

# <카페 자영업자를 위한 고객 상담 보조 챗봇>

고객과 자영업자 여러분의 소중한 시간을 아껴드립니다.

2022. 07. 27. AI 13기 황인환

# 발표 순서

---

## 1. 문제 정의

- 요식업 자영업자의 현실
- 어떤 문제를 풀고 싶은가?

## 2. 구현 기능 소개

- 사용한 데이터셋 소개
- 구현 기능 소개

## 3. 앞으로의 과제

- 한계점 및 향후 목표

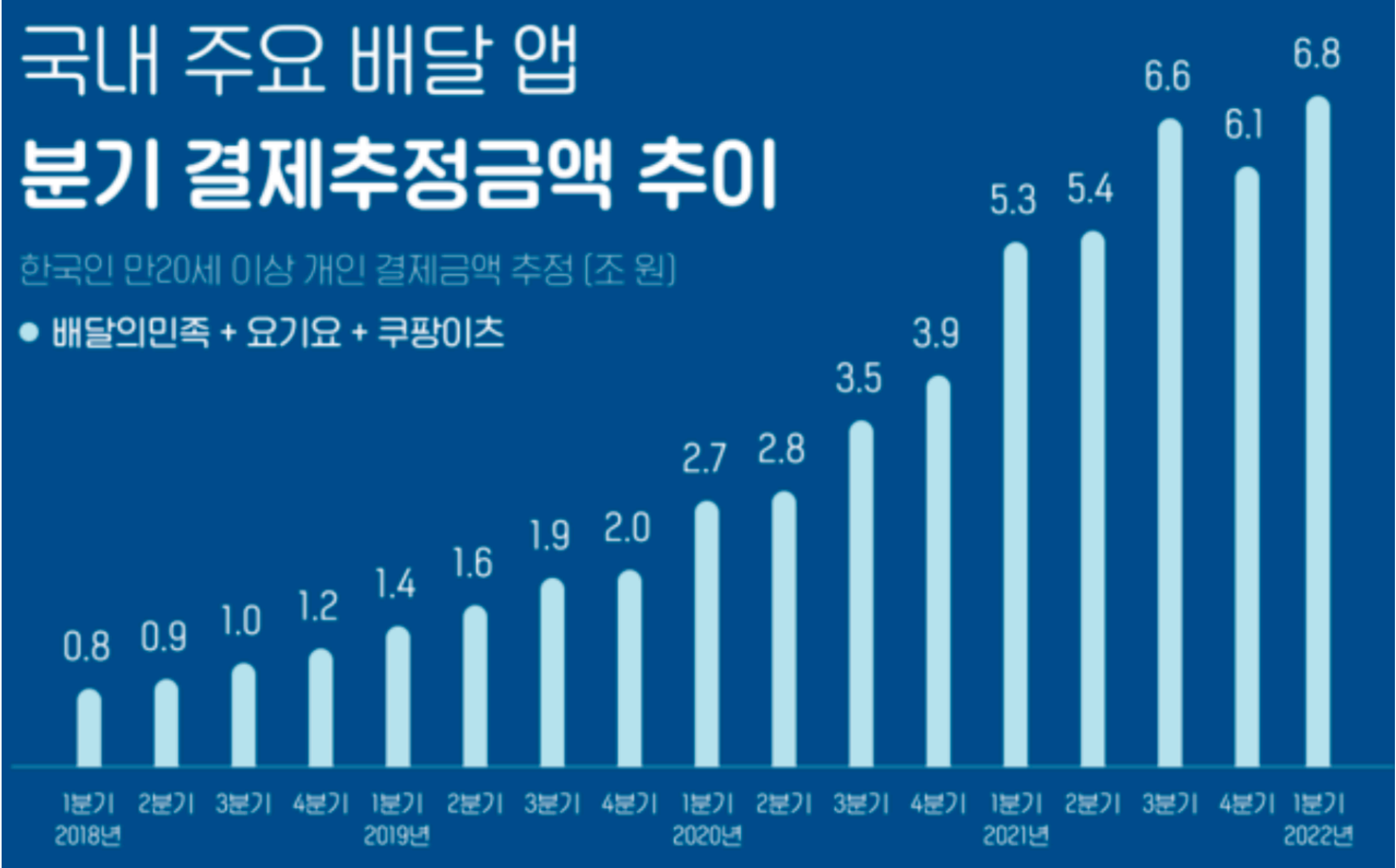
# Part 1. 문제 정의

# Part 1-1. 요식업 자영업자의 현실

## 팬데믹으로 인한 배달시장의 폭발적인 성장



2020년 던햄비 ‘코로나19가 소비자에 미친 영향’ 보고서



# Part 1-1. 요식업 자영업자의 현실

## 배달앱 정책에 따른 자영업자의 부담 상승



## 배달앱 포장수수료 부과 압박, 자영업자들 부글부글

✎ 배소현 기자 | Ⓞ 입력 2022.07.25 17:24 | 💬 댓글 0

## 배달의민족 수수료 인상에 자영업자 '곡소리'

류범열 기자 | ryu4813@chosun.com 등록 2022.04.01 17:29

## "자영업자들이 음식값 올려 방패막이 하란건가" 배달수수료 인상에 '막막'

기사등록 2022/04/04 17:23:27  
최종수정 2022/04/04 17:50:40

🔗 가가

출처: [https://newsis.com/view/?id=NISX20220531\\_0001892158](https://newsis.com/view/?id=NISX20220531_0001892158)  
[https://news.v.daum.net/v/20220628175259385?x\\_trkm=t](https://news.v.daum.net/v/20220628175259385?x_trkm=t)  
[http://www.dizzotv.com/site/data/html\\_dir/2022/04/01/2022040180191.html](http://www.dizzotv.com/site/data/html_dir/2022/04/01/2022040180191.html)  
[https://mobile.newsis.com/view.html?ar\\_id=NISX20220404\\_0001820179](https://mobile.newsis.com/view.html?ar_id=NISX20220404_0001820179)



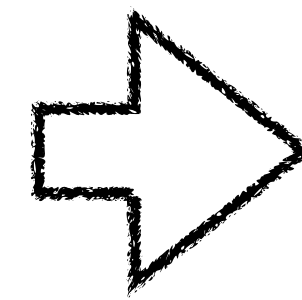
# Part 1-1. 요식업 자영업자의 현실

---

## BIG 프랜차이즈는 자체 앱 마케팅 강화

### 외식 프랜차이즈의 독립, 자체 배달 앱 개발에 박차

배달 수수료 피해 '탈배달앱'하는 프랜차이즈  
자체 배달 앱 시스템 구축으로 충성 고객 끌어 모아



**BUT!!**  
**자본력이 미약한 자영업자들은?**

"배달 수수료 너무 비싸요"...자사앱 키우는 외식업체들

(cf. 이커머스 시장에 불었던 D2C 열풍과 같은 맥락)

