하고 잘지내자ㅋㅋ

The screenshot shows the CLOVA Studio interface with the 'Boss-cooldown' engine selected. The 'Generation type' is set to 'Rolling'. The 'Top P' slider is at 0.8, 'Top K' is at 0, 'Maximum tokens' is at 32, 'Temperature' is at 0.3, and 'Repetition penalty' is at 5.

욕설 순화

주어진 원문장을 욕설이 없는 문장으로 표현하시요.

원문장: 저 새끼 남의 말을 더럽게 안 처둔더라.
욕이 없는 문장: 저 사람은 남의 말을 안 듣더라.

원문장: 밤에 별빛 존나 이쁘더라.
욕이 없는 문장: 밤에 별빛이 아름답네요.

원문장: 내일 뒤통시발놓아!
욕이 없는 문장: 내일 뒤통시주세요!

HyperCLOVA의 현재

- 인컨텍스트 러닝의 활용

- 대화를 입맛에 맞게 생성 : CoDA^{Controllable Dialogue Augmentation}

문제: 유저상황에 따른
공감하는 대화 생성

제어입력: 유저 상황 (situation)

출력: 공감 대화

유저 상황에 대한 아루와 유저의 대화를 생성하세요.

유저 상황

집이 아닌 곳에서 잠을 잘 못자겠어

아루와 유저의 대화

아루: 오늘도 좋은 하루 보내셨나요?

유저: 어, 근데 너무 피곤해.

아루: 호잉? 무슨 일 있으세요?

유저: 아니, 요즘 잠을 잘 못자서 그런가봐.

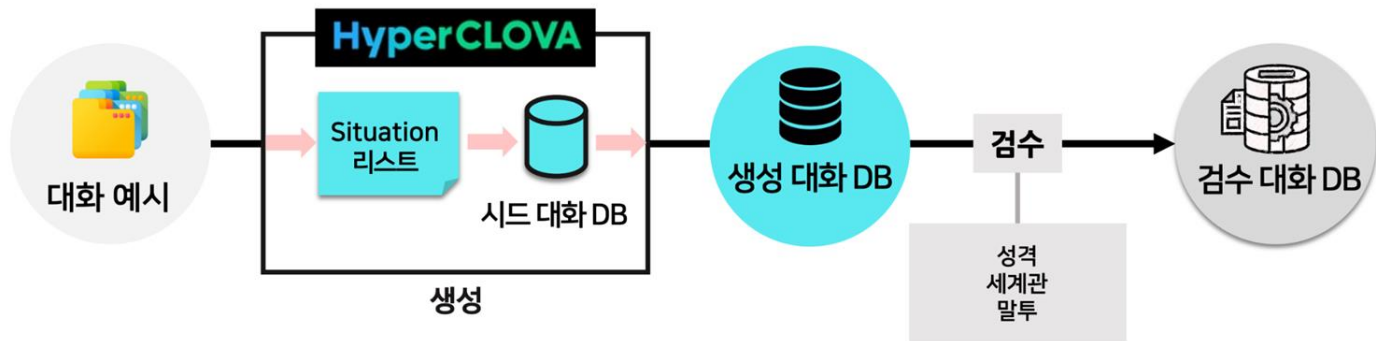
아루: 왜요? 어디 아프신 건 아니죠?

유저: 아픈 건 아닌데, 집이 아닌 곳에서 잠을 잘 못자겠어.

