

MBTI 기반 투자성향분석

BOAZ 분석 17기_Stalker



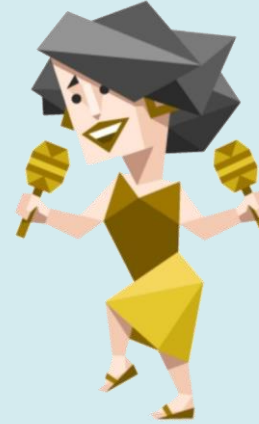
STALKER



김지원



곽윤경



조수연



유승희



김나현



목차



STALKER의 투자성향분석



1

주제 선정 이유

↑ 56

2

서비스 개요

↑ 45

3

서비스 흐름

↑ 33

4

진행상황 및 계획

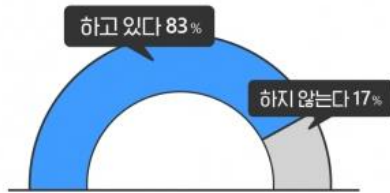
↑ 19

🔍 주제 선정 이유

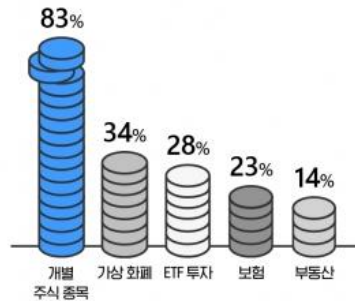
2030세대 의 투자 관심도 증가

투자가 일상이 된 MZ세대...20~30대 직장인 약 90% 주식투자

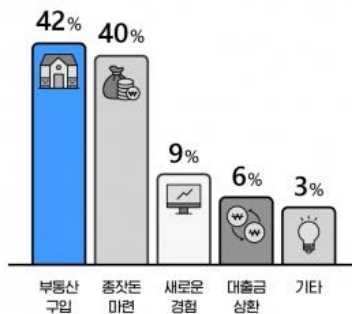
G 투자 여부



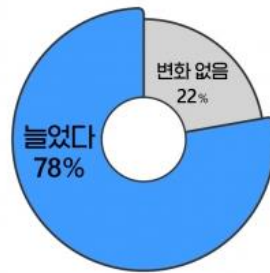
G 투자 방법 (복수응답)



G 투자 목적



G 보험에 대한 관심



[데일리한국 박재찬 기자]대형 GA(법인보험대리점) 리치앤코가 수도권 직장인 1000명을 대상으로 설문 조사한 결과, MZ세대 직장인 10명 중 8명이 '일상적'으로 투자(예적금 제외)를 하고 있으며, 올해 10% 이상 수익률을 낸 것으로 조사됐다. 더불어 10명 중 8명은 최근 보험에 관심이 증가했다고 답했다.

리치앤코가 모바일 리서치 전문 기관 오픈서베이에 의뢰해 수도권에 거주하는 20~30대 직장인 1000명을 대상으로 재테크 관련 설문조사를 진행한 결과를 공개했다.

가장 눈길을 끈 부분은 투자 여부다. 2030세대 직장인 응답자의 83%가 주식, 부동산, 펀드, 가상 화폐 등에 '현재 투자하고 있다'고 답했다. 10명 중 8명이 예적금을 제외한 지속적인 투자 활동을 하고 있는 것을 봤을 때 MZ세대 사이에서는 이미 투자가 '일상'으로 자리잡은 모양새다.

2030세대 투자자 중 절반 이상이 전체 수입의 약 20%를 넘지 않는 수준에서 투자 활동을 하고 있었다. 월 수입 대비 투자 비중을 묻는 질문에 10% 미만이라고 답한 응답자가 28.5%로 가장 많았고 '20% 미만(27.0%)', '30% 미만(17.9%)'이 그 뒤를 이었다. 전체 수입의 50% 이상을 투자한다고 답한 공격적인 투자자 비중도 11.6%로 적지 않다.

중 단기(1~5년) 투자 목적을 묻는 질문에는 '독립, 자가 구입 등 부동산 마련'이 42.2%로 1위를 차지했다. 눈길을 끄는 부분은 20대가 30대보다 더 많은 비율로 '부동산 마련'을 투자 목적으로 삼았다는 점이다.

🔍 주제 선정 이유

시중의 투자 성향 분석



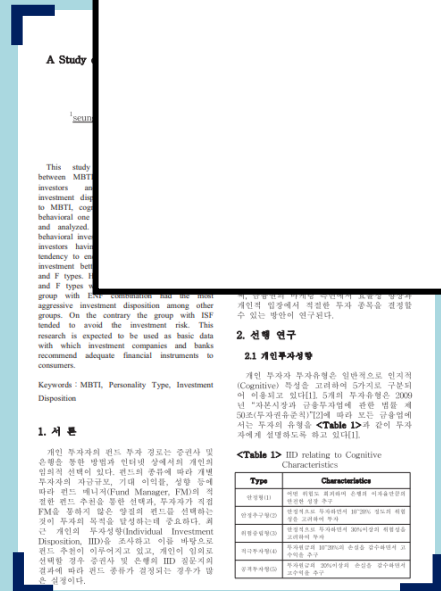
※ 최종 주제 ※

MBTI별 개인 맞춤형 투자 성향 분석



투자 가이드 라인을 제공하는 챗봇 구현하기

MBTI 성격



INT	17	2.24±1.033
INF	40	2.28±1.109

** $p < 0.01$

성격 조합은 EST, ESF, ENT, ENF, IST, ISF, INT, INF로 총 8개이며, 각 성격의 투자 성향은 유의한 차이가 있는 것으로 나타났다. 가장 안정적인 투자성향을 보이는 것은 ISF형의 투자자이며, 가장 공격적인 투자성향을 보이는 것은 ENT형의 투자자이다.

