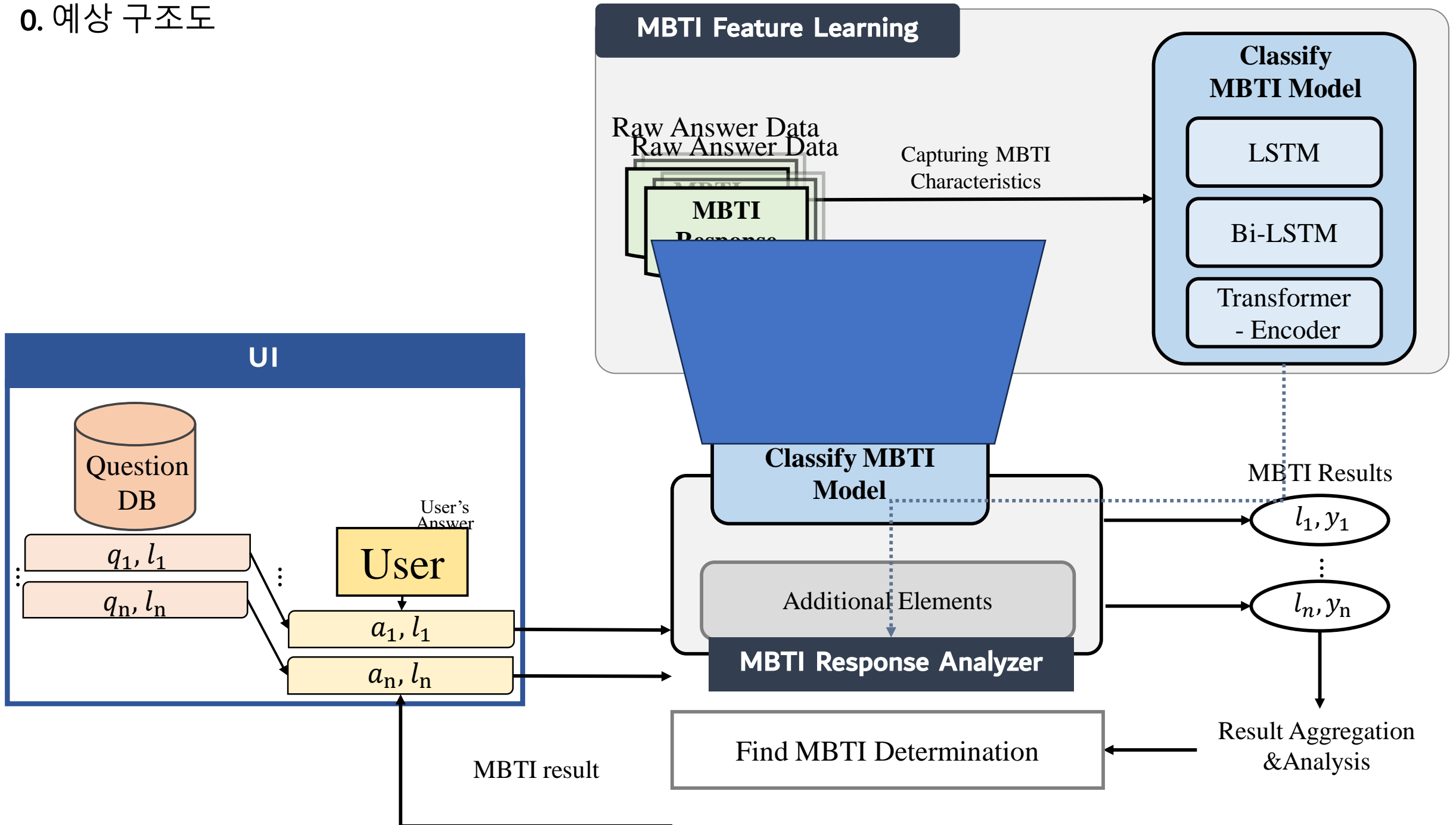
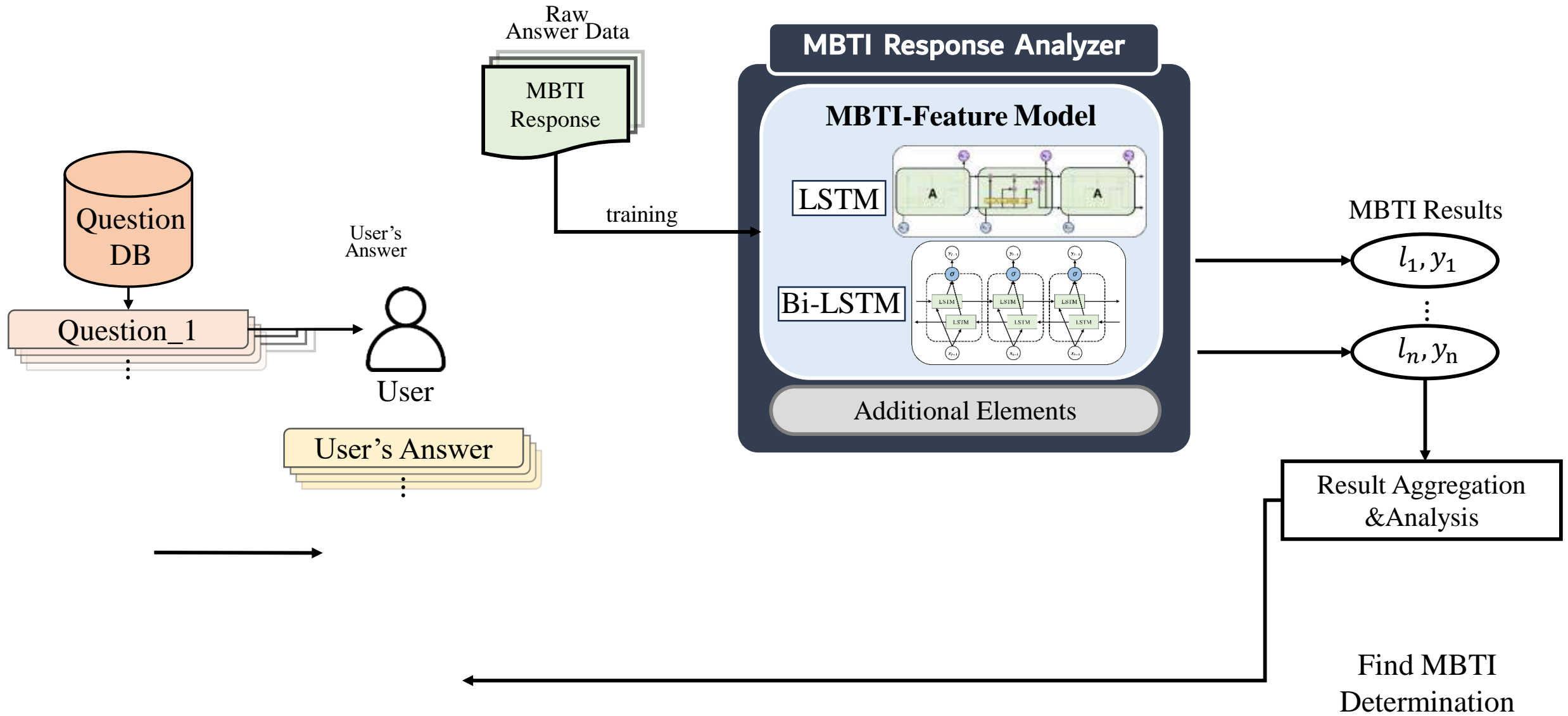