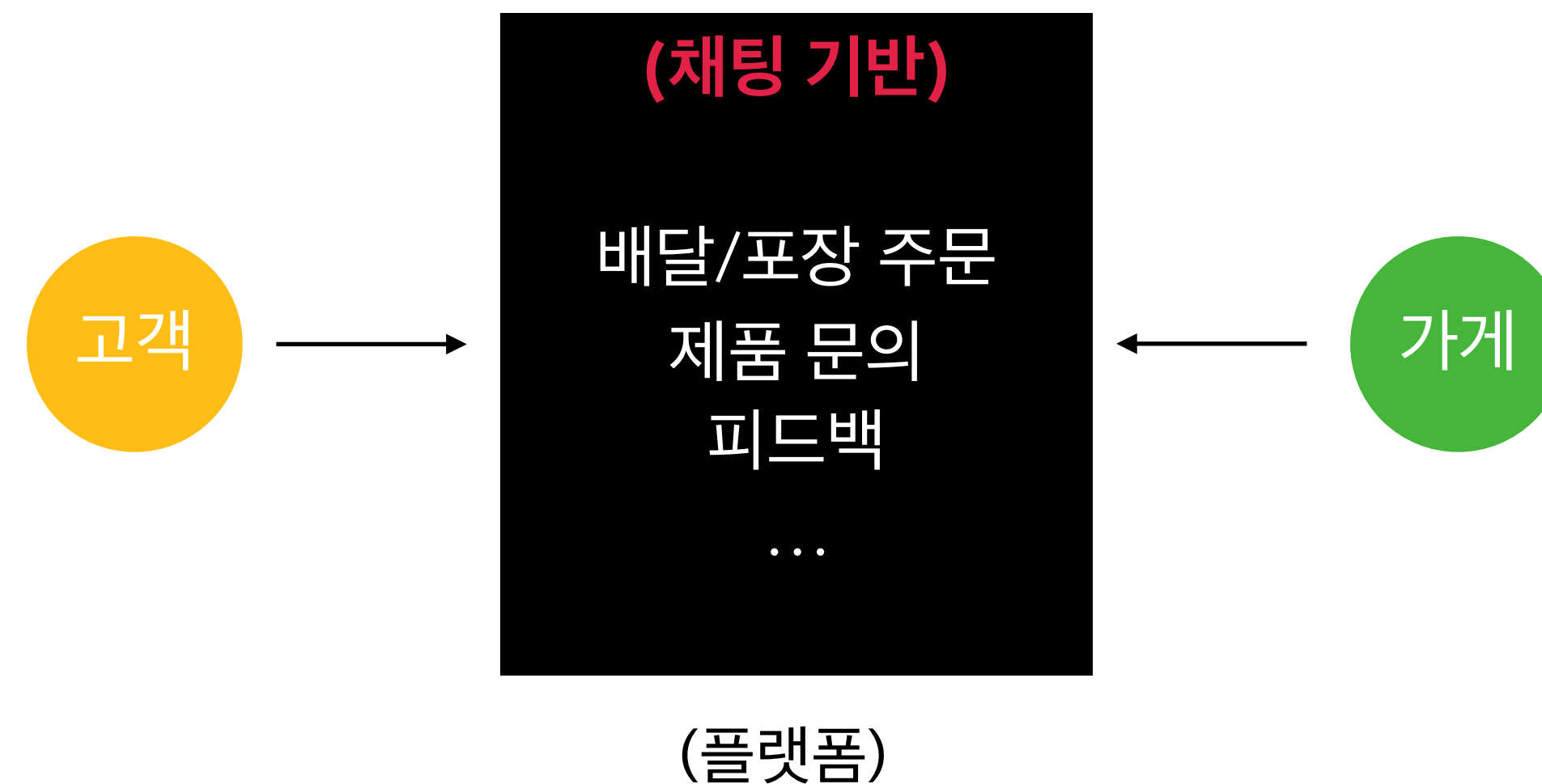
## Part 1-2. 어떤 문제를 풀고 싶은가?

---

### 아이디어 - 소통 플랫폼

자영업자들이 고객과 직접 얘기할 수 있는 소통 플랫폼이 있다면 어떨까?