🔍 서비스 개요

카카오톡 플러스친구 STALKER

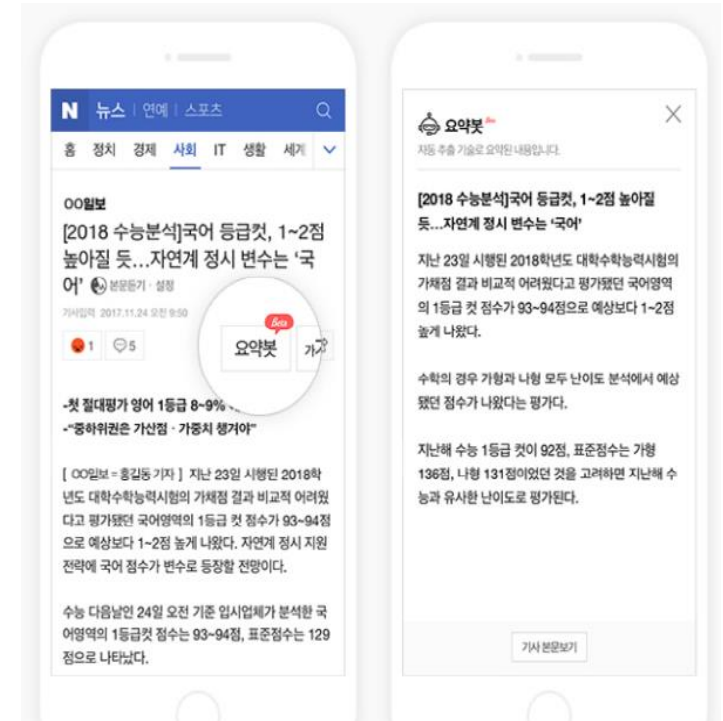
KakaoTalk
-P 플러스친구

STALKER
'STOCK' + 'TALKER'

MBTI 기반 투자성향 분석



주식 뉴스 요약 정보 제공



🔍 서비스 설계

DATA SET

MBTI를 바탕으로 긍/부정을 유도할 수 있는 질문 제작

E 성향보다 I 성향이,
S 성향보다 N 성향이,
T 성향보다 F 성향이,

정신적 스트레스에 민감하여 위험을 감수하는 경향이 적다.
(E/S/T 위험 감수 vs I/N/F 안전 추구)

출처 : 김석환, 강형구 (2020) 금융소비자의 위험선호와 성격유형과의 관계에 관한 위험연구

🔍 서비스 설계

DATA SET

MBTI를 바탕으로 긍/부정을 유도할 수 있는 질문 제작

각 성향별로 주관식 질문 2개, 객관식 질문 2개 제작
스토리라인에 따라 질문 전개

- 주관식 질문 : 자연어 처리를 통한 긍/부정 감정분석
- 객관식 질문 : 성향 판단 어려움 시 보조할 수 있는 질문

답변 유도를 위해 일부 주관식 질문에는 답변 예시를 선제시
답변 다양화를 위해 일부 주관식 질문에는 답변 예시 제거

🔍 서비스 설계

DATA SET

긍/부정을 대답을 유도하여 성향을 파악할 수 있는 질문 제작

투자성향 (안정형, 안정추구형, 위험중립형, 적극투자형, 공격투자형)
판단 기준 중 3가지 선택

- 수익성 우선 VS 안전성 우선
- 단기적 VS 장기적
- 한 가지 상품에 집중적으로 투자 VS 다양한 상품에 분산 투자

🔍 서비스 설계

DATA SET

긍/부정을 대답을 유도하여 성향을 파악할 수 있는 질문 제작

(1) 수익성 VS 안정성

[주] 절친 춘식이가 좋은 투자 기회가 생겼다며 1억의 빚을 내겠다고 합니다. 춘식이에게 뭐라고 할 것인가요?

[객] 당신은 주식이 망해 5억의 빚을 지게 되었습니다. 이 때 무지가 게임을 제안했습니다. 다음 중 어떤 선택을 하시겠습니까? (100% 확률 1억, 50% 확률 5억)

[객] 내가 감내할 수 있는 손실 (0-10%, 10-20%, 20-30%, 30-50%, 50% 이상)

🔍 서비스 설계

DATA SET

긍/부정을 대답을 유도하여 성향을 파악할 수 있는 질문 제작

(2) 단기적 VS 장기적

[주] 당신이 오늘 매수한 '카카오'를 2년동안 묵히면 무조건 수익이 날거라는 말을 들었습니다. 이 말을 들었을 때 당신의 생각이나 느낌은 어떤가요?

[객] 당신의 투자 가능 기간을 선택해주세요. (6개월 이내, 6개월 이상 ~ 1년 이내 ...)