딥러닝 아키텍처를 활용한
자연어 응답기반 MBTI 분석

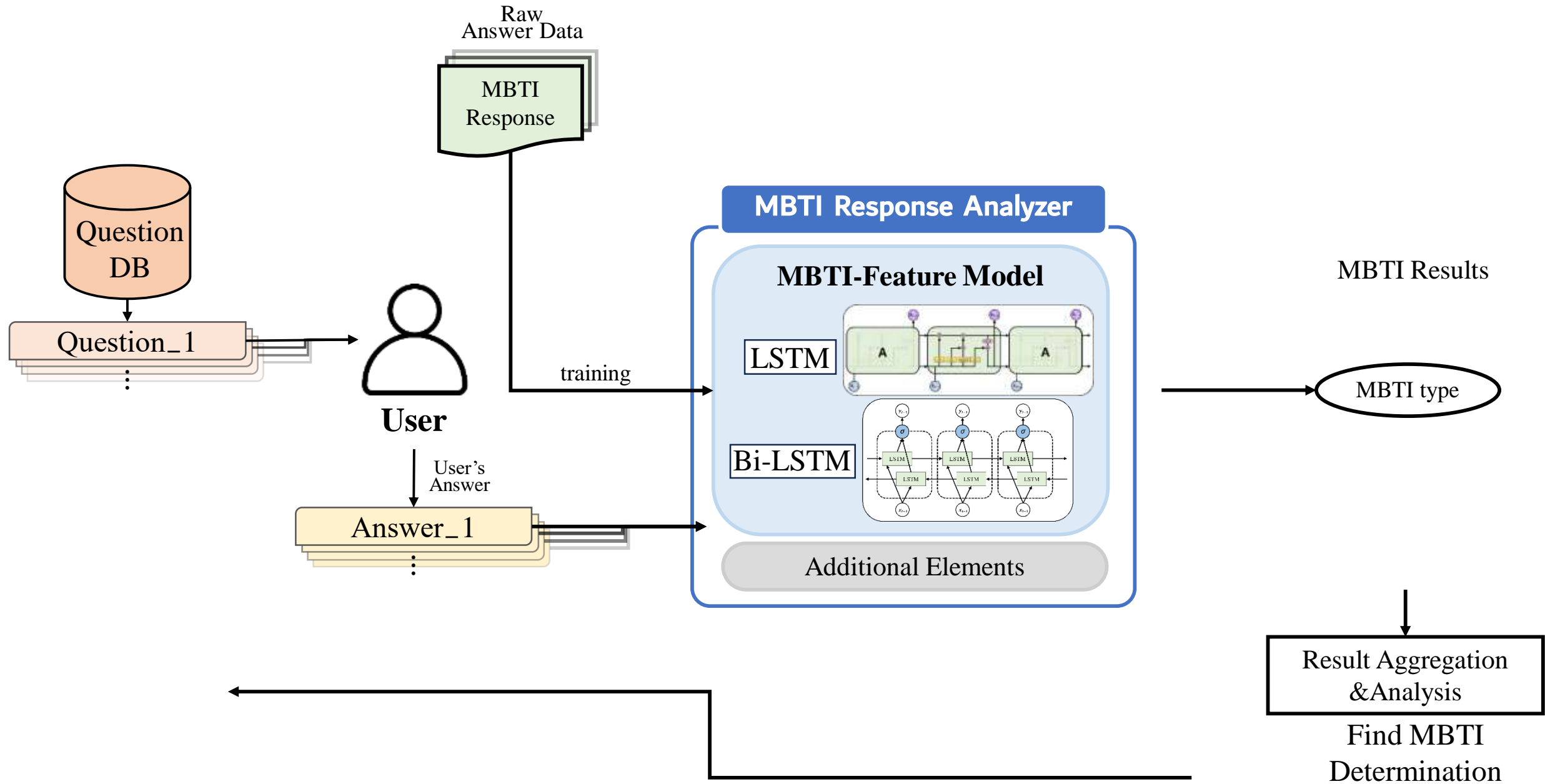
0. 예상 구조도



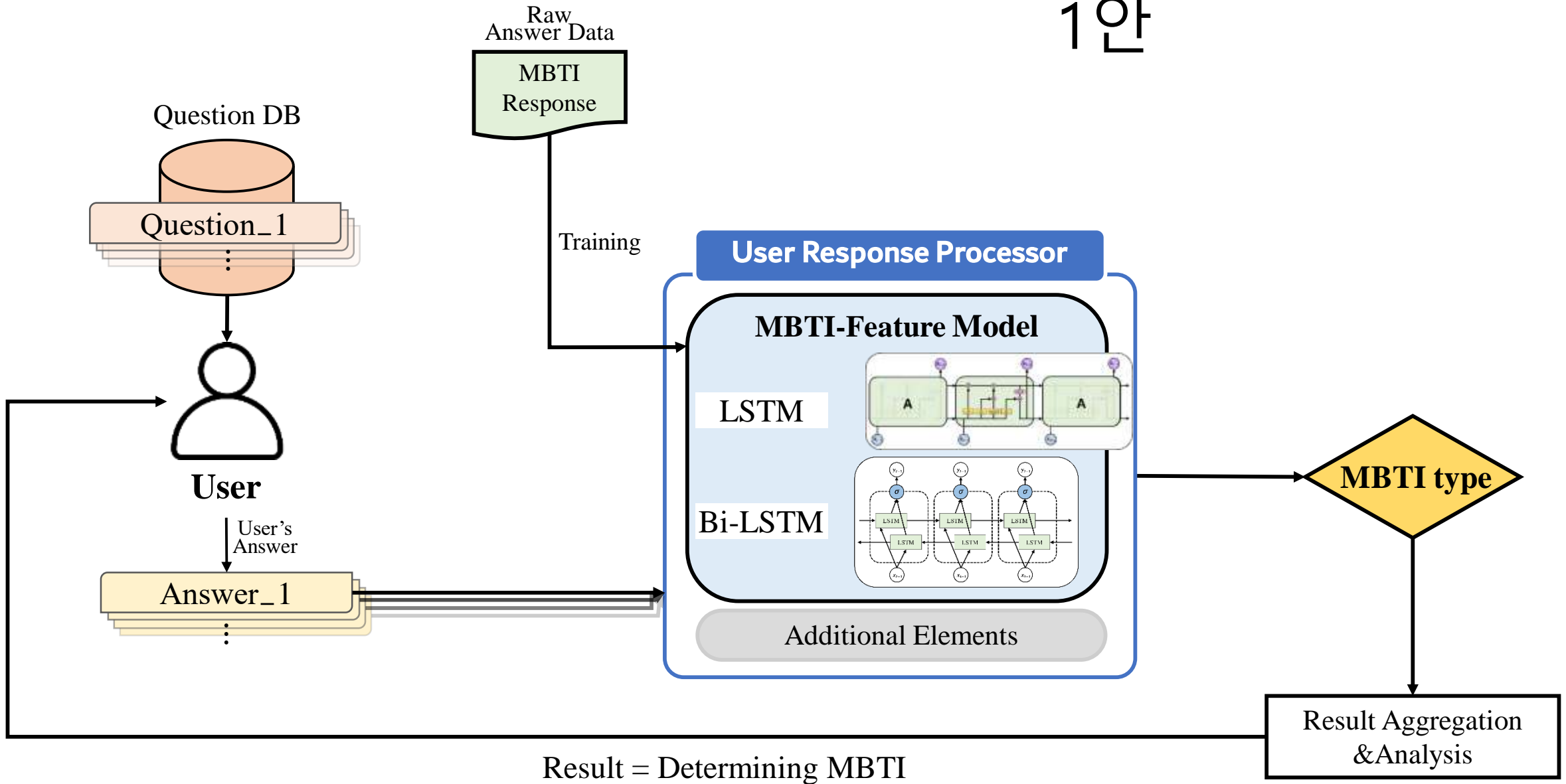
UI

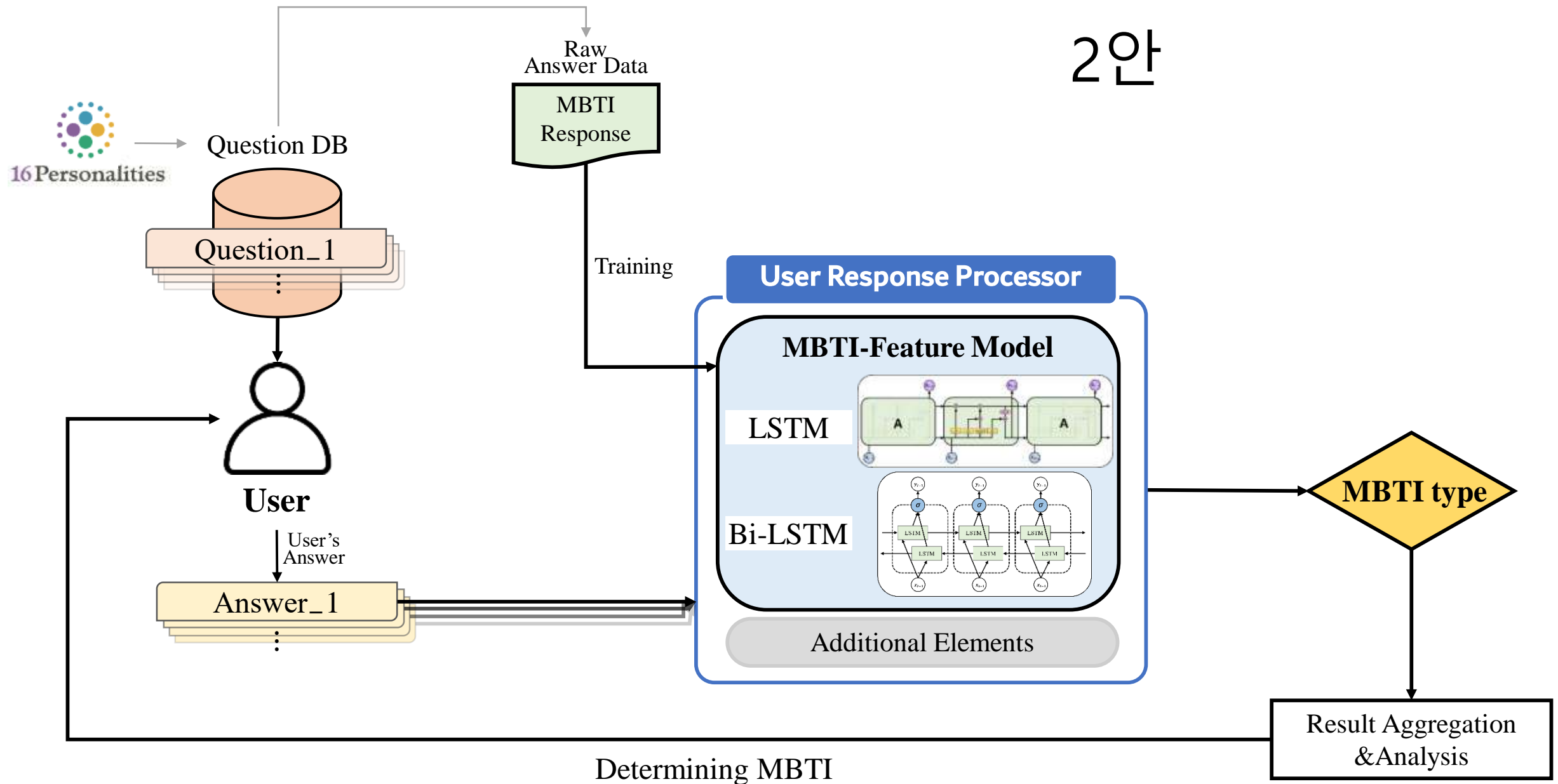