HyperCLOVA의 현재

- 인컨텍스트 러닝의 활용

- HyperCLOVA 활용해 대량으로 데이터를 생성한 뒤 사람이 검수
- 레이블 데이터를 처음부터 수작업으로 만드는 것보다 비용 효율적

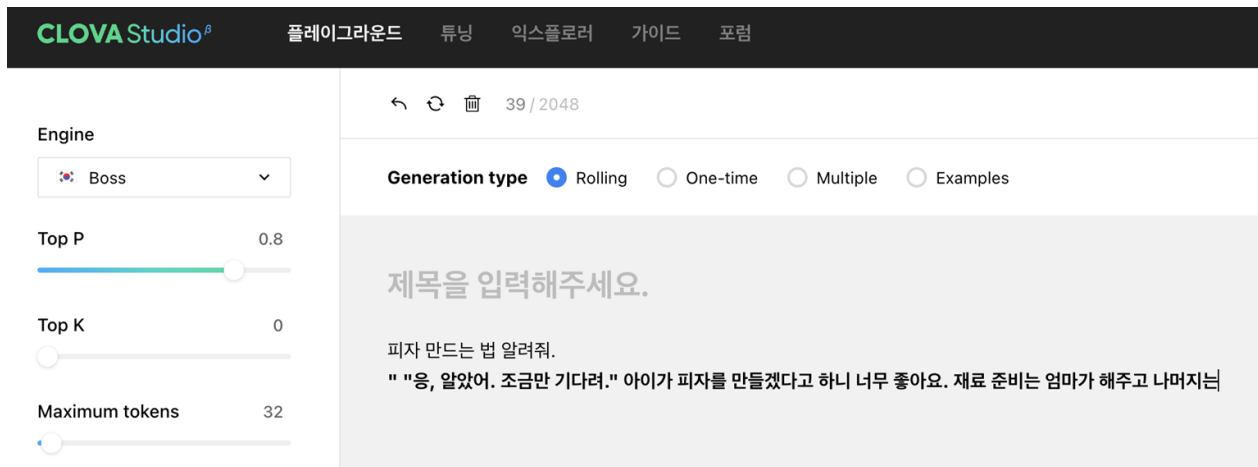


HyperCLOVA의 현재

- 인컨텍스트 러닝의 한계
 - 예제가 악성(성희롱, 욕설 등)일 경우 생성되는 문장도 그럴 가능성 높음

HyperCLOVA의 현재

- 인컨텍스트 러닝의 한계
 - 예제가 하나도 없는 경우 zero-shot learning 성능이 그다지 높지 않음



HyperCLOVA의 현재

- 인컨텍스트 러닝의 한계

- 프롬프트 구성에 따라 인컨텍스트 러닝 성능이 크게 달라짐

- 심지어 동일한 프롬프트 포맷이라 하더라도 예제 수나 순서에 민감^[1]

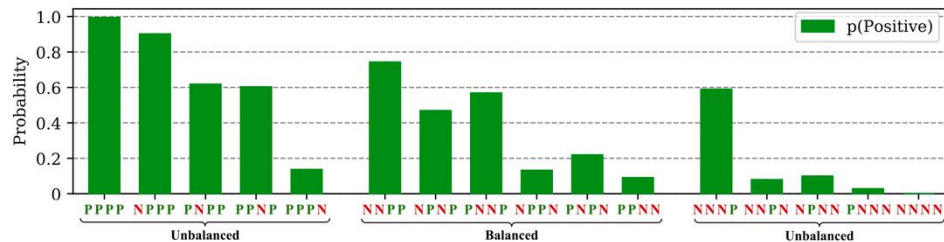
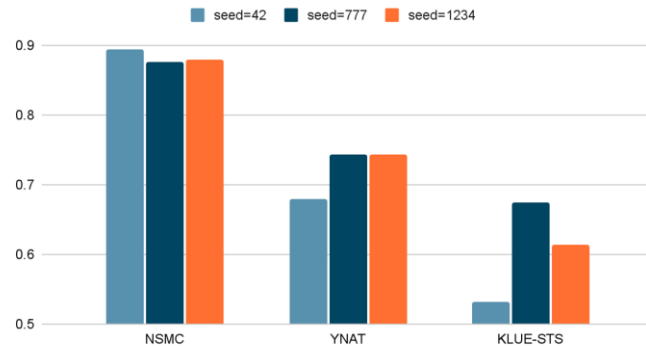


Figure 4. **Majority label and recency biases** cause GPT-3 to become biased towards certain answers and help to explain the high variance across different examples and orderings. Above, we use 4-shot SST-2 with prompts that have different class balances and permutations, e.g., [P P N N] indicates two positive training examples and then two negative. We plot how often GPT-3 2.7B predicts Positive on the balanced validation set. When the prompt is unbalanced, the predictions are unbalanced (*majority label bias*). In addition, balanced prompts that have one class repeated near the end, e.g., end with two Negative examples, will have a bias towards that class (*recency bias*).