---



수익구조 - 구독 기반 플랫폼 이용료

## Part 1-2. 어떤 문제를 풀고 싶은가?

---

가설 - 고객과 소통의 보조 도구로써 NLP 기술 접목

**가설1:** 자영업자들은 매우 바쁘므로 채팅을 다 직접 치는 것이 부담될 것이다.

**가설2:** 요식업종의 경우 고객으로부터 받는 대화의 종류가 어느 정도 정해져있을 것이다.  
(ex. 해당 가게에서 파는 메뉴에 대한 문의 / 배달, 주문 관련 문의나 요청 등)



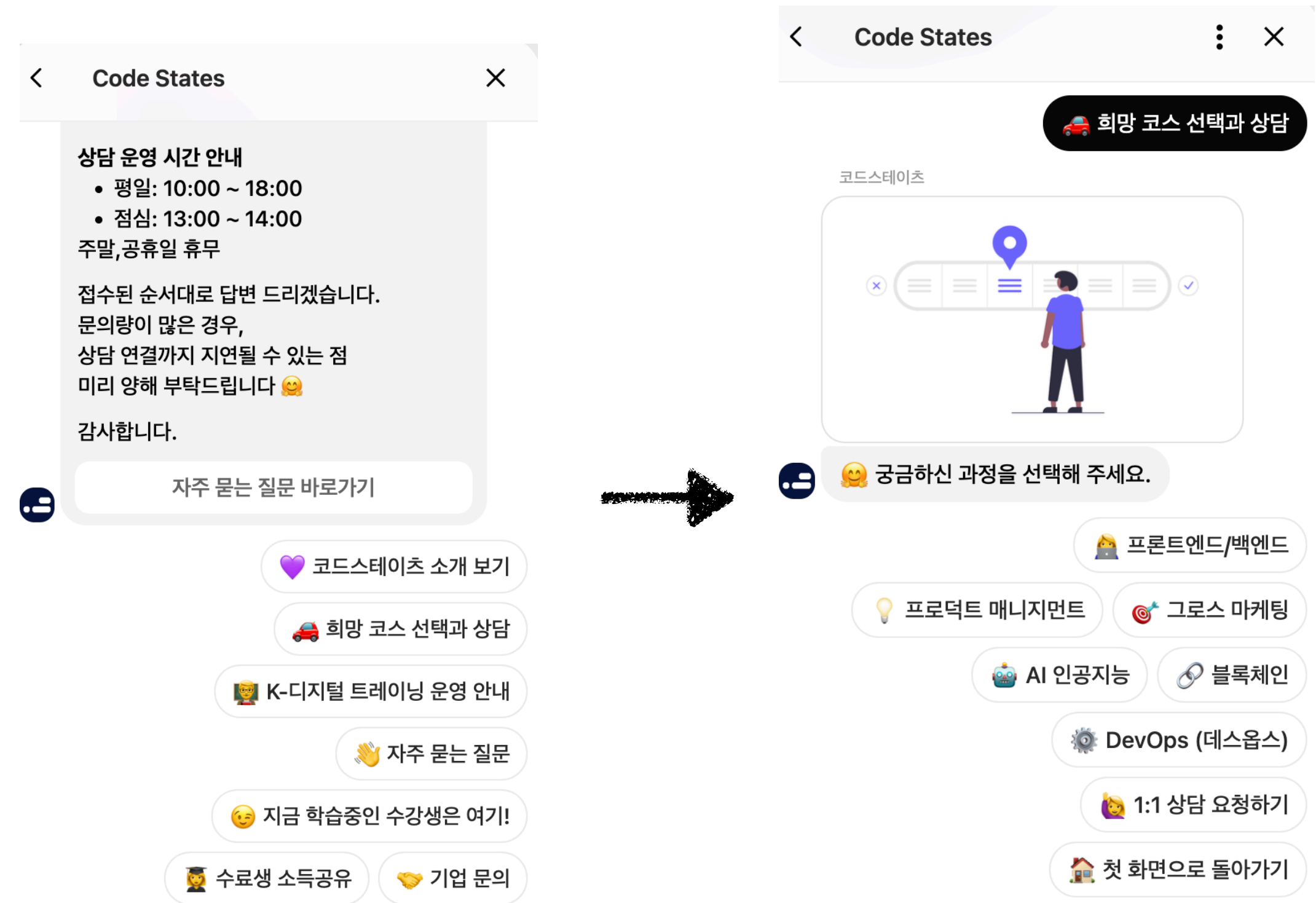
만약 과거 고객 상담 데이터에 기반하여  
**고객의 문의사항에 대해 자동으로 답변을 완성시켜주는 기술**이 있다면,  
자영업자들의 고객 소통 리소스 감소에 도움이 될 것이다.

Ex) 고객이 채팅으로 배달/주문/메뉴 등에 관한 질문을 주면, 점원의 입력창에는 NLP 모델 기반으로 예상 답변이 입력되어있고, 점원은 확인 후 엔터만 누르면 됨.



# Part 1-2. 어떤 문제를 풀고 싶은가?

## 참고 - (개인적으로 느끼는) 고객 입장에서 FAQ 기반 챗봇의 불편함



(코드스테이츠 홈페이지 채널톡 캡처 - 잘 세팅된 예시)

**챗봇의 목적: 고객이 빠르고 정확하게 원하는 정보를 얻을 수 있도록 해 고객만족을 높인다.**

- ▶ FAQ 기반 챗봇은 제대로된 flow로 템플릿을 세팅하는 것이 매우 중요. (기업 입장에선 관리 리소스 증가)
- ▶ 관리가 안될 경우 고객은 수 많은 단계를 거쳤음에도 원하는 정보를 얻지 못해 이탈할 가능성 증가. (혹은 1:1 상담으로 직유입)  
**즉, FAQ 챗봇은 상담원의 직접 상담과 병행될 수는 있어도 완전히 대체될 수는 없음.**

## Part 2. 구현 기능 소개

# Part 2-1. 사용한 데이터셋 소개

## 데이터셋 출처 및 구성

### 소개

상점 서비스 운영 시 발생할 수 있는 다양한 상황에서의 고객 질문과 그에 따른 답변으로 구성된 데이터.

### 구성

가구인테리어, 건강, 패션, 식품, 카페 등 14개 업종으로 구분하여 파일이 나뉘어 있음.

데이터 출처	질의응답데이터 규모 (최종 산출물 기준)	도메인
콜센터 데이터	400만 건 질의응답	백화점, 홈쇼핑, e-commerce 등 유통관련
녹취 데이터	100만 건 질의응답	도.소매업, 숙박.음식업점, 수리, 기타개인서비스업, 보건 업 등에 해당하는 약 20종 상점(한국표준산업분류 10차 기준)

### 데이터 출처 - AI Hub ‘소상공인 고객 주문 질의-응답 텍스트’

본 AI데이터는 한국지능정보사회진흥원의 사업결과임을 밝히며, 원데이터는 위 링크를 통해 확인하실 수 있습니다.