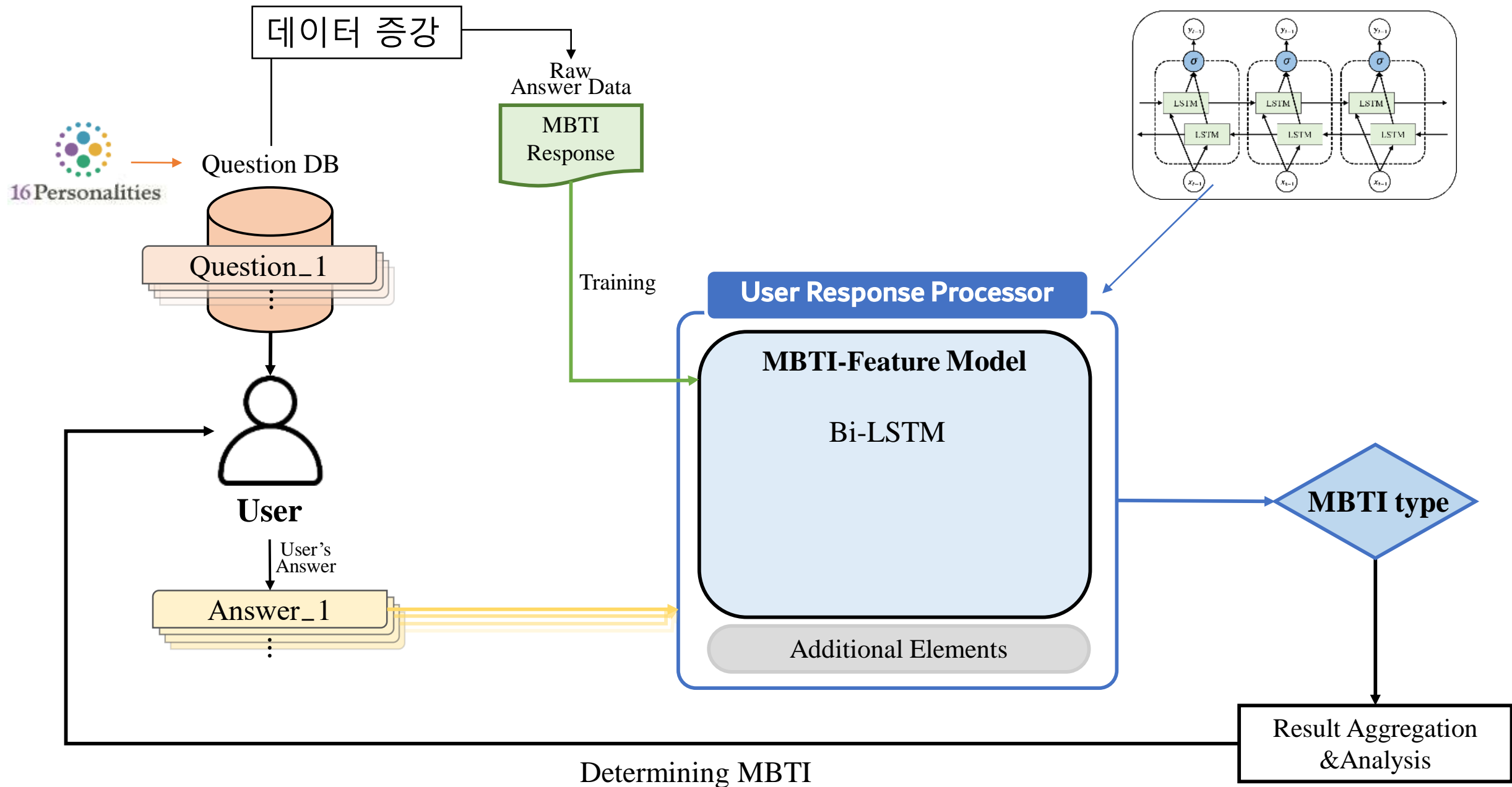
0. 예상 구조도

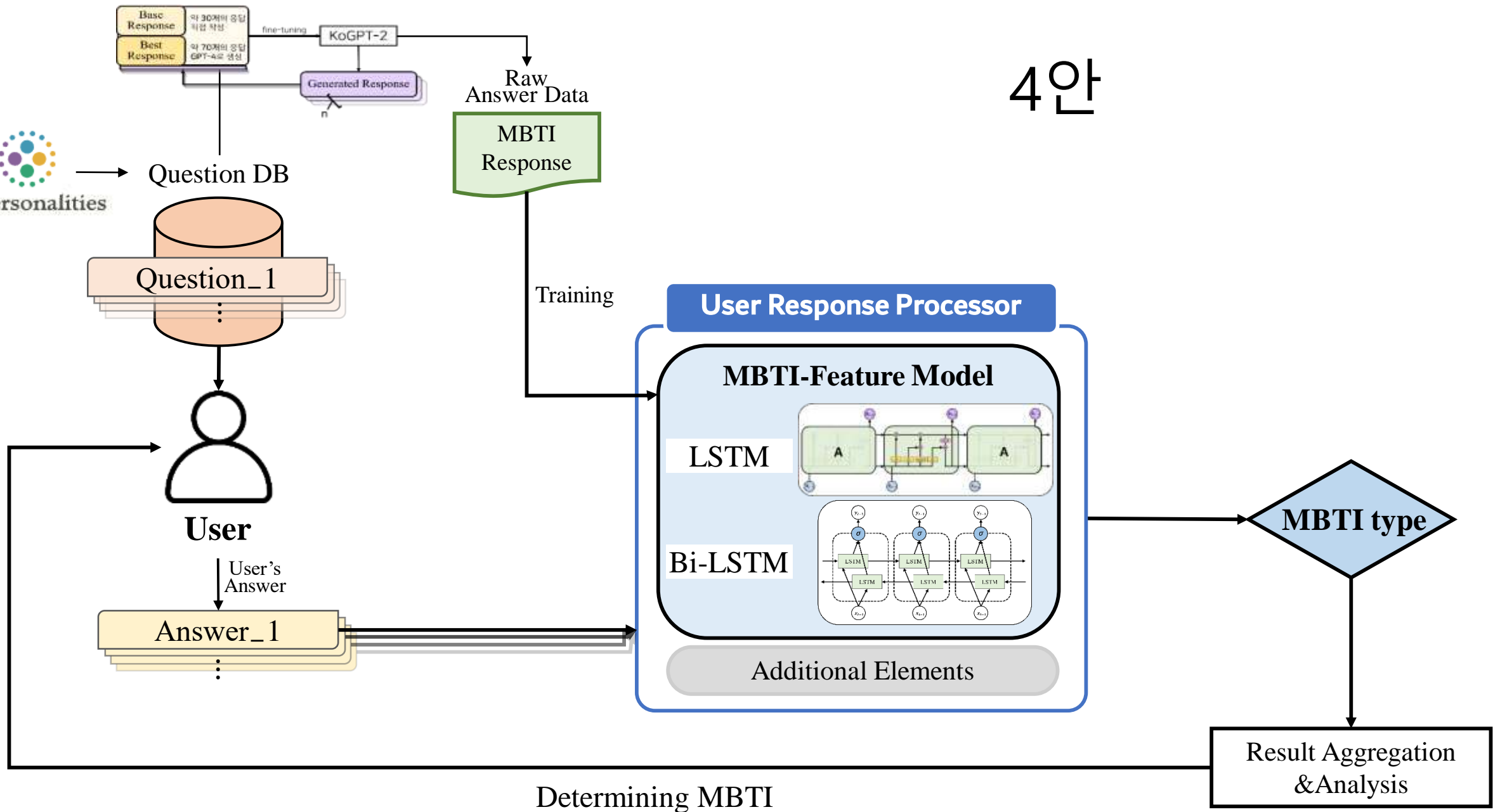


1안

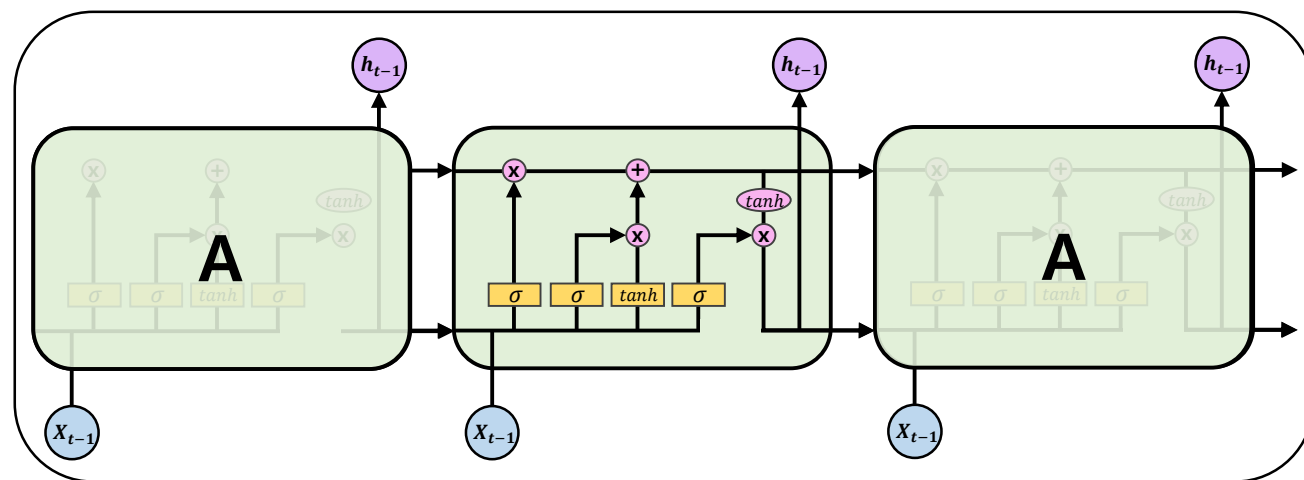