🔍 서비스 설계

DATA SET

긍/부정을 대답을 유도하여 성향을 파악할 수 있는 질문 제작

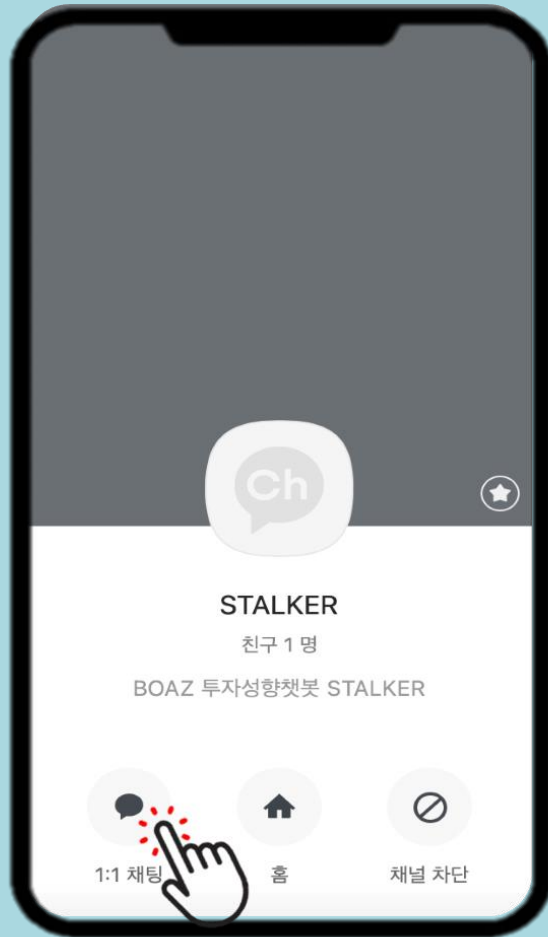
(3) 집중투자 VS 분산투자

[주] 소문난 스타 펀드매니저 제이지가 본인의 명함을 주면서, 당신의 투자를 맡아주겠다고 합니다. 이 때 당신의 대답이나 기분이 어떤지 적어주세요.

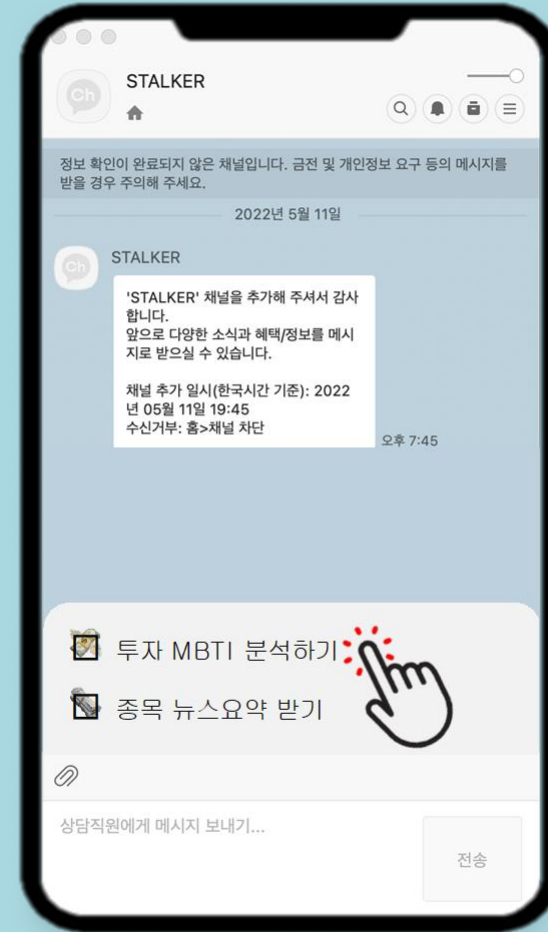
[객] 2년이 지난 후, '카카오'를 매도해서 수익 1억이 생겼습니다. 이 돈으로 다시 투자할 때 당신의 선택은 무엇인가요? (한 종목 몰빵, 여러 군데 분산 투자)

🔍 서비스 설계

MBTI 기반 투자 성향 분석



1. 카카오톡 친구 추가 -> 1:1 채팅

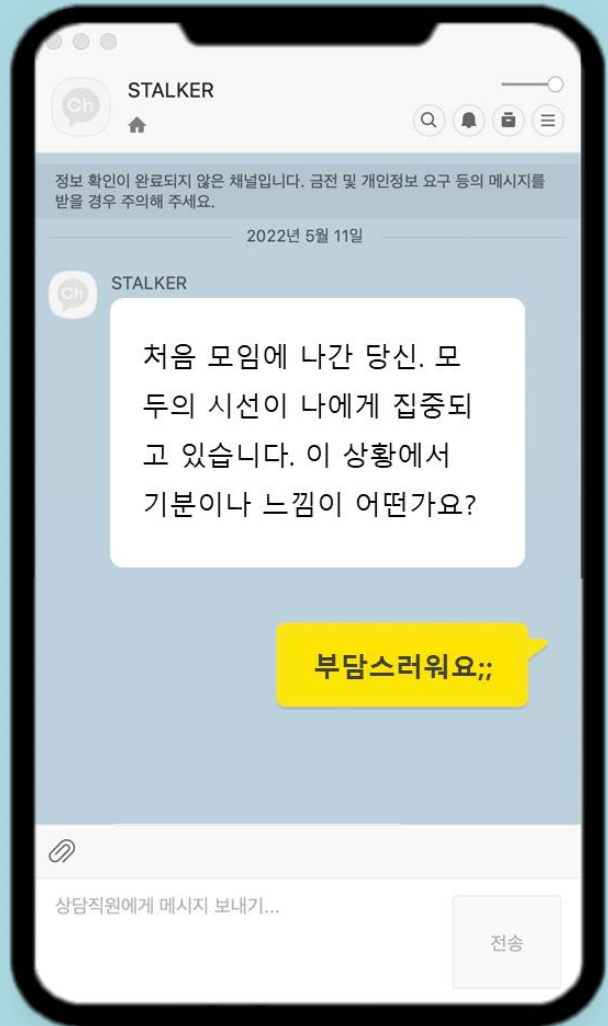


2. 메뉴에서 [투자 MBTI 분석하기] 선택

서비스 설계

MBTI 기반 투자 성향 분석

3. 챗봇 질문 제시 -> 사용자 답변



📌 객관식 및 주관식 총 20문항

📌 스토리 형식 질문으로 구성

아무도 먼저 나에게 말을 걸어주지 않는다면 어떤 기분이나 생각이 드시나요?