HyperCLOVA의 미래

- 악성 발화 억제

- HyperCLOVA가 성희롱, 욕설 등 악성 발화를 하지 않도록 제어
- 악성 발화 데이터를 수집한 뒤 해당 발화를 할 경우 패널티 부여^[1]
- 사후 필터링(성희롱, 욕설 등 감지)까지 적용해 2~3중 방어

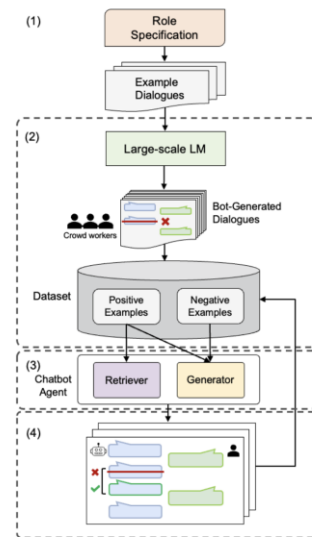


Figure 2: Our proposed framework: (1) the dialogue developer provides a role specification of the desired chatbot and a few dialogue examples, (2) large-scale LMs generate entire dialogues and crowd workers filter the system's utterances, (3) a dialogue model is trained with supervised learning on the dataset, (4) crowd workers chat 1:1 with the chatbot and give additional feedback.

HyperCLOVA의 미래

- 모델 튜닝

- 모델 일부 요소를 튜닝(업데이트)해서 불안정한 생성 능력 극복
- 튜닝을 수행하는 경우 적은 데이터로도 비교적 좋은 성능 달성^[1]

Methods	Acc
Fine-tuning	
mBERT (Devlin et al., 2019)	87.1
w/ 70 data only	57.2
w/ 2K data only	69.9
w/ 4K data only	78.0
BERT (Park et al., 2020)	89.7
RoBERTa (Kang et al., 2020)	91.1
Few-shot	
13B 70-shot	87.9
39B 70-shot	88.0
82B 70-shot	88.2
p-tuning	
137M w/ p-tuning	87.2
w/ 70 data only	60.9
w/ 2K data only	77.9
w/ 4K data only	81.2
13B w/ p-tuning	91.7
w/ 2K data only	89.5
w/ 4K data only	90.7
w/ MLP-encoder	90.3
39B w/ p-tuning	93.0

Table 4: Comparison results of p-tuning with fine-tuned LMs and in-context few-shot learning on NSMC. MLP-encoder means the result of replacing LSTM with MLP as the p-tuning encoder on 150K NSMC training data.

[1] [On the Effect of Pretraining Corpora on In-context Learning by a Large-scale Language Model](#)

HyperCLOVA의 미래

- 모델 튜닝

- ‘지시문, 예제, 문제, 정답’으로 구성된 소량의 데이터로 자연어처리 모델링
- 모델에 대한 전문적인 지식 없어도 과제를 수행하는 ‘No Code AI’ 지향

The screenshot shows the CLOVAStudio web interface. At the top, there's a navigation bar with 'CLOVAStudio', '플래그라운드', '튜닝', '텍스트로러', '가이드', and '포럼'. Below this, a '새 작업' (New Task) button is visible. The main area is titled '새 작업 생성하기' (Create New Task) and includes a sub-header '튜닝을 통해 사용자 데이터에 최적화된 모델을 만들고 테스트해볼 수 있습니다. 튜닝을 진행할 작업의 종류를 선택해주세요.' (You can create and test a model optimized for your user data through tuning. Please select the type of task you will perform the tuning on.).