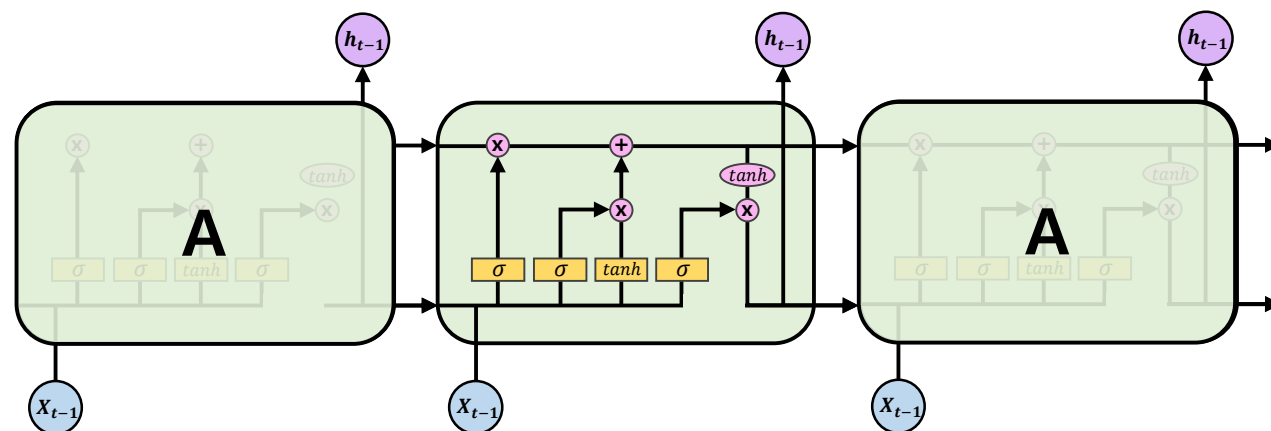
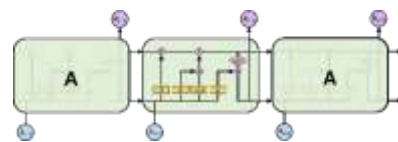


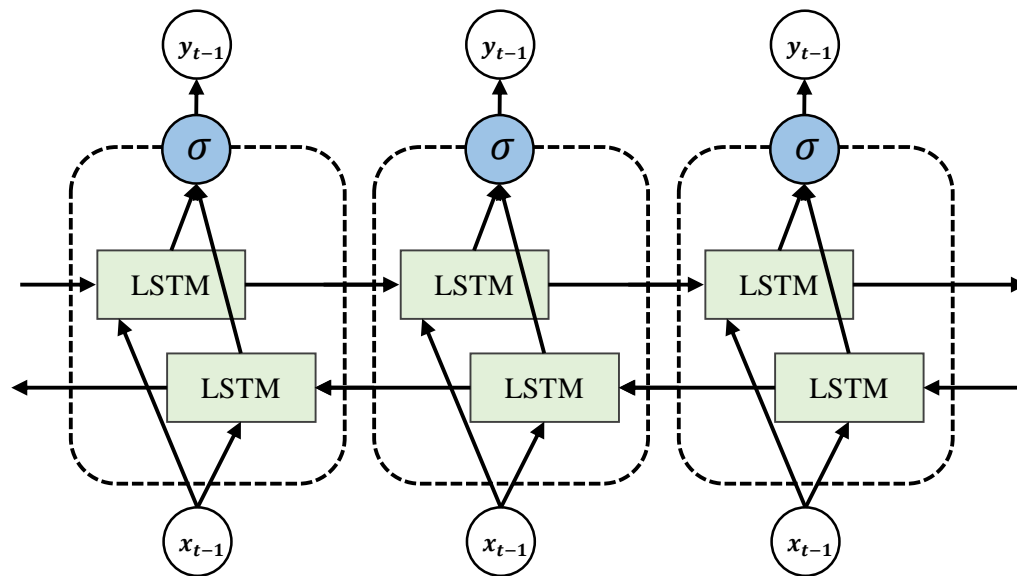
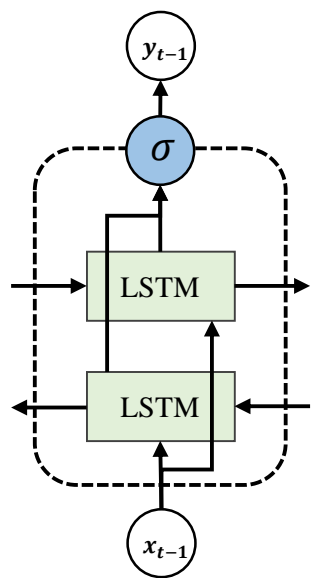