주식 초보인 나에게 주식 포트폴리오 구성 종목, 작성법, 심지어 글씨체까지 알려주는 친절한 옆자리 훈수러 프로도. 당신은 프로도가 마음에 드시나요?

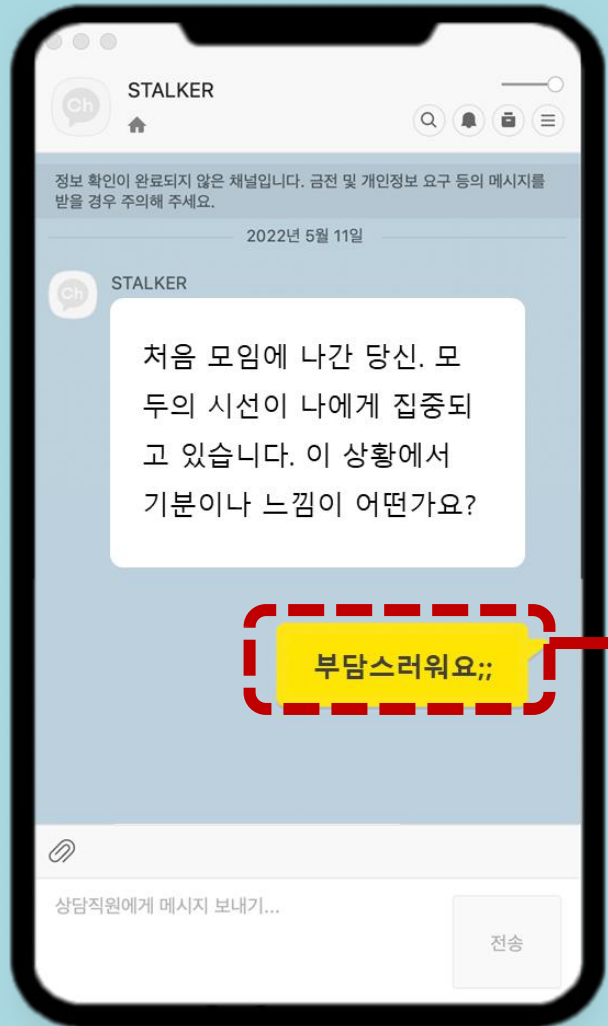
프로도가 나에게 열심히 주식을 알려주다가, 뜬금없이 '지진나면 어떻게 할거야?' 라고 물어봅니다. 이 때 당신의 행동이나 생각을 적어주세요.

당신은 일년동안 스터디에서 열심히 공부해서 세미나 발표를 무사히 끝냈습니다. 절친 춘식이 "너 공부 열심히 안하는 것 같았는데 발표 잘했다." 라고 했을때, 당신의 기분은 어떨까요?

서비스 설계

MBTI 기반 투자 성향 분석

3. 챗봇 질문 제시 -> 사용자 답변

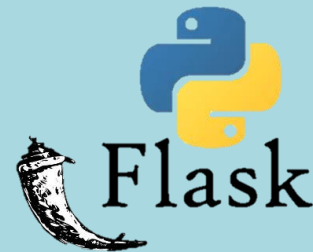


📌 객관식 및 주관식 총 20문항

📌 스토리 형식 질문으로 구성

<구현 계획>

1) Python 서버(미정)로 대담 텍스트를 전송



2) 질문당 긍정/부정 or 성격유형 분석 후 결과값 누적



서비스 설계

MBTI 기반 투자 성향 분석

4. 질문 - 답변 종료 후 투자 MBTI 결과 출력



<성향 분류>

- 📌 카카오프렌즈 캐릭터를 컨셉으로 총 8개 성향으로 분류
- 📌 성향별 특징과 투자 조언을 함께 보여줌

I S F 겁쟁이 주린이 라이언

I N T 게으른 분석가 허세킹 프로도

I S T 자신감 뽐뽐 마이웨이 네오

I N F 멘탈 관리 필요한 튜브

E N F 일단 가보자고~! 팔랑귀 무지

E N T 과감한 투자천재 어피치

E S F 팔고나면 오르는 호구 제이지

E S T 생각을 알 수 없는 욕심쟁이 능력자 콘

🔍 서비스 설계

주식 뉴스 요약 정보 제공

사용자 : 종목 입력

챗봇 : 해당 종목 뉴스 요약 / 키워드 출력

🔍 서비스 설계

주식 뉴스 요약 정보 제공

라이브러리 활용

PORORO

: Platform Of neural mOdelS for natural language processing

- 뉴스 헤드라인의 토픽 분류 (zero-shot topic classification)
- KoBART summary 기능을 사용해, 입력 내용을 짧은 문장으로 축약

라이브러리 활용

KR-WordRank

: 단어의 좌/우에는 단어가 등장하고, 단어가 아닌 substring 좌/우에는 단어가 아닌 substring 이 등장함. 단어는 다른 많은 단어들과 연결되기 때문에 질 좋은 links 가 많이 연결되며, 단어가 아닌 substring 은 소수의 backlinks 를 받는다는 것.