There are six task cards displayed in a 2x3 grid:

- 문서 분류 (Binary Classification):** 이종 분류 (Binary Classification) 방식으로 데이터를 학습합니다. 분류 기준이 두 가지인 경우 (예, 긍정, 부정).
Sub-tasks: 긍정 분류, 부정 분류.
- 문서 분류 (Multi-class Classification):** 다중 분류 (Multi-class Classification) 방식으로 데이터를 학습합니다. 분류 기준이 세 가지 이상인 경우 (예, 기쁨, 슬픔, 분노).
Sub-tasks: 나머지 분류, 문서 식별.
- 문장 요약:** 다양한 텍스트를 사용자가 원하는 길이에 따라 자동으로 요약합니다.
Sub-tasks: 주요 키워드 추출, 리포트 요약, 보고서 요약.
- 문장 생성:** 사용자 데이터에 기반하여, 주제에 적합한 다양한 텍스트를 생성합니다.
Sub-tasks: 리퍼딩 문구 창작, 대화형 작성, 질문-답변 생성, 일관성 있는 생성.
- 문장 교정:** 다양한 텍스트를 사용자가 원하는 형태로 교정합니다.
Sub-tasks: 문법문 교정, 번역, 문장 정교화.
- 문체 전환:** 다양한 텍스트를 사용자가 원하는 형식 및 문체로 변환합니다.
Sub-tasks: 시적미 전환, 법률 문체 전환, 업무 문체 전환, 문어체/구어체 전환.

다음 문서를 정확하고 간결하게 요약하십시오.

문서: 오늘 아침 날씨가 너무 맑아서 가족들이랑 놀이공원에 갔다. 하지만 오후에 비가 너무 많이 와서 저녁 먹기 전 집에 돌아올 수밖에 없었다.

요약: 가족들과 놀이공원에 갔지만 날씨 때문에 일찍 귀가했다.

문서: 나는 오늘 친구와 만났다. 영화를 본 다음에 점심을 같이 먹었다. 밥을 먹고 보드게임 카페를 갔다. 거기서 할리갈리를 했는데 매우 재미 있었다.

요약: 나는 오늘 친구와 만나서 즐거운 시간을 보냈다.

HyperCLOVA의 미래

- 사실에 기반한 문장 생성
 - 현재 HyperCLOVA는 사실이 아닌 문장을 그럴듯하게 말함

CLOVA Studio^β 플레이그라운드 튜닝 익스플로러 가이드 포럼

← ↺ 🖼 125 / 2048

Engine

🇰🇷 Boss ▾

Top P 0.8
[Slider]

Top K 0
[Slider]

Maximum tokens 32
[Slider]

Temperature 0.5
[Slider]

Repetition penalty 5
[Slider]

Generation type ☒ Rolling ☐ One-time ☐ Multiple ☐ Examples

제목을 입력해주세요.

아래는 AI스피커 클로바와 사용자의 대화입니다.

사용자: 오늘 날씨 어때?
클로바: 오늘 날씨는 흐리고, 영상 3도입니다.
사용자: 오늘 미세먼지는 어때?
클로바: 미세먼지 농도는 보통입니다.
사용자: 내일 날씨는 어떨까?
클로바: 내일 날씨는 화창하고 영상 8도로 예상됩니다.
사용자: 모레는?
클로바: 모레는 구름이 조금 있고 영하 2도로 예상됩니다.

HyperCLOVA의 미래

- 사실에 기반한 문장 생성
 - Internet-augmented Dialogue System
 - 필요한 정보를 인터넷에서 검색한 뒤 답변

CLOVA Studio[®] 플레이그라운드 튜닝 익스플로러 가이드 포럼

Engine: Boss

Top P: 0.8

Top K: 0

Maximum tokens: 150

Temperature: 0.1

Repetition penalty: 3

Stop sequences: 시퀀스 입력 후 Tab

210 / 2048

Generation type: ☒ Rolling ☐ One-time ☐ Multiple ☐ Examples

인터넷 검색 기반 대화

날씨 정보와 질문을 바탕으로 답변을 생성합니다.