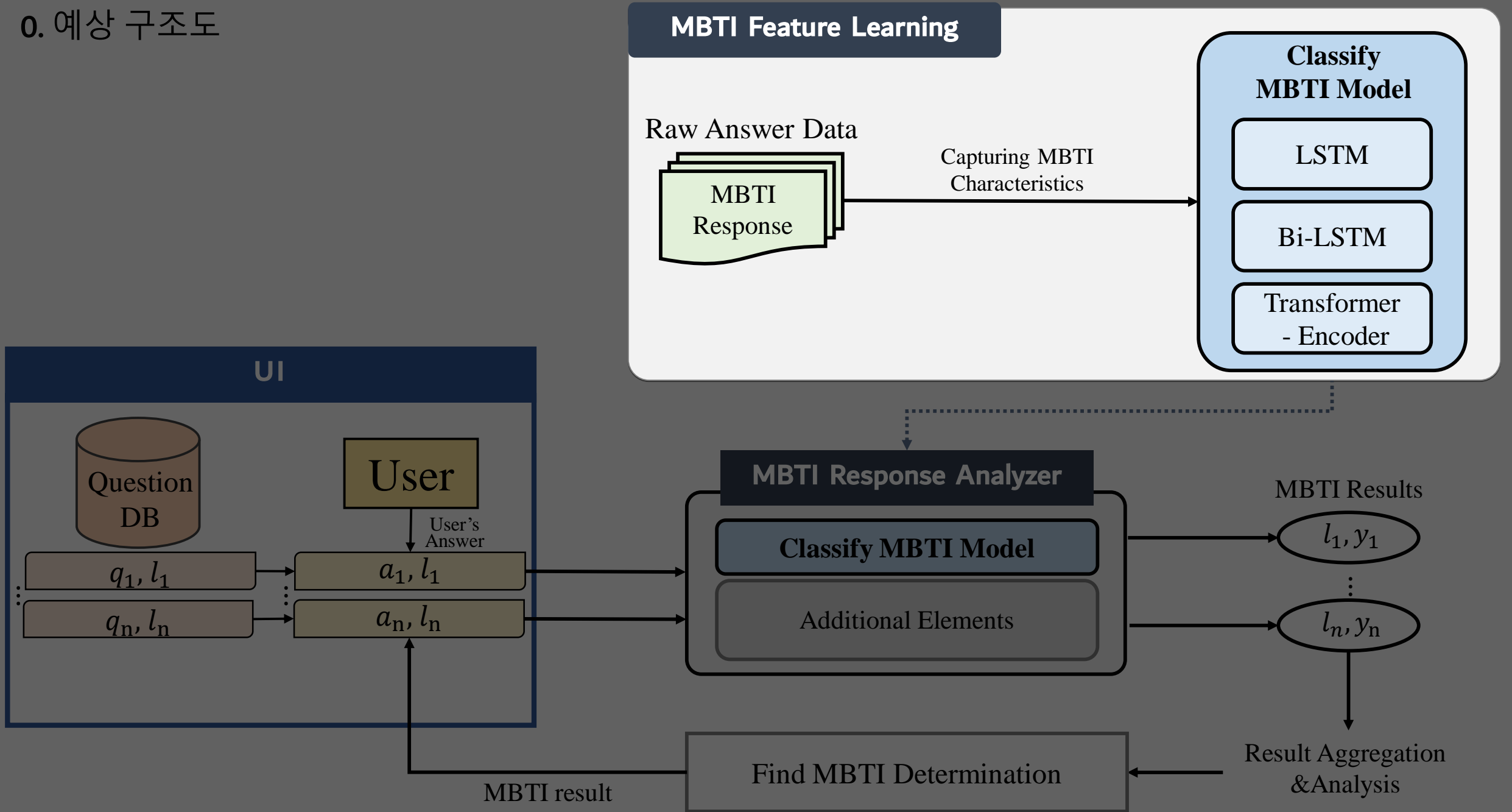


4안

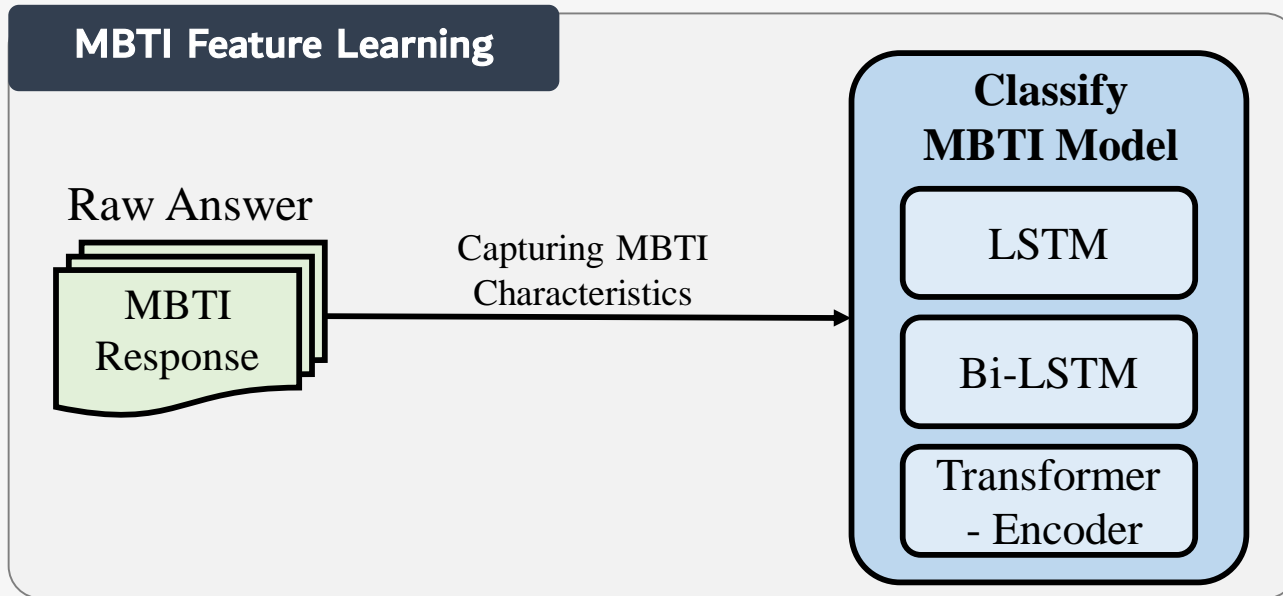




0. 예상 구조도



I . MBTI Feature Learning

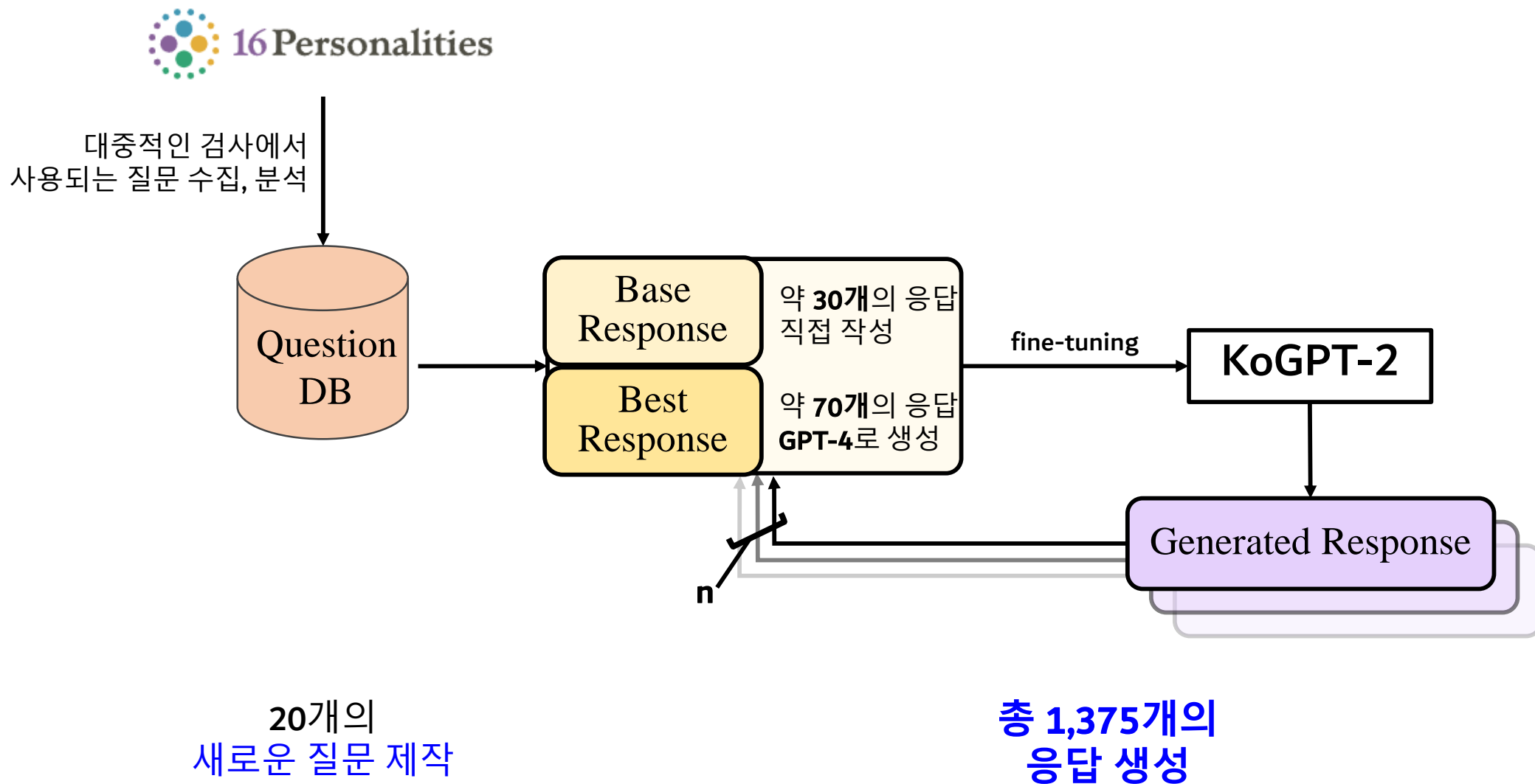


1. 데이터 수집

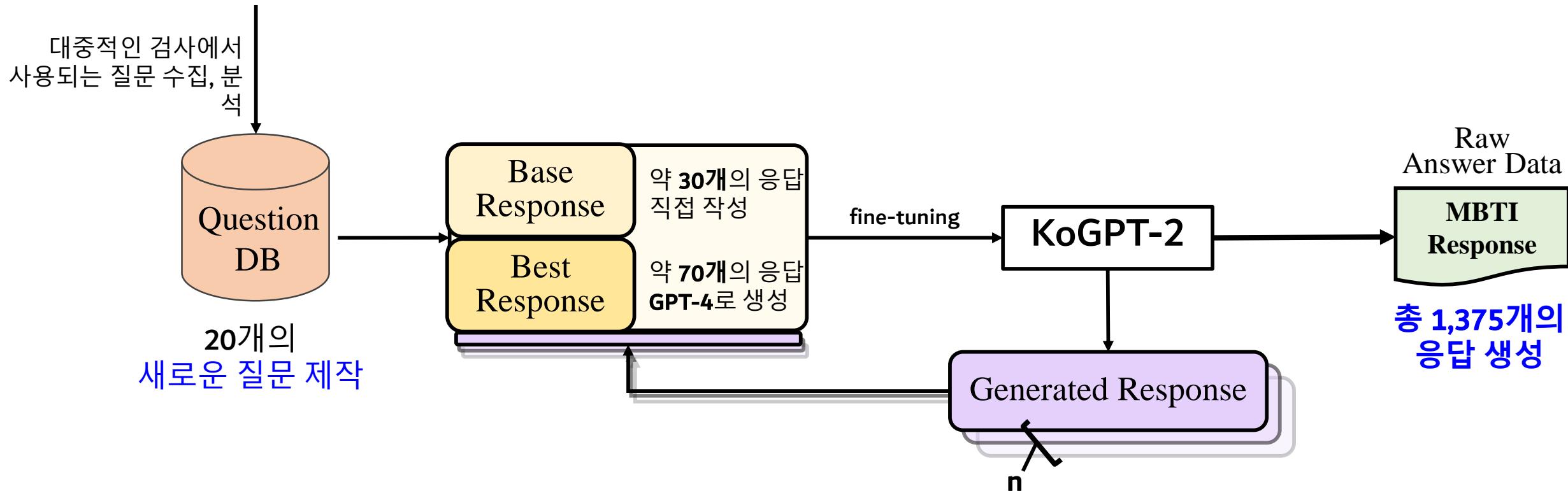
2. **LSTM/Bi-LSTM**을 이용한

MBTI 특징 학습

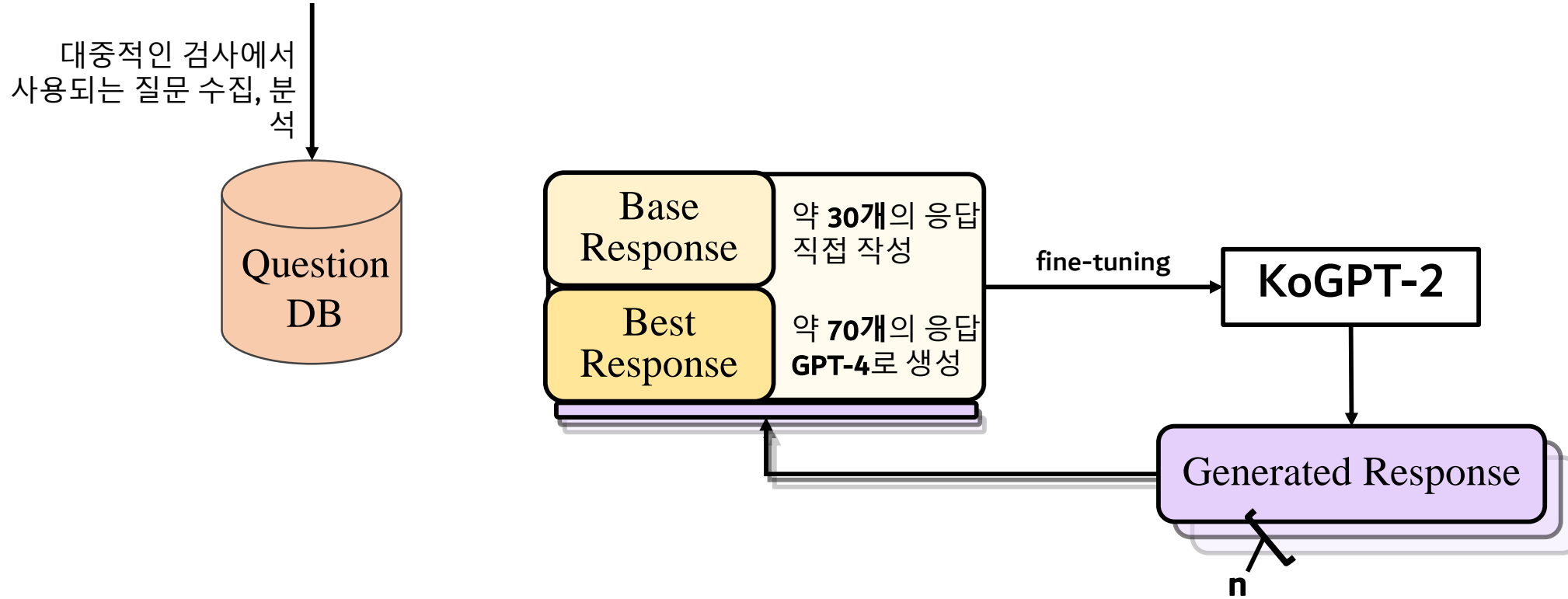
1. 데이터 수집



1. 데이터 수집



1. 데이터 수집

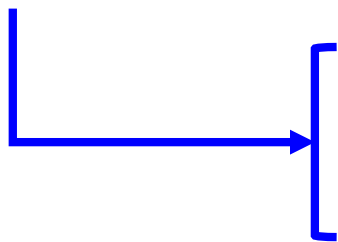


1. 데이터 수집

Raw Answer Data

MBTI
Response

총 1,375개의
응답 생성



Training set + Data Augmentation	1,000개	2,965개
Validation set	375개	-
Test set	38개	-

사람들이 자유롭게 작성한
비정제된 응답 수집하여 사용

우선 친한 사람을 찾고 친한 사람 없는 곳이라면.....그냥 영혼없이 구석에 박혀있거나 말 걸어줄 때까지 기다림

준비 다 하고 나왔는데 갑자기 취소되면 빠치지만 나오기 전에 취소되면 오히려 좋아? 그래도 짜증나고 싫다기보단 좋아,,,행복,,느낌이 더 썬듯??

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

LSTM/Bi-LSTM에 2가지 접근법 시
도

1. 임베딩 레이어 선택에 따른 성능 비교

기본 임베딩 사용 vs 사전학습된 임베딩 사용

Pre-trained
fastText -
KO

2. 학습 데이터 선택 범위에 따른 성능 비교