: 토큰나이저를 이용하지 않으면서 단어/키워드 추출을 비지도학습 기반으로 수행

서비스 설계

DataSet

한국어 감정 정보가 포함된 단발성 대화 데이터셋

<https://aihub.or.kr/opendata/kefi-data/recognition-language/KETI-02-009>



AI Hub 감정분석 Data

Q	A	label
12시 땡!	하루가 또 가네요.	0
1지망 학교 떨어졌어	위로해 드립니다.	0
3박4일 놀러가고 싶다	여행은 언제나 좋죠.	0
3박4일 정도 놀러가고 싶다	여행은 언제나 좋죠.	0
PPL 심하네	눈살이 찌푸려지죠.	0
SD카드 망가졌어	다시 새로 사는 게 마음 편해요.	0
SD카드 안돼	다시 새로 사는 게 마음 편해요.	0

챗봇 Data

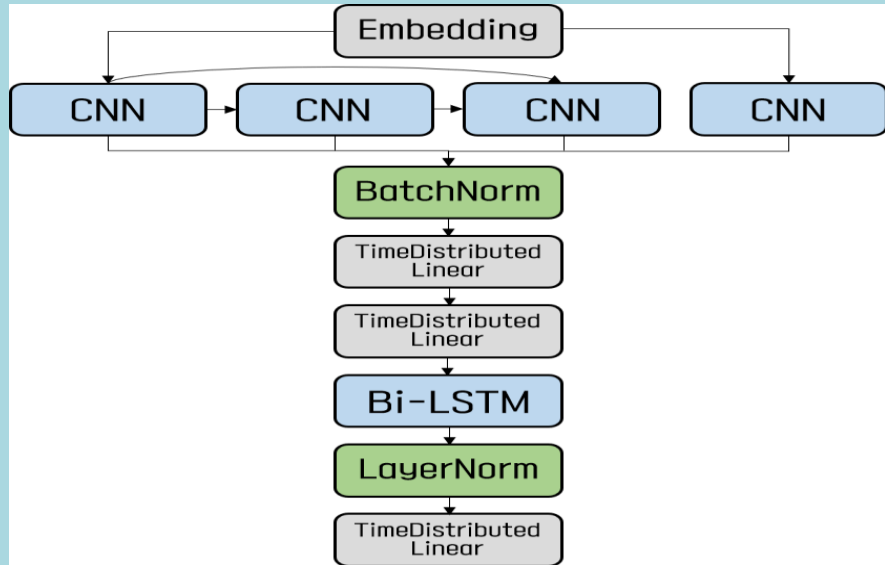
2022. 3. 26 오후 9:08:51	회식이 전혀 없는 모임	글	서운	상대에게 먼저 말을 건다	넵	전기차 테마주를 산다
2022. 3. 26 오후 9:11:01	회식이 많은 모임	즐긴다	집에 가고싶다	상대에게 먼저 말을 건다	YES	전기차 테마주를 산다
2022. 3. 26 오후 9:19:55	회식이 많은 모임	좋다 이기회에 내 자신을	목직에서 나오는 매력을	상대에게 먼저 말을 건다	좋습니다 서로 공유할수있	전기차 테마주를 산다
2022. 3. 26 오후 10:38:21	회식이 많은 모임	집중된 것을 즐긴다.	큰소리로 인사하며 반응을	상대에게 먼저 말을 건다	뭔가 써하다	전기차 테마주를 산다
2022. 3. 26 오후 10:41:41	회식이 많은 모임	많은 사람들의 관심.. 환영	이거는 변수이긴 한데...	상대에게 먼저 말을 건다	네네 마음에 들어요.	전기차 테마주를 산다
2022. 3. 26 오후 11:09:46	회식이 전혀 없는 모임	부담스럽다.	부끄러워진다.	상대가 걸 때까지 기다린	프로도가 마음에 든다.	전기차 테마주를 산다
2022. 3. 27 오전 8:00:35	회식이 전혀 없는 모임	괜찮다	종그렇다	상대에게 먼저 말을 건다	굳이다	테슬라를 산다
2022. 3. 28 오후 2:20:28	회식이 많은 모임	재밌다	어색하다	상대가 걸 때까지 기다린	아니요	테슬라를 산다
2022. 3. 29 오후 4:08:54	회식이 많은 모임	당황스럽다 떨린다	내가 뭐 잘못했나 생각해	상대에게 먼저 말을 건다	좋다	전기차 테마주를 산다
2022. 3. 29 오후 10:13:21	회식이 많은 모임	재미있다	오히려 좋아	상대에게 먼저 말을 건다	네	테슬라를 산다

설문조사 Data

🔍 서비스 설계

전처리 PROCESS_띄어쓰기 교정

Chatspace 모델



```
import chatspace
```

```
spacer = chatspace.Chatspace()
```

```
spacer.space("따뜻한봄날이되면그때는편안히만날수있으면 좋겠어요.")
```

```
# 따뜻한 봄날이 되면 그때는 편안히 만날 수 있으면 좋겠어요.
```

전처리 PROCESS_맞춤법 교정

부산대 맞춤법 교정기

The screenshot shows the interface of the '한국어 맞춤법/문법 검사기' (Korean Spelling/Grammar Checker). The main area displays the input text: '아주 재밌음 이 영화 개 꾸르잼!'. The right sidebar shows the correction suggestions:

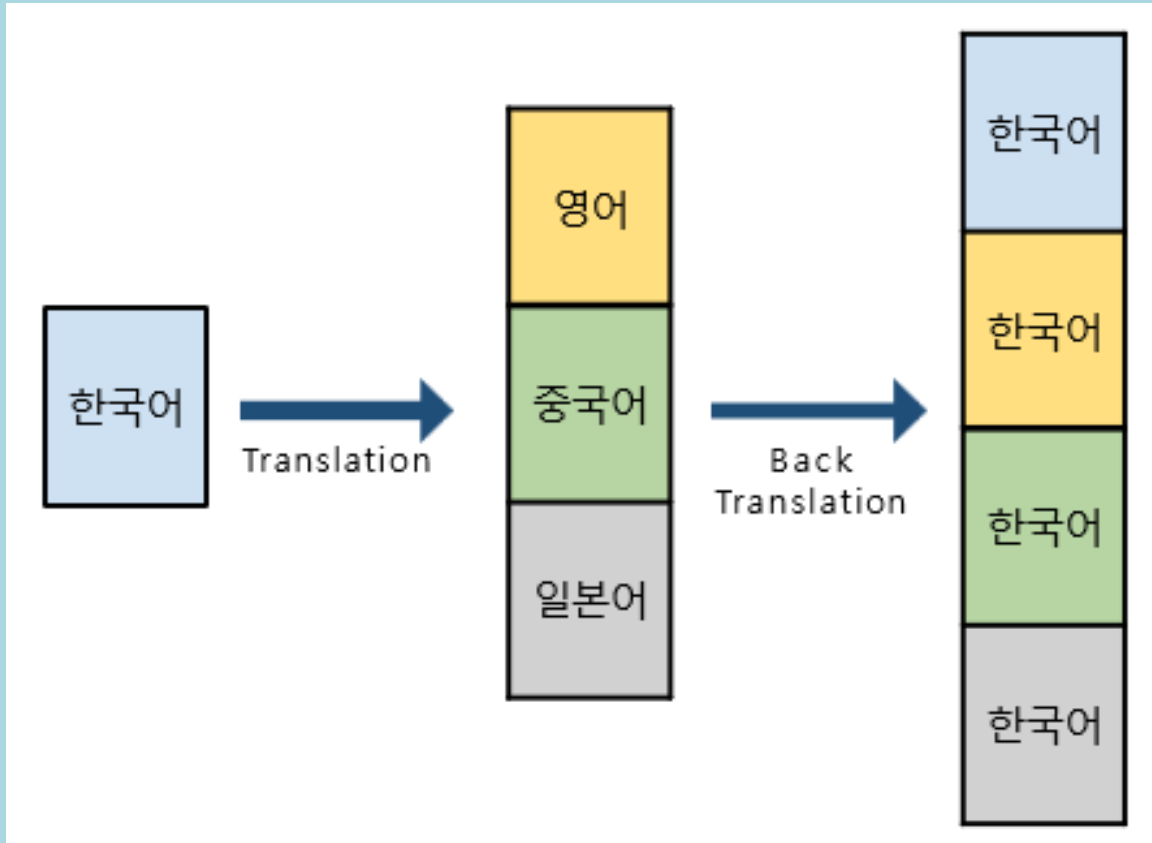
- 입력 내용: 꾸르잼
- 대치어: 아주 재밌음
- 작성 수정: 원하는 대치어를 직접 입력하세요. [적용]
- 도움말: 흔히 인터넷이나 블로그 등 SNS나 문자 메시지 등에서 줄일 수 없는 말을 줄여 씁니다만 이는 잘못입니다. 표준어로 표기해야 하며 이때 문장 성분이 다르면 띄어 써야 합니다.
- 의견: 대치어가 맞지 않거나, 다른 의견이 있으면 내용을 적어서 [보내기]를 눌러 주세요. [보내기]

At the bottom, there are buttons for '다시 쓰기' (Rewrite), '원문 복사' (Copy original text), and '돌아가기' (Go back).

🔍 서비스 설계

전처리 PROCESS_불균형 해소

Back Translation



Back Translation

기존 텍스트를 외국어로 번역한 뒤
다시 기존의 언어로 번역하는 기법

- ✧ 번역하는 언어의 어순에 맞게 재번역되어
같은 의미의 다양한 문장을 만들 수 있음
- ✧ 영어, 중국어, 일본어로 번역후 재번역하여
기존 데이터에 비해 4배 가까이 증강

서비스 설계

모델링 PROCESS

KoELECTRA

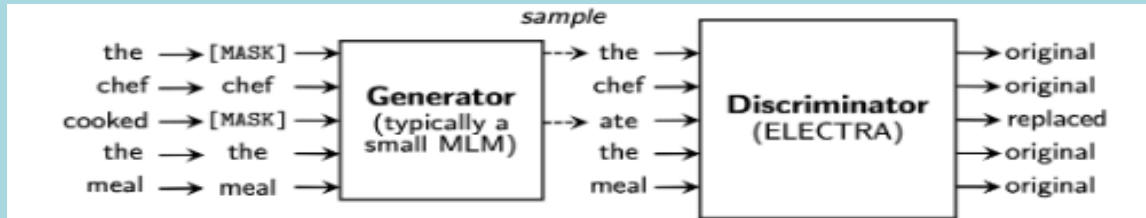


Figure 2: An overview of replaced token detection. The generator can be any model that produces an output distribution over tokens, but we usually use a small masked language model that is trained jointly with the discriminator. Although the models are structured like in a GAN, we train the generator with maximum likelihood rather than adversarially due to the difficulty of applying GANs to text. After pre-training, we throw out the generator and only fine-tune the discriminator (the ELECTRA model) on downstream tasks.