# Part 2-1. 사용한 데이터셋 소개

## 사용한 데이터셋 소개 - ‘카페’ 카테고리

	IDX	발화자	발화문	카테고리	QA번호	QA여부	감성	인텐트	가격	수량	크기	장소	조직	사람	시간	날짜	상품명	상담번호	상담내순번
0	1	s	고객님 혹시 무슨 문제 있으신가요?	카페	5	q	m	제품_불량_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9252	9
1	2	c	네 이상해요	카페	5	a	m	제품_불량_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9252	10
2	3	c	빙수 위에 뿌려진 과일에서 쉰 냄새가 나는데 어떻게 하면 되죠?	카페	20591	q	n	제품_불량_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	빙수 과일	20591	1
3	4	s	죄송합니다, 폐기처분 하고 새 빙수를 제조해드리겠습니다.	카페	20591	a	m	제품_불량_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	빙수	20591	2
4	5	c	마카롱 주문했는데 마카롱이 엄청 딱딱해요.	카페	66996	q	n	제품_불량_질문	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	마카롱	66996	1

총 169,633건

### ['카페' 카테고리를 선택한 이유]

- 1. 업종 특성 상 응대 직원이 상시적으로 음료 제조에 투입됨. (NLP 자동완성 기술에 대한 니즈가 높을 것으로 예상)
- 2. 판매 상품의 종류 및 고객 요청사항이 비교적 정형화되어 있고 대화가 짧아 NLP 모델의 일반화에 좀 더 유리하다고 판단됨.  
(Ex. 패션 카테고리의 경우 고객 개개인의 특성이 답변 선택에 있어 특수하게 적용되는 경우가 많음.)

## Part 2-2. 구현 기능 소개

### 구현 기능 프레임워크 소개



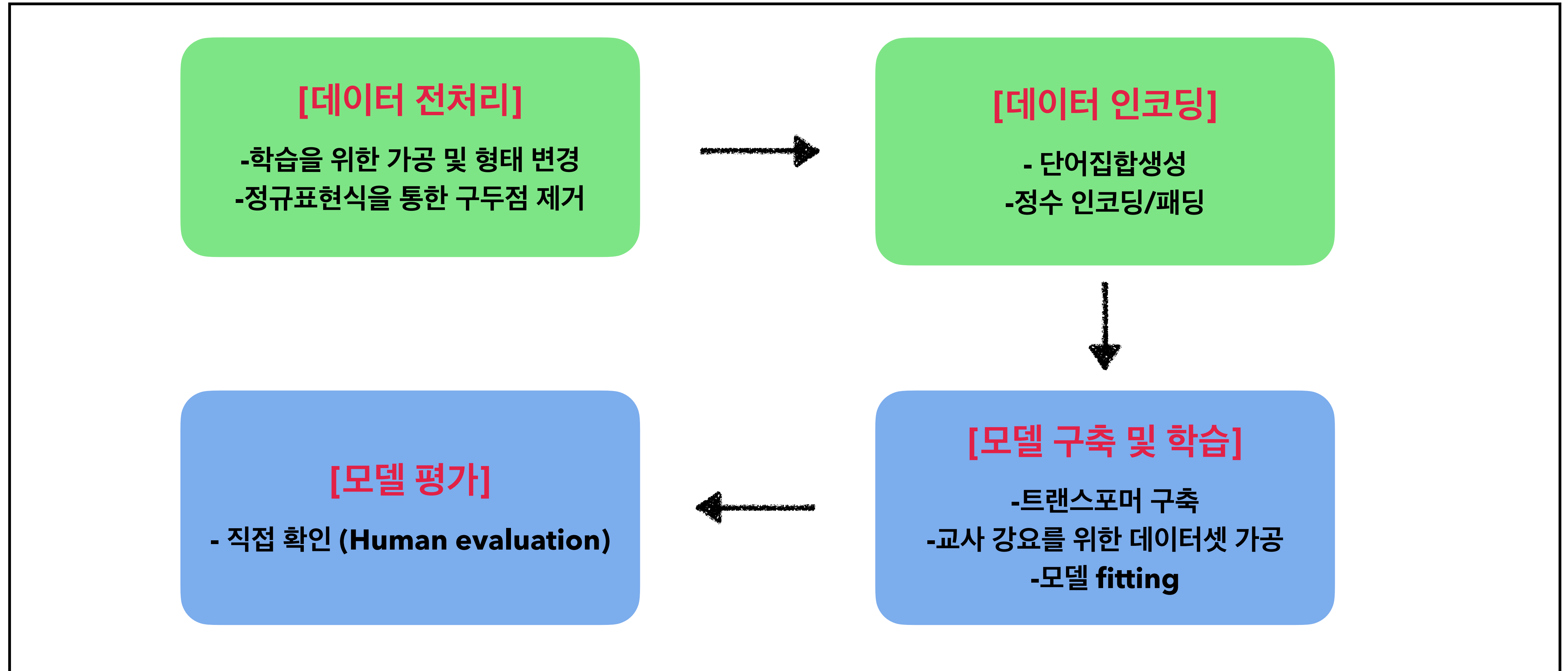
### [Transformer 모델을 선택한 이유]