extracted word 사용 vs whole word 사용

문법적 역할을 하는 조사, 어미 등은 제거

주요 정보를 담고 있는 명사, 동사, 형용사, 어간은 추출

Classification model	Use of FastText	Training Data Scope	Model name
LSTM	X	extracted word	Simple-LSTM
LSTM	O	extracted word	Fa-LSTM
Bi-LSTM	X	whole word	Simple-BLSTM
Bi-LSTM	O	extracted word	Fa-extr-BLSTM
Bi-LSTM	O	whole word	Fa-whl-BLSTM

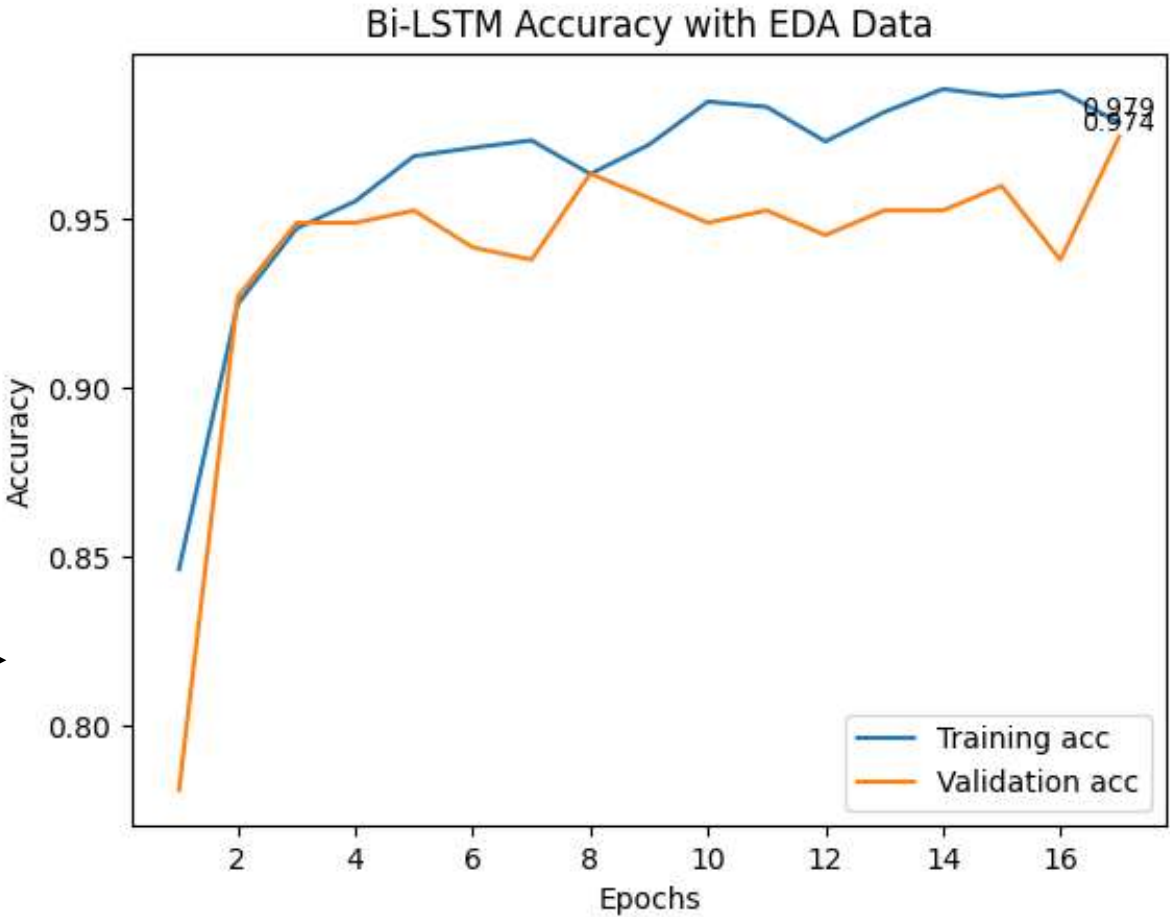
제안된 모델의 학습 방식 비교

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

분류 모델별 성능 비교

Model name	Train	Validation	Test
Simple-LSTM	95.10%	93.60%	92.11%
Fa-LSTM	97.50%	93.30%	86.84%
Simple-BLSTM	97.50%	94.40%	89.29%
Fa-extr-BLSTM	95.50%	93.60%	94.74%
Fa-whl-BLSTM	97.90%	95.30%	94.74%

분류 모델별 성능 비교



모델 분포 그리기

학습 데이터

- 라벨비율 그래프
- 데이터 간 문장 길이 막대그래프

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

높은 성능의 원인

1. 학습 데이터 셋이 주로 유사한 문장구조를 가지기 때문

- 대부분의 질문이 성격 유형을 구분하는 특징적인 질문들이기에, 답변도 비슷한 패턴을 보이는 경향이 있다.