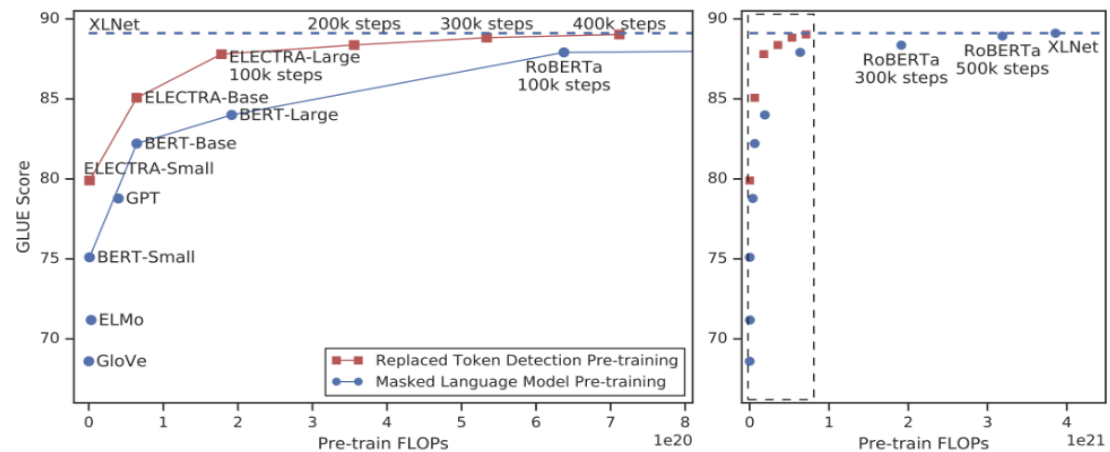
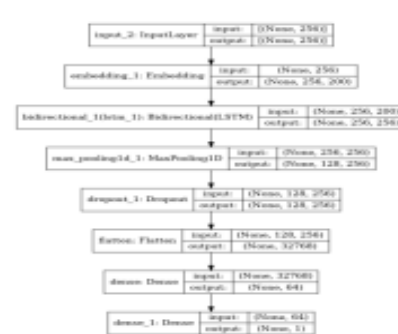


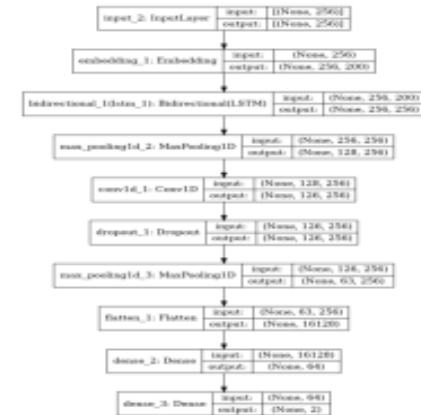
Figure 1: Replaced token detection pre-training consistently outperforms masked language model pre-training given the same compute budget. The left figure is a zoomed-in view of the dashed box.