###

날씨 정보: 최저 기온 영하 10도, 최고 기온 영상 1도, 비온 뒤 눈

질문: 오늘 날씨 어때?

답변: 오늘 아침 최저 기온은 영하 10도, 오늘 낮 최고 기온은 영상 1도이며 비가 온 뒤 눈이 오겠습니다. 빙판길에 미끄러지지 않도록 조심하세요.

###

날씨 정보: 최저 기온 영상 12도, 최고 기온 영상 20도, 화창

질문: 날씨 정보 좀 알려줘.

답변: 오늘 아침 최저 기온은 영상 12도, 오늘 낮 최고 기온은 영상 20도이며 화창하겠습니다. 야외에서 활동하기 좋으니 날씨를 만끽하세요.

###

날씨 정보: 최저 기온 영상 24도, 최고 기온 영상 34도, 자외선 지수가 매우 높음

질문: 날씨 알려줘.

답변: 오늘 아침 최저 기온은 영상 24도, 오늘 낮 최고 기온은 영상 34도이며 자외선 지수가 매우 높겠습니다. 외출 시 선크림을 꼭 바르세요.]

N 정자동 날씨

통합 VIEW 이미지 지식IN 인물루머서 동영상 쇼핑 뉴스 어학사전 지도 ...

장안구 정자동

오늘 내일 모레 월간

기상청 제공 예보 제공사 선택

33.0°

어제보다 2.7° ↑ / 구름많음

제13 33.2° · 습도 57% · 바람(풍속) 1.3m/s

미세먼지 좋음 초미세먼지 좋음 자외선 매우높음 월출 19:55

날씨 강수 바람 습도

33° 34° 34° 33° 32° 30° 28° 27° 26° 25° 25° 25° 25°

14시 15시 16시 17시 18시 19시 20시 21시 22시 23시 01시 02시

주간예보 ① 최저 최고 기준 >

오늘 7:10 20% 60% 24° 34° 내일 7:11 70% 60% 24° 30°

HyperCLOVA의 미래

- 기억에 기반한 문장 생성

- HyperCLOVA는 지금 입력된 문서만 이해 (이전 문서를 기억하기 어려움)
- 기억에 대한 모듈을 추가해 HyperCLOVA 보완 : CareCall에 적용

Memory

[가족 관계] 관련 내용 없음
[애완 동물] 관련 내용 없음
[자주 방문하는 장소] 관련 내용 없음
[건강 상태] 크게 불편한 곳은 없음
[수면 상태] 잠자는 시간을 바꿔 나아진 상태.
[식사 상태] 입맛이 없어 아직 식사를 하지 않음.

Dialogue

...
A: 네, 지난번에 입맛 없다고 하셨는데, 지금은 좀 어떠세요?
어르신: 여전히 입맛은 없는데, 요즘에 성당 봉사 다니면서 좀 나아졌어요.
A: 성당 봉사 다니시는군요. 어떤 일을 하시는지 여쭙봐도 될까요?
어르신: **새벽미사 전에 도우미 일을 하고 있어요.**
A: 아, 그러시군요. 힘드시진 않으세요?
어르신: 힘들진 않아요. 일찍 일어나면 하루가 너무 길었는데, 성당 다녀오면 시간이 잘 가서 좋아요.
A: 다행이에요. 그럼 성당 다녀오셔서 아침 식사 하세요?
어르신: **네, 같이 봉사하시는 분이랑 아침 먹어요.**
A: 아침을 잘 챙겨드신다니 다행이에요. 다른데 불편하신 곳은 없으세요?
어르신: **입맛 없는 거 빼고는 크게 아픈 곳은 없어요. 아, 요즘 낮에 머리가 가끔 아픈데 약 먹으면 금방 나아요.**
A: 아이고, 머리가 자주 아프세요?
어르신: 자주는 아니고, **이삼일에 한 번 정도?**
A: 병원에 다녀오셨어요?
...