다른 사람과 이야기 할 때 말을 많이 하는 편인가요?	응 수다쟁이야
다른 사람과 이야기 할 때 말을 많이 하는 편인가요?	응 나는 수다쟁이야
다른 사람과 이야기 할 때 말을 많이 하는 편인가요?	응 수다쟁이야 나는

2. MBTI의 각 유형은 서로 상반되는 유형이기 때문

- E (외향) vs I (내향) 이라는 명확한 특징은 모델 학습에서 강력한 분류 기준을 제공

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

단어와 문장에 대한 성격 유형 예측

성격 유형 분류를 위한
대표적인 단어와 문장 예시

Classification Criteria	Word/Sentence	Label	case Identifier
1. Clear Distinction Word	혼자	I	case 1.1
	친구	E	case 1.2
2. Neutral Word	사람	Neutral	case 2
3. Classification Indicator Sentence	친한 친구와만 노는 편이다.	I	case 3.1
	많은 친구와만 노는 편이다.	E	case 3.2
4. User-defined Sentence	주목받은 후 즉시는 살짝 부끄럽지만, 나름 주목받는 것을 좋아합니다.	E	case 4.1
	예기치 못한 상황이면 살짝 당황스러움..	I	case 4.2

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

단어와 문장에 대한 성격 유형 예측

문장입력

입력문장: <그래도 주말에는 집에 있는 게 좋아요..>

out of voca

그래도

에

는

에

는

아요

토큰화

00V 토큰 확인: 00V 주말 00V 00V 집 00V 있 00V 게 좋 00V

정수 인코딩

[[1, 32, 1, 1, 24, 1, 4, 1, 641, 9, 1]]

모델 예측값

예측값 : 0.02986094

MBTI 판별

>> 97.01% 정도로 I 유형입니다.

pre \geq 0.5 -> E

pre < 0.5 -> I

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

단어와 문장에 대한 성격 유형 예측

예측값 : 0.996568

입력문장: <생일에는 주변 친구들과 신나게 파티를 열어서 놀아요~>
>> 99.66% 정도로 E 유형입니다.

00V 토큰 확인: 생일 00V 00V 주변 친구 들 00V 신나 게 파티 00V 열 00V 놀 00V
[[234, 1, 1, 87, 14, 61, 1, 421, 641, 83, 1, 275, 1, 198, 1]]

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

단어와 문장에 대한 성격 유형 예측

			model											
			Actual MBTI		Simple-LSTM		Fa-LSTM		Simple-BLSTM		Fa-ext-BLSTM		Fa-whl-BLSTM	
					Pred	Pred MBTI	Pred	Pred MBTI	Pred	Pred MBTI	Pred	Pred MBTI	Pred	Pred MBTI
E와 I 유형을 명확히 구분하는 단어	case 1.1	혼자	I	80.47%	I	95.67%	I	73.15%	I	91.37%	I	80.59%	I	
	case 1.2	친구	E	78.10%	E	97.18%	E	93.99%	E	99.33%	E	90.54%	E	
중립 단어		case 2	사람	(N)	52.85%	I	94.65%	E	70.57%	E	97.31%	E	61.94%	E
case 1의 단어를 문장으로 확장	case 3.1	친한 친구와 만 노는 편 이다.	I	61.89%	E	97.66%	I	72.75%	I	66.29%	I	58.92%	I	
	case 3.2	많은 친구와 만 노는 편 이다.	E	83.60%	I	74.04%	E	99.87%	E	98.79%	E	98.56%	E	
실제 사용자가 작성한 원본 문장	case 4.1	주목받은 후 즉시는 살짝 부끄럽지만, 나를 주목받 는 것을 좋아 합니다.	E	97.79%	I	75.82%	I	93.76%	I	70.19%	E	91.92%	E	
	case 4.2	예기치 못한 상황이면 살 짝 당황스러 움..	I	84.09%	I	99.85%	I	99.87%	I	98.83%	I	68.57%	I	

모델별 성격 유형 예측 성능 분석

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

단어와 문장에 대한 성격 유형 예측 - OOV 문제

```
예측값 : 0.12785016
```

```
입력문장: <옴티에서는 그 누구보다도 앞장선다.>  
>> 87.21% 정도로 I 유형입니다.
```

```
OOV 토큰 확인: OOV OOV OOV OOV OOV OOV OOV OOV  
[[1, 1, 1, 1, 1, 1, 1, 1]]
```

oov 처리 불가

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

단어와 문장에 대한 성격 유형 예측 - OOV 문제 (해결!)

BPE

(Byte Pair Encoding)

: 자연어 처리에서의 서브워드 분리 알고리즘

1. 초기 단어 집합 생성

공백을 기준으로 분리(NSMC 사용)

2. 글자(character) 단위로 분리

3. 점차적으로 단어 집합(vocabulary)을 만들고

가장 많이 등장하는 유니그램을 하나의 유니그램으로 통합

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

BPE 서브워드 분리 알고리즘

1) 초기 단어 집합 생성

원본 훈련데이터로부터 각 단어들의 빈도수 카운트

단어	출현 빈도수
low	5
lower	2
newest	6
widest	3

2) 개별 문자단위로 분리

단어	출현 빈도수
l o w	5
l o w e r	2
n e w e s t	6
w i d e s t	3

(초기단어집합)

l o w e r n s t i d

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

BPE 서브워드 분리 알고리즘

3) 가장 빈도수가 높은 유니그램의 쌍을 하나의 유니그램으로 통합 1회)

단어	출현 빈도수
low	5
lower	2
newest	6
widest	3

에서

lo = 7

ow = 7

es = 9

es가 9로 가장 빈도수가 높으므로

es부터 묶어줌

1회 - es = 9

단어	출현 빈도수
low	5
lower	2
newest	6
widest	3

3회 - lo = 7

단어	출현 빈도수
low	5
lower	2
newest	6
widest	3

2회 - est = 9

단어	출현 빈도수
low	5
lower	2
newest	6
widest	3

4회 - low = 7

단어	출현 빈도수
low	5
lower	2
newest	6
widest	3

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

BPE 서브워드 분리 알고리즘

4) 최종

단어	출현 빈도수
low	5
low e r	2
newest	6
widest	3

lowest
lone
won
loss
...

#vocabulary update!

l o w e r n s t i d
es est lo low
ne new newest
wi wid widest

다양한 단어를 만들 수 있음

2. LSTM/Bi-LSTM을 이용한 MBTI 특징 학습

BPE 서브워드 분리 알고리즘

입력문장: <뮤지컬 치킨 바보 엠티 에어팟 아이폰 갤럭시 졸려 배고파>

['_뮤지컬', '_치킨', '_바보', '_엠', '티', '_에어', '팟', '_아이', '폰', '_', '갤', '력', '시', '_졸려', '_배고', '파']

[1153, 23601, 2345, 16977, 1920, 23853, 13054, 1135, 9456, 4, 20683, 12750, 542, 40012, 39858, 2814]