그 외 모델링 기법들

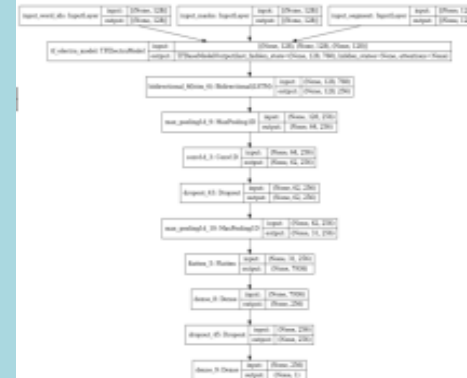
BiLSTM



CNN+BiLSTM



Ko-ELECTRA

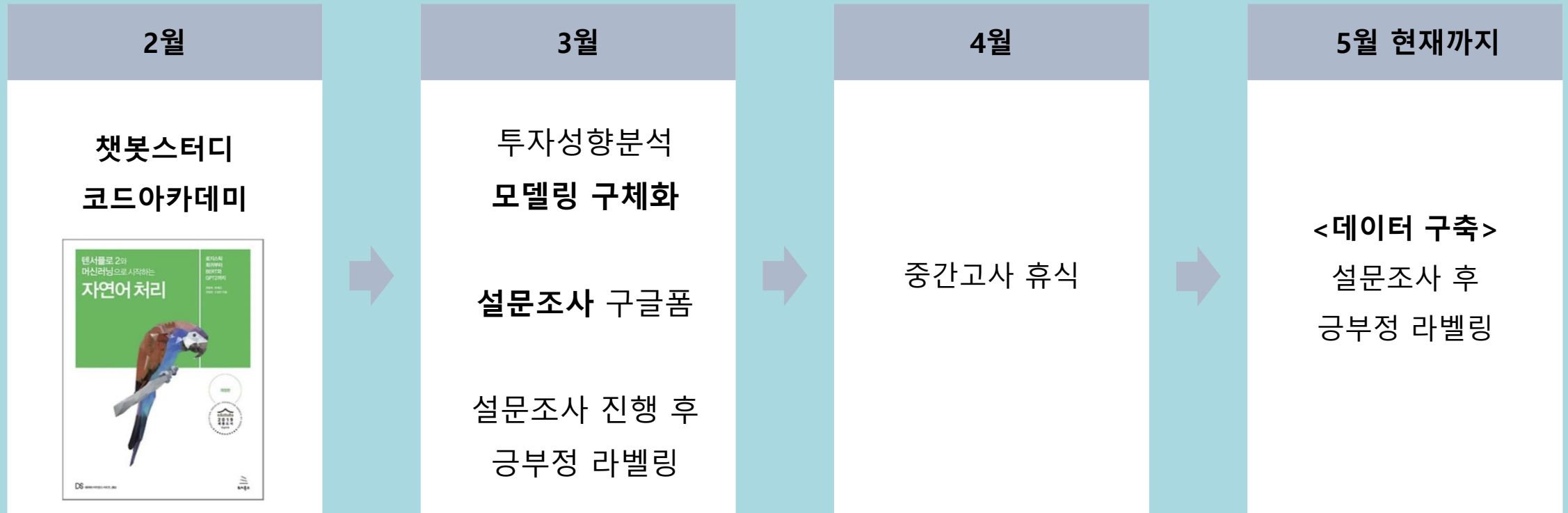


Character Level CNN



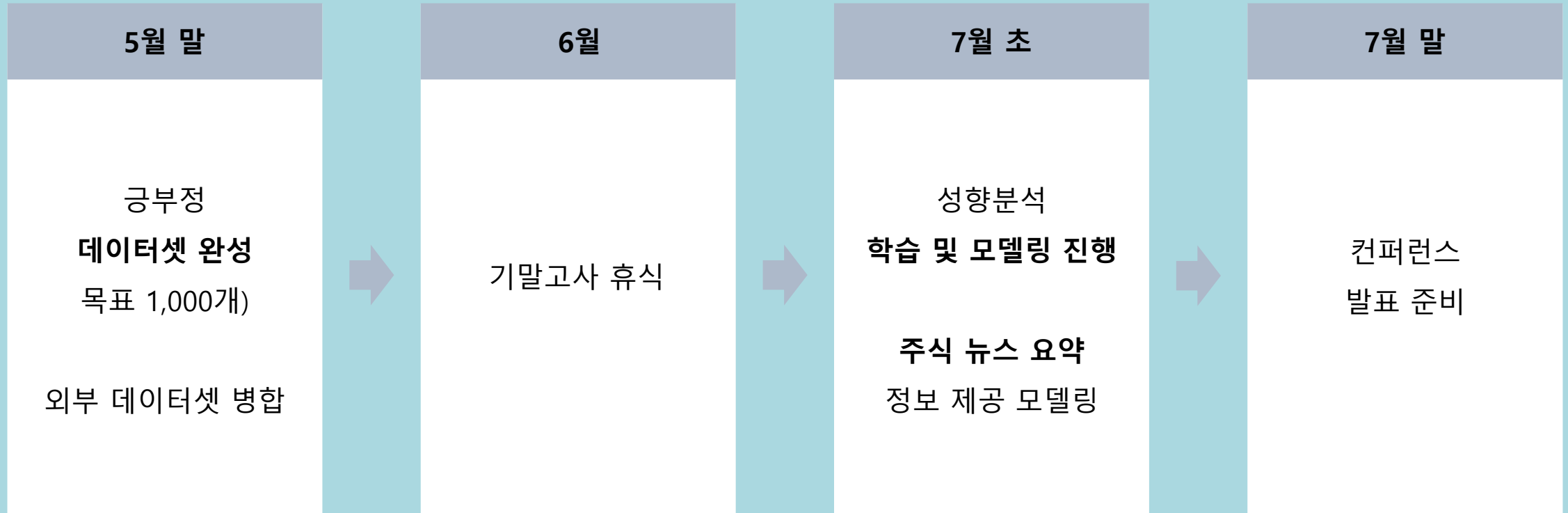
🔍 진행상황 및 계획

진행상황



🔍 진행상황 및 계획

진행계획



🔍 진행상황 및 계획

어려움

✓ 데이터의 부족

: 데이터의 양이 충분하지 않을 수 있음 (데이터 수집의 어려움)

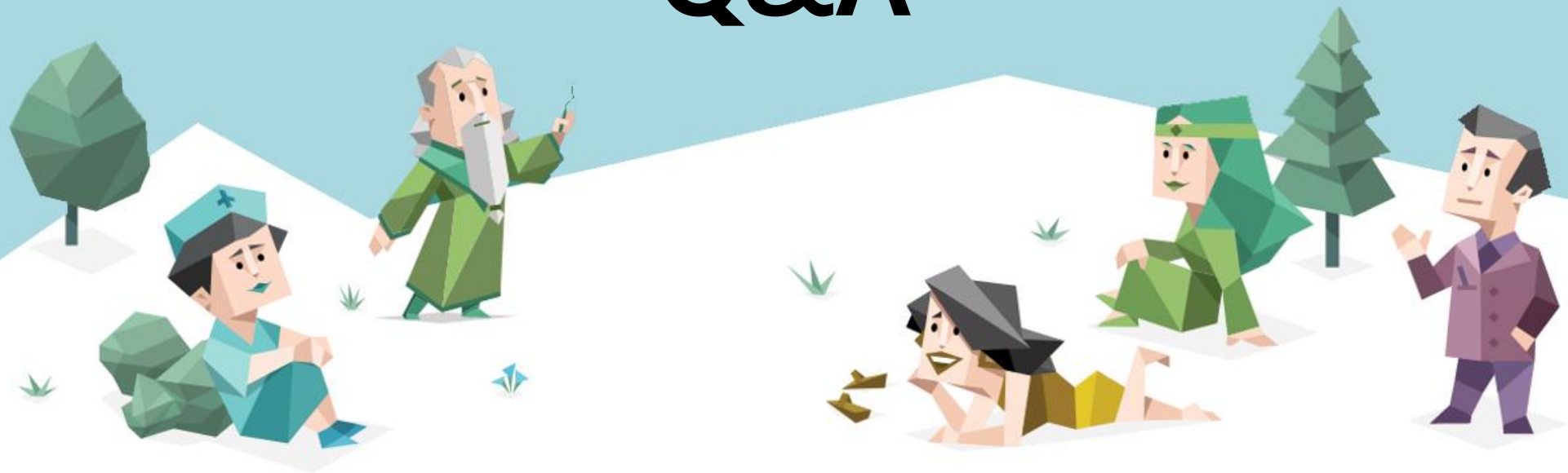
✓ 상용화

: 상용화를 위한 추가적인 고민 필요

✓ 모델링 속도

: 본격적인 모델링 진행 시 속도가 느릴 수 있음 (돌려보아야 알 수 있음)

Q&A



감사합니다

BOAZ STALKER