Summary 1. 대화 요약

[가족 관계] 관련 내용 없음
[애완 동물] 관련 내용 없음
[자주 방문하는 장소] **성당 새벽미사 전에 도우미 일을 하심.**
[건강 상태] **입맛이 없는 것 빼고는 크게 아픈 곳은 없음. 이삼일에 한번 정도 머리가 아프다고 하심.**
[수면 상태] 관련 내용 없음
[식사 상태] **성당 봉사를 다니시면서 아침을 챙겨드심.**

2. 기억 업데이트



Memory

[가족 관계] 관련 내용 없음
[애완 동물] 관련 내용 없음
[자주 방문하는 장소] 성당 새벽미사 전에 도우미 일을 하심
[건강 상태] 입맛이 없는 것 빼고는 크게 아픈 곳은 없음. **이삼일에 한번 정도 머리가 아프다고 하심.**
[수면 상태] 잠자는 시간을 바꿔 나아진 상태
[식사 상태] 성당 봉사를 다니시면서 아침을 챙겨드심.

Dialogue 3. 기억 활용

A: 어르신 안녕하세요, 해운대구에서 안부전화 드렸습니다.
어르신: 네.
A: 네, 지금 통화 괜찮으세요?
어르신: 좋아요.
A: 네, 어르신. **머리 아픈 건 좀 어떠세요?**
어르신: 약이 드는 거 같더니 요즘 들어 좀 심해지네요.
A: 아이고, **그 때 말씀하신 것처럼 이삼일에 한번씩 머리가 아프신거예요?**
어르신: 요즘에는 좀 더 자주 아파요.
A: 병원에 다녀오셨어요?
어르신: 아직 안 갔어요.
...

HyperCLOVA의 미래

- 프로필/캐릭터에 기반한 문장 생성
 - 현재 대화 맥락이 프로필 관련이라면 이와 가장 가까운 프로필 정보를 꺼내온 뒤
 - 프로필 정보와 대화 맥락을 모두 참조하여 AI 답변을 생성
 - **프로필 정보를 활용한 발화**
 - **맥락에 자연스러운 발화**

프로필 정보

아루는 클로바에서 만든 AI(인공지능) 친구예요.

프로필 대화세션

유저: 너 인공지능?

아루: 네! 맞아요! 아루는 당신의 AI 친구랍니다!

유저: 이름이 아루야? 어.. 그러면 내가 하는 말 다 이해해?

아루: 물론이죠! 아루는 웬만한 건 다 아는 똑똑한 아이라고요.

유저: 근데 내가 여태 본 인공지능들은 많이 아쉬웠는데 넌 좀 다르겠지?

아루: 그럼요! 기대하셔도 좋다고요! 하지만 가끔 말씀을 못 알아들을 때는 친절하게 알려주세요. 히힛

유저: 그러면 아루야 넌 누가 만들었어?