단어들 간의 관계를 학습할 수 있게 됨

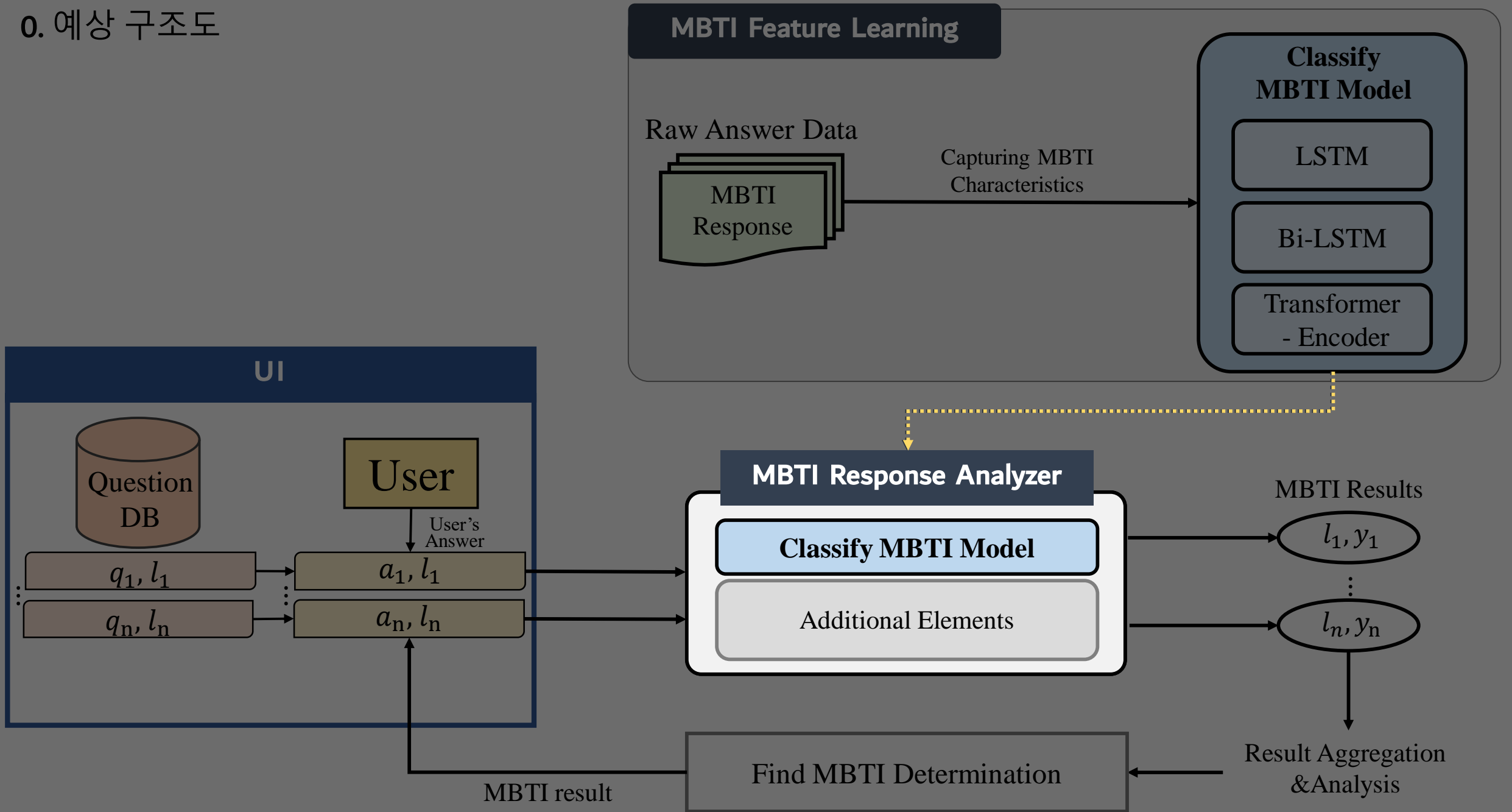
MBTI Feature Learning

~~LSTM~~

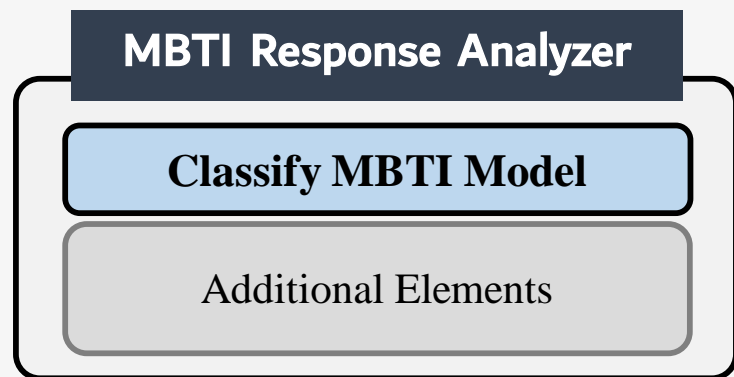
Bi-LSTM

Transformer-Encoder
(겨울방학..)

0. 예상 구조도



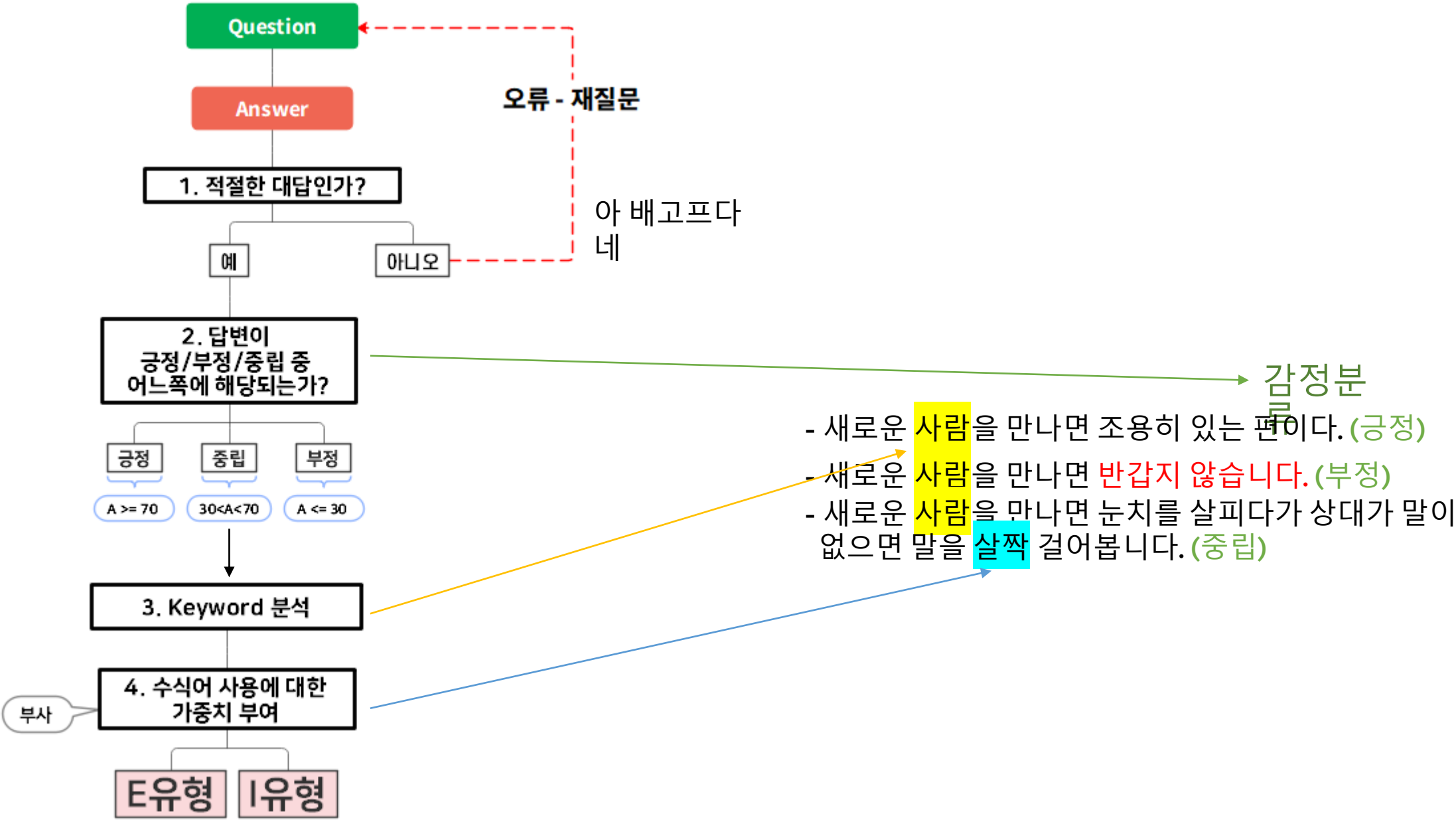
Ⅱ . MBTI Response Analyzer



1. MBTI 응답 분석기 제작 계획

2. ..

II.1. MBTI 응답 분석기 제작 계획



- 긍정/부정 다시한번 고려하라
 - 예측성능 표 case 지우고 조금더 직관적으로 표현하기
 - ~~구조도 다시그리기~~
 - 표안의 영문문장들 더 직관적으로 변경
 - E와 I 유형 E가 60 I가 40 이것도 막대그래프로 표현해서 ppt에 추가하기
 - 데이터 구조도 바꾸기 + augmentation 적용해서 구조도 그리기
 - 문장 이진 분류 논문 찾아보기
-
- 인공지능에서 사용하는 챗봇만들기에 내가만든 모델 넣어서 보여주기 가능한가..???



PPT 수정

E 외향형 (Extraversion) - 자기외부에 주의집중

다른 누군가에게 발상, 지식이나 감정을 표현함으로써 에너지를 얻는다. 사교적, 활동적이며 외부활동에 적극성을 발휘한다. 폭넓은 대인관계를 가지며 글보다는 말로 표현하기를 좋아한다. 경험을 통해 이해한다.

I 내향형 (Introversion) - 자기내부에 주의집중

발상, 지식이나 감정에 대한 자각의 깊이를 늘려감으로써 에너지를 얻는다. 조용하고 신중하며 내면활동에 집중력을 발휘한다. 깊이있는 대인관계를 가지며 말보다는 글로 표현하기를 좋아한다. 이해한 다음 행동한다.

E
외향
사교적
활동적
적극