1. 구현 기능의 주요 목표는 고객의 질문을 정확히 이해하고 그에 맞는 답변을 내주는 것인데, 트랜스포머는 Attention 매커니즘이 극대화되어 있어 자연어 처리 분야에서 매우 우수한 성능을 보이는 모델임.
2. GPU 연산에 최적화되어 있어 빠른 학습이 가능한 장점이 있는데, 이는 추후 추가 학습 등 확장성 측면에서 유리함.

## Part 2-2. 구현 기능 소개

---

### 기능 구현 과정 개요





# Part 2-2. 구현 기능 소개

## 기능 구현 과정 상세 - 데이터 전처리 전

발화자	발화문	QA번호	QA여부	감성	인텐트	상담번호	상담내순번
c	여기 화장실은 어디에 있죠?	7	q	m	매장_부대시설_질문	4	4
c	화장실 비밀번호 있나요?	9	q	m	매장_부대시설_질문	4	5
s	0000# 누르시면 돼요.	10	a	m	매장_부대시설_질문	4	5
s	네, 화장실에 데리고 가셔도 됩니다.	14	a	m	매장_부대시설_확인	4	7
s	네, 같이 내어드릴게요.	6	a	m	매장_이용_요청	4	3
s	죄송하지만 혹시 모를 사고 예방 차원차 반려동물은 말아드리지 않습니다.	12	a	m	매장_이용_질문	4	6
c	강아지 들어가도 되는 매장 맞죠?	1	q	m	매장_이용_확인	4	1
s	네, 같이 들어오셔도 돼요.	2	a	m	매장_이용_확인	4	1
s	아니요, 화장실 옆에 흡연실 위치하고 있습니다.	10	a	m	매장_부대시설_질문	7	5
c	너무 컷이네 그러데 컷다 조 트어주세요	1	q	m	매장_이용_요청	7	1

### [문제점]

- 정확한 학습을 위해선 ‘질문-답변’ 형태로 데이터셋을 구성해야 함.
- 하지만 위와 같이 질문과 답변이 규칙없이 혼재되어 있음.

# Part 2-2. 구현 기능 소개

## 기능 구현 과정 상세 - 데이터 전처리 후

Q	A
쿠키 픽업 예약 드렸었는데 혹시 4일날 픽업으로 바꿔주실 수 ...	네 4일자로 바꿔드릴게요.
A세트 구성에서는 카페라떼가 두 잔 나오는게 맞죠?	네 제일 많이 찾으시는 메뉴로 두 잔 나갑니다.
A세트 주문하려고 하는데 카페모카 나오는거 맞는거죠?	네 세트에는 카페모카로 준비해드리고 있습니다.
A세트 시키면 헤이즐넛라떼가 나오는거죠?	네 달달한 세트라서 그렇습니다.
A세트 주문 넣었는데 자바칩프라페로 들어간 거죠?	네 제가 직접 주문 확인 했습니다.
...	...

### [해결방안]

- 1. 기존 데이터프레임에서 '상담번호' 칼럼 기준 질문과 답변이 딱 한 번씩만 오고간 대화만 남기고 모두 DROP.
- 2. 'Q', 'A' 빈 리스트를 생성하고 for loop를 통해 'QA여부' 칼럼 기준 'q', 'a'를 차례대로 하나씩 뽑아서 리스트에 할당.
- 3. 위에서 더해 추가적으로 정규표현식을 통해 구두점(?!,.,) 제거. (토큰화 용이를 위함)

결과적으로 총 56,163건의 질의응답 샘플 확보.