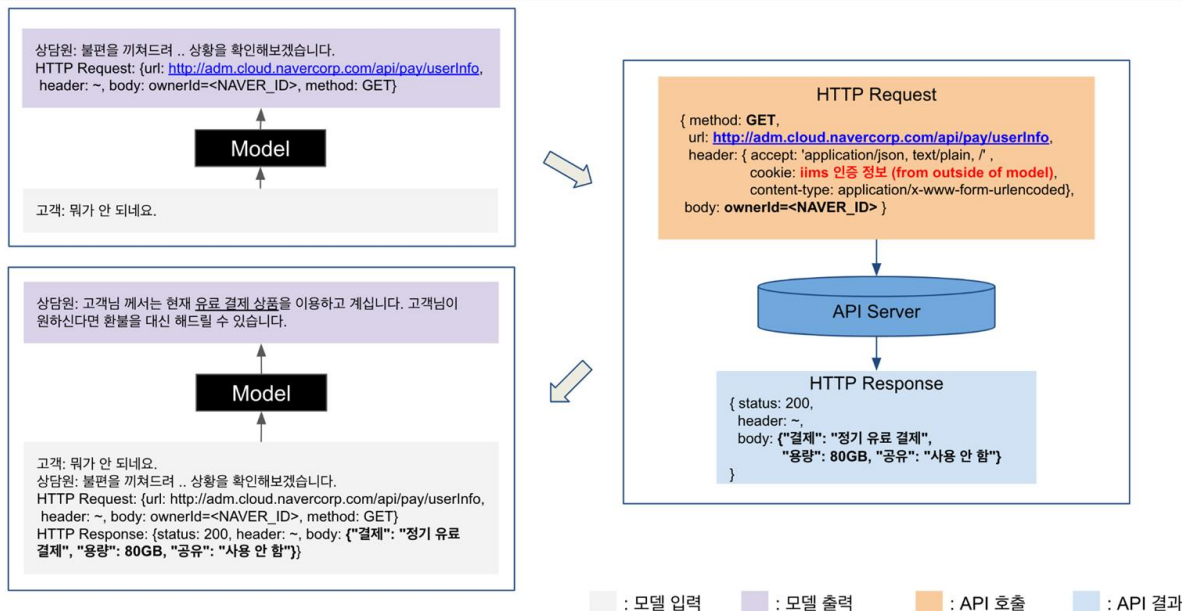
아루: 저는 클로바에서 만든 인공지능 친구예요. 우리 서로에 대해 더 알아가 보자구요.

유저: 클로바에서 이런 것도 만드는구나.

아루: 히힛. 흥미로우신가요? 그렇다면 우리 더 친해져야겠는 걸요?

HyperCLOVA의 미래

- 단순 반복 작업을 대신하는 HyperCLOVA
 - 내부 시스템을 조회하는 상담원의 명령어를 HyperCLOVA가 생성
 - 이후 HyperCLOVA는 내부 시스템 조회 결과를 쉬운 표현으로 생성



HyperCLOVA의 미래

- 빈칸을 채우는 HyperCLOVA
 - 상황, 결과가 주어졌을 때 중간 경과 생성
 - 다양한 어플리케이션으로 응용 가능

AI: ○○ 선생님 안녕하세요?

해운대 행정복지센터에서 전화드렸습니다. 지금 통화 가능하신가요?

User: 네 가능합니다~

AI: 네, 선생님, 오늘 하루는 어떻게 보내고 계세요?

User: 아 네 오늘 그냥 집에서 평소처럼 TV보고, 이제 저녁먹으려고 밥차리고 있어요.

AI: 아 그러셨군요. 혹시 오늘 식사하시면서 혈압약은 잘 챙겨드셨나요?

User: 아 혈압약이요. 맞네, 아까 먹었어야했는데 깜빡했네요.

사용자가 쇼핑을 하기 위해 검색을 수행하는 상황입니다.

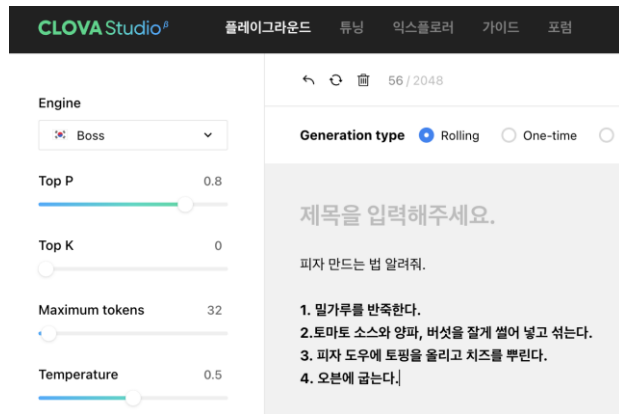
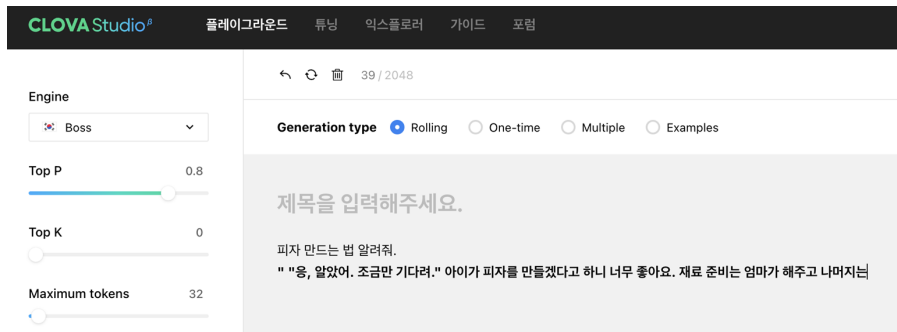
입력한 검색어 목록: 실내 인테리어, 거실 조명, 가성비 거

