

# 2024년 지능화 파일럿 프로젝트 프로젝트 계획서

*Project*

**BERT 모델을 활용한 고객 서비스 챗봇 구현**  
Implementation of a Customer Service Chatbot Using BERT Model

사수진(2023254010)

2024. 09. 25.

# Contents



- 연구 배경 및 목표
- 기존 연구의 한계
- 연구 방법
- 데이터셋 구축 및 모델 학습 평가
- 결론 및 향후 연구과제

# 연구 배경 및 목표

## 1. 배경 및 필요성

- 챗봇으로 고객상담 업무를 대체 및 보완하여 업무 부담 감소.
- 시간과 장소에 구애받지 않고 지속적인 고객 문의 응답.
- 대화 형식을 통해 이용자의 의도와 요구를 파악하여 맞춤형 정보 제공.
- 고객 지원 인력 비용 절감과 대량의 고객 요청 효율적 대응 가능.

## 2. 문제 정의

- 고객 서비스팀의 업무 부담 증가로 인한 비효율성.
- 팀원의 경험과 숙련도에 따라 응답의 일관성 결여.
- 대규모 고객 지원 인력 유지에 따른 높은 비용 발생.
- 대량 문의 처리에 시간 소요로 인한 고객 응답 지연.



# 연구 배경 및 목표

## 3. 연구 목표

- 오피스 고객 지원 챗봇 구현
  - ✓ 오피스 제품 관련 고객 지원을 자동화하는 챗봇 연구 개발
  - ✓ 매뉴얼 데이터를 기반으로 한 질의 응답(Q&A) 데이터 구축
  - ✓ BERT 모델을 파인튜닝하여 고객 문의 이해 및 응답 생성

## 4. 연구 기대 효과

- 빠른 문제 해결로 고객 만족도 증가
- 연중무휴로 언제 어디서나 고객 지원 가능
- 일관되고 정확한 정보 제공으로 신뢰성 증대

# 기존 연구의 한계

## 1. 기존 연구의 한계

- 일반 도메인에서 높은 성능
- 의료, 법률 등 특수 도메인에서 어휘 외 단어 문제 및 도메인 특화 용어 이해 부족 자연어 처리 성능 저하
- 다양한 분야에 특화된 사전학습 언어 모델 개발
- 특수 도메인의 요구를 충족시키기 위한 연구 및 활용

## 2. 기존 기술의 현황, 비교

모델	기반 기술	특징	장점	단점
BERT	Transformer, Bidirectional Attention	사전학습된 양방향 인코더 기반 모델, 문맥 이해 능력 강화	양방향 문맥 이해, 다양한 NLP 태스크에 적합	대규모 모델로 인해 계산 자원 소모 큼, 한국어 지원 부족
KoBERT	BERT의 한국어 특화 버전	한국어 데이터로 사전학습된 BERT, 한국어 문맥 이해 강화	한국어 문맥을 잘 이해, BERT의 장점 활용	한국어 데이터에 특화되어 있어 다른 언어에는 비효율적
KcBERT	BERT의 한국어 특화 버전	한국어 데이터로 학습, 한국어 질의응답 태스크에 최적화	한국어의 문맥과 의미를 잘 이해, 특정 도메인에 효과적	특정 도메인에 특화되어 있어서 일반적인 언어 모델에 비해 범용성이 부족
KoGPT2	GPT-2의 한국어 특화 버전	GPT-2 기반의 언어 생성 모델, 단방향 어텐션 사용	자연어 생성 성능 뛰어남, 대화 생성에 유리	단방향 문맥 이해, 긴 문맥 처리에 어려움, 한국어 데이터의 다양성 문제