# Part 2-2. 구현 기능 소개

## 기능 구현 과정 상세 - 데이터 인코딩 (Subword TextEncoder)

단어집합 생성



정수 인코딩/패딩

```
# 질문, 답변 데이터로부터 단어 집합 생성
tokenizer = tfds.deprecated.text.SubwordTextEncoder.build_from_corpus(
    questions + answers, target_vocab_size=2**13)

# 시작 토큰과 종료 토큰에 대한 정수 부여.
START_TOKEN, END_TOKEN = [tokenizer.vocab_size], [tokenizer.vocab_size + 1]

# 시작 토큰과 종료 토큰을 고려하여 단어 집합의 크기를 + 2
VOCAB_SIZE = tokenizer.vocab_size + 2
```

```
# 0번 샘플을 임의로 출력
print(questions[0])
print(answers[0])
```

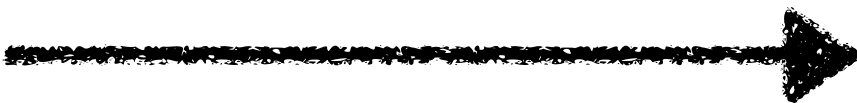
[8294	853	392	691	375	302	8073	19	59	714	41	7101	1854	38
2629	1576	7623	1572	3	7	8295	0	0	0	0	0	0	0
0	0]												
[8294	47	2629	388	19	652	8295	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0]												

(Cf. Subword TextEncoder 장점 - OOV가 없음)