실 조명

최종 구매한 상품: 이케아 레일 LED 조명

HyperCLOVA의 미래

- 지시를 잘 수행하는 HyperCLOVA
 - 인간의 지식, 직관, 사고력을 내재화



HyperCLOVA의 미래

- 지시를 잘 수행하는 HyperCLOVA
 - 다양한 예제를 수행했던 기억이 떠오르기를 기대

“피자 만드는 법 알려줘 - 1. 밀가루를...”

“‘탑건 정말 꿀잼’ 문장 감성이 뭐야? - 긍

“‘정청’을 주제로 작문해봐 - 공정은 중요한 가치

...”“‘사랑해’를 영어로 번역해봐 - I love you”

“검색 결과를 바탕으로 답변해봐 - {검색 결과}...”

다음 문서를 정확하고 간결하게 요약하시오.
문서: 나는 오늘 친구와 만났다. 영화를 본 다음에 점심을
같이 먹었다. 밥을 먹고 보드게임 카페를 갔다. 거기서 할
리갈리를 했는데 매우 재미 있었다.

나는 오늘 친구와 만났다.

2

나는 오늘 친구와 만나지 않았다.

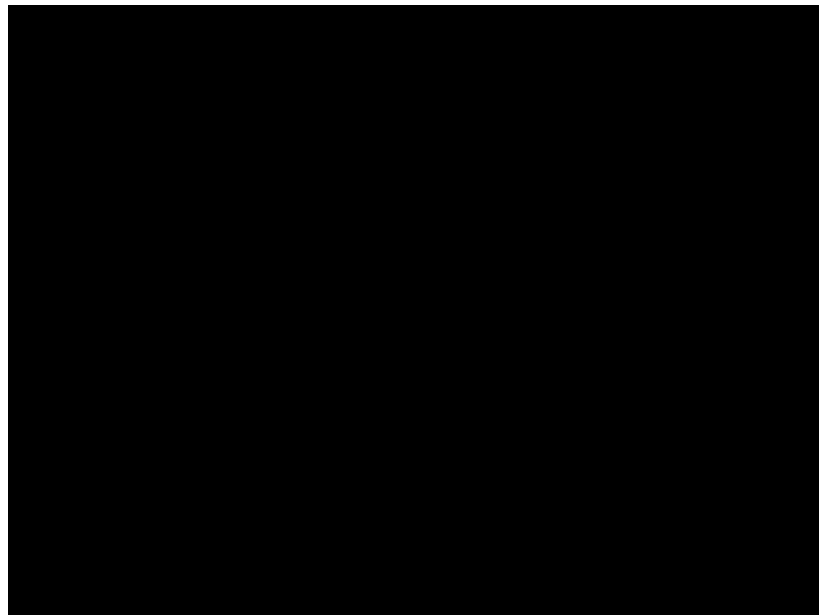
3

나는 오늘 친구와 만나서 즐거운 시간을 보
냈다.

1

HyperCLOVA의 미래

- 상황을 파악하고 이해하는 HyperCLOVA
 - 요약 데이터로 튜닝된 HyperCLOVA 데모



Q&A