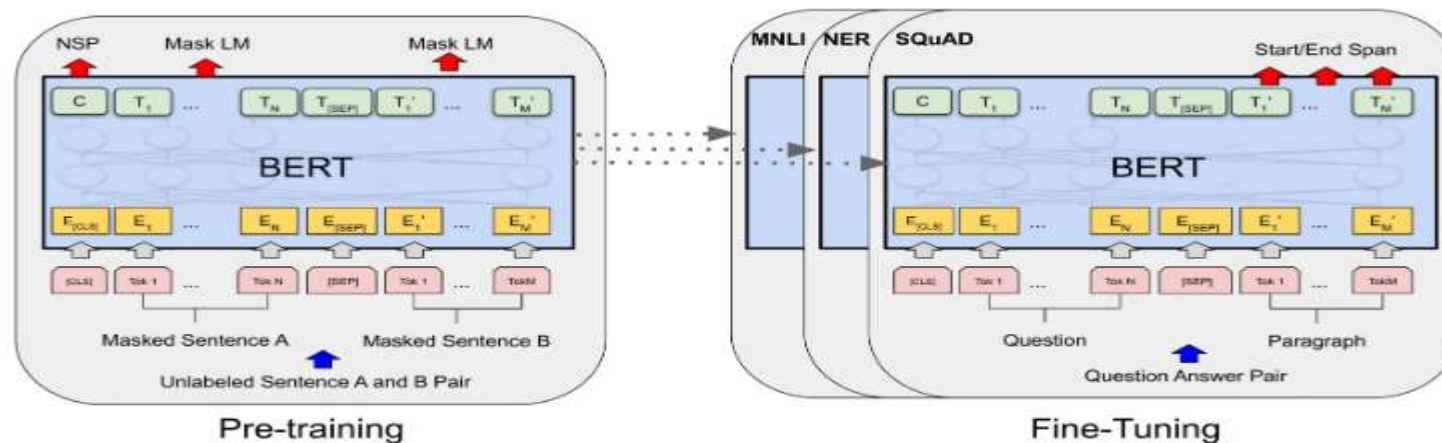
# 연구 방법

## 1. BERT 모델 개요

- 구글에서 2018년에 발표한 언어 모델로, Transformer 아키텍처를 기반으로 양방향 학습으로 뛰어난 성능
- 대규모 텍스트 데이터로 사전 학습되며 특정 자연어 처리 작업에 파인튜닝 가능

## 2. BERT 주요 특징

- 양방향 문맥 이해: BERT는 문장의 앞뒤 문맥을 모두 이해하여 더 정확한 문맥 이해를 가능하게 합니다.
- 사전학습과 미세 조정: BERT는 대량의 텍스트 데이터를 이용해 일반적인 언어 패턴을 학습한 후, 다양한 특정 작업에 맞게 미세 조정할 수 있습니다.
- Transformer 기반: BERT는 Attention 메커니즘을 사용하는 Transformer의 인코더 부분만 사용하여 긴 문맥의 이해를 가능하게 합니다.





# 연구 방법

## 3. BERT 다국어 모델

- 여러 언어를 동시에 지원하여 다양한 언어 처리 작업에서 뛰어난 성능을 발휘
- 글로벌 사용자에게 일관된 고객 지원을 제공하기 위해 다국어 이해 및 응답 생성에 최적화

## 4. KoBERT 모델

- BERT의 한국어 성능 한계를 극복하기 위해 SKT Brain에서 개발한 모델
- KoBERT는 위키피디아나 뉴스 등에서 수집한 수백만 개 한국어 문장의 대규모 말뭉치 (Corpus)를 학습

## 5. Ko-Sentence-BERT-SKTBERT 모델

- 한국어 문장 임베딩에 최적화된 Sentence-BERT 기반 모델
- 문장 유사도 측정 및 분류 작업에서 우수한 성능 발휘

# 연구 방법

## 6. 모델의 성능 평가

- Exact Match(EM)
  - ✓ 예측값이 정답과 완전히 일치하는 비율을 측정하는 지표
  - ✓ 질문-답변 시스템에서 예측된 답변이 실제 답변과 완전히 일치하는지 평가할 때 사용

$$EM = 100 * \frac{\text{예측 정답 수}}{\text{실제 정답 수}}$$

수식 1 Exact Match

- F1 스코어
  - ✓ 실제 정답과 예측이 일부만 겹쳐도 점수를 부여하는 지표
  - ✓ Precision은 정확도를 의미하며 Recall은 재현율을 의미