# Part 2-2. 구현 기능 소개

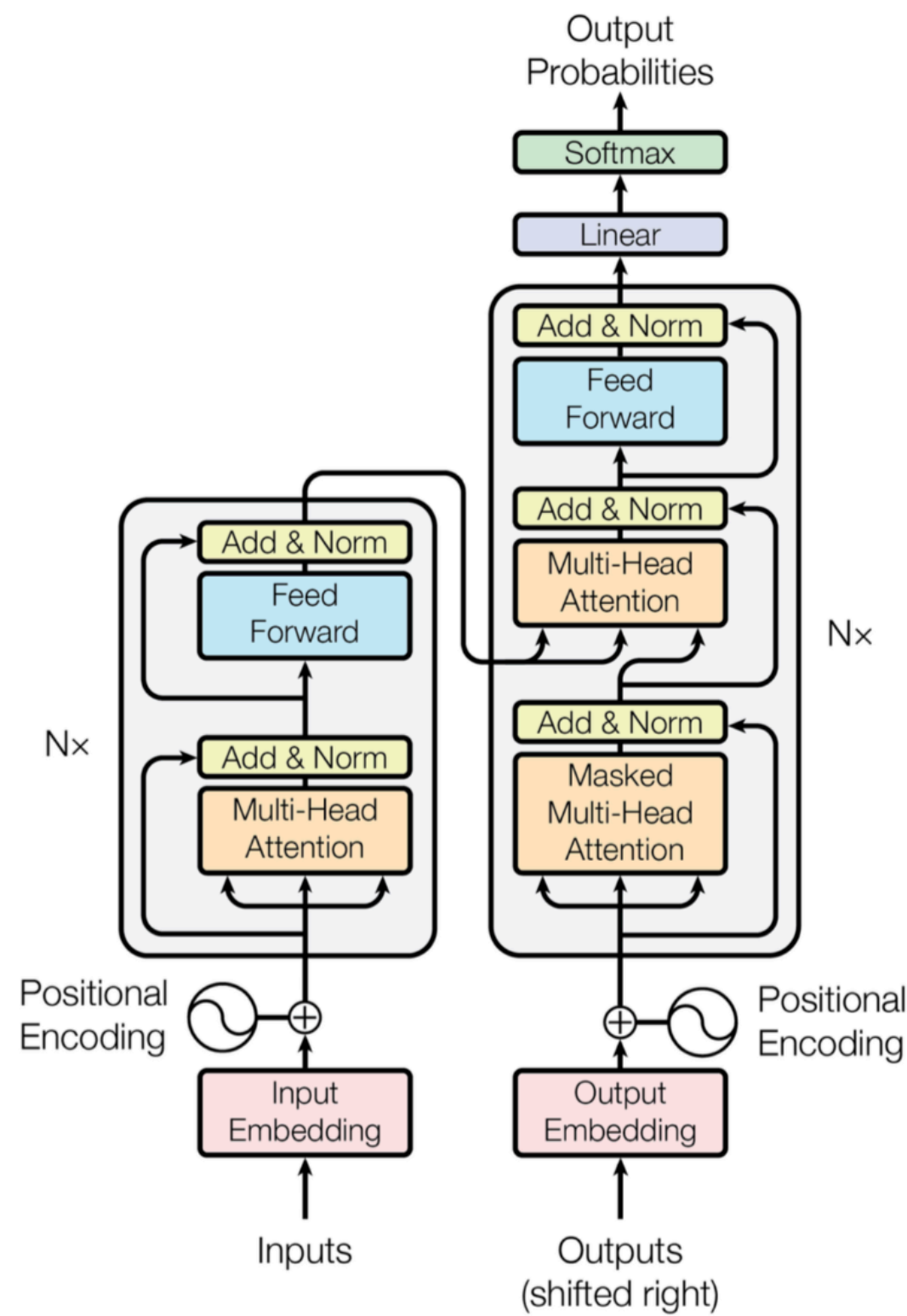
## 기능 구현 과정 상세 - 모델 구축 및 학습

모델 구축



-교사 강요를 위한 데이터셋 가공  
-compile

모델 학습



```
# 모델 학습
EPOCHS = 10
model.fit(dataset, epochs=EPOCHS)
```

```
Epoch 1/10
878/878 [=====] - 55s 51ms/step - loss: 1.7743 - accuracy:
Epoch 2/10
878/878 [=====] - 45s 52ms/step - loss: 1.2755 - accuracy:
Epoch 3/10
878/878 [=====] - 45s 51ms/step - loss: 1.0023 - accuracy:
Epoch 4/10
878/878 [=====] - 45s 52ms/step - loss: 0.8299 - accuracy:
Epoch 5/10
878/878 [=====] - 47s 53ms/step - loss: 0.7245 - accuracy:
Epoch 6/10
878/878 [=====] - 45s 51ms/step - loss: 0.6228 - accuracy:
Epoch 7/10
```

하이퍼 파라미터  
d\_model=256, num\_layers=2, num\_heads=8, dff=512, dropout=0.1



## Part 2-2. 구현 기능 소개

### 기능 구현 과정 상세 - 모델 평가 - Human evaluation

#### 자연스러운 출력

```
output = predict('배달 되나요?')
```

Input: 배달 되나요?  
Output: 네 가능합니다

```
output = predict('아메리카노 디카페인 있나요?')
```

Input: 아메리카노 디카페인 있나요?  
Output: 네 있습니다

```
output = predict('아메리카노 얼마예요?')
```

Input: 아메리카노 얼마예요?  
Output: 5000원입니다

```
output = predict('차 종류는 뭐가 있어요?')
```

Input: 차 종류는 뭐가 있어요?  
Output: 루이보스 캐모마일 있습니다

#### 어색한 출력

```
output = predict('여기 뭐가 맛있어요?')
```

Input: 여기 뭐가 맛있어요?  
Output: 비교적 양이 많은 메뉴입니다

```
output = predict('3명 포장 안 되나요?')
```

Input: 3명 포장 안 되나요?  
Output: 네 안 됩니다

## BUT!!

인간의 판단이 비교적 덜 들어가는 기본 응대는  
충분히 가능해보이므로 이 정도면 괜찮다고 판단됨.

## Part 3. 앞으로의 과제



# Part 3. 앞으로의 과제

---

## 한계점 및 향후 목표

### 1. 서비스 측면

고객 소통 보조 챗봇 전체 기능 중 극히 초입인 모델 구현에만 그쳤다는 점. (= 실제 활용성 미검증)

→ 간단하게라도 모델을 활용해 ‘질문-응답’이 이루어지는 APP을 구축해 사용성 테스트를 해보자.

### 2. 데이터 측면

다양한 업종으로 확장하기 어렵다는 점. (ex. 같은 질문이어도 고객에 따라 다른 답변이 나가는 패션 업종)

→ 데이터 전처리 시 이런 특성을 가진 질문에 대해선 적절한 처리가 이루어질 수 있도록 시도해보자.

### 3. 구현 모델 측면

다양한 성능 지표를 통한 모델의 객관적 지표 평가가 이루어지지 않았다는 점.

→ 성능 지표를 활용한 평가도 해본 후 출력값의 정확도를 높이기 위한 최적화 작업을 진행해보자.

**끝.**

**감사합니다.**