$$F1 = 100 * \frac{2 * Precision * Recall}{Precision + Recall}$$

수식 2 F1 score



# 데이터셋 구축 및 모델 학습 평가

## 1. 데이터셋 구축 방법

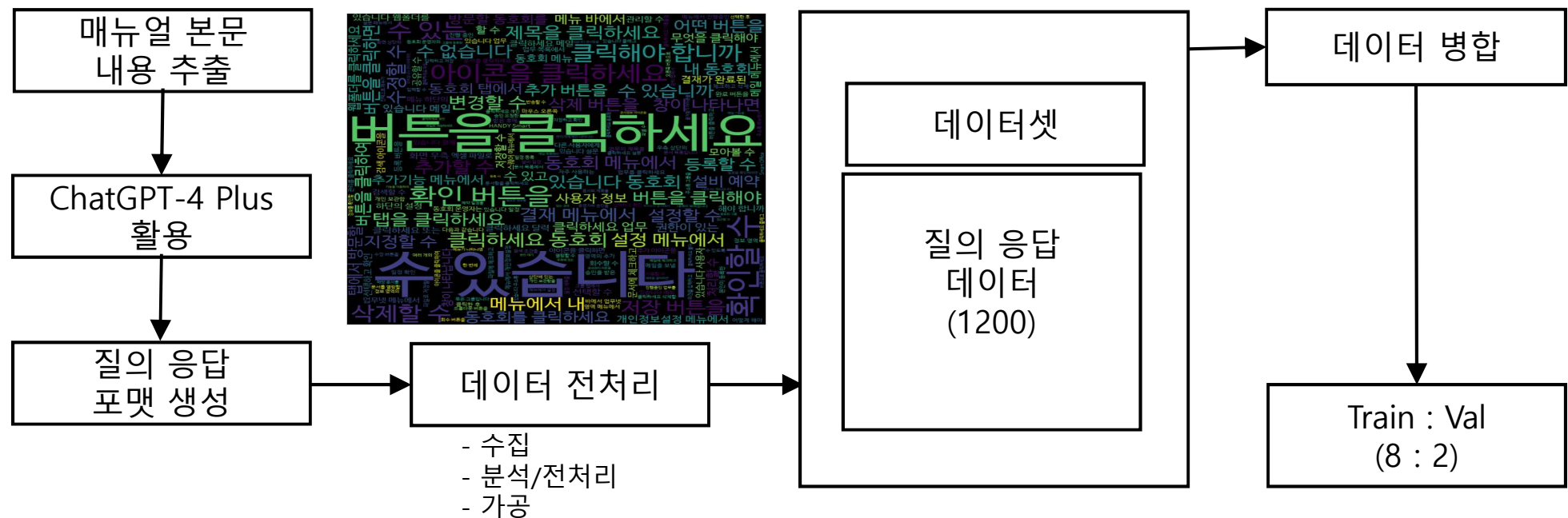
- 오피스 제품 매뉴얼 데이터를 기반으로 한 질의 응답(Q&A) 데이터셋 JSON 파일 형식으로 생성  
데이터셋 포맷 예시(KorQuad 한국어 질의 응답 데이터셋 참고)

```
{  
  "context": "문서나 글의 일부, 질문의 배경이 되는 내용.",  
  "question": "질문 내용",  
  "id": "고유한 id값",  
  "is_impossible": "문맥 내에서 질문에 대한 답변이 존재 유무(true/false)",  
  "answers": {  
    "text": ["답변 내용"],  
    "answer_start": [문서 내에서 답변이 시작되는 위치 (index)]  
  }  
}
```

# 데이터셋 구축 및 모델 학습 평가

## 2. 데이터 수집

- ChatGPT-4 Plus를 활용하여 매뉴얼 문서의 본문을 단락 단위로 분할한 후, 데이터셋 형식으로 **1,200개** 생성



# 데이터셋 구축 및 모델 학습 평가

## 3. 데이터 분석 및 전처리

- 질의 응답 데이터셋 id값을 고유한 값으로 일괄 보정 작업
- 질의 응답 데이터셋 문서내에서 답변이 존재하지 않는 경우 보정 작업
- 질의 응답 데이터셋 문서내에서 답변이 시작되는 위치(**answer\_start**) 데이터 검증 및 보정 작업
  - ✓ ChatGPT-4 Plus를 사용하여 데이터셋을 구축하였으나, 응답 시작 지점의 정확성이 부족

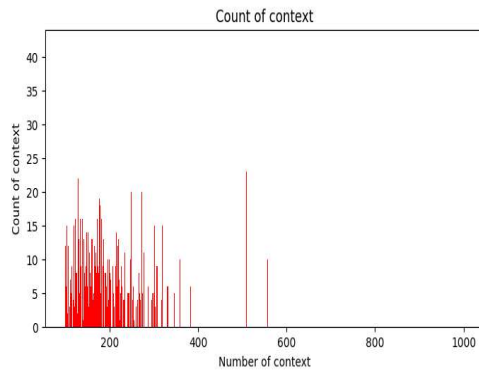
	pharagraph	question	answer	answer_start	is_impossible
0	감사안함 감사를 받지 않습니다 감사안함을 선택하면 추가되었던 감사 결재경로가 모두 ...	감사안함을 선택하면 어떤 일이 발생합니까	추가되었던 감사 결재경로가 모두 삭제됩니다	28	0
1	감사안함 감사를 받지 않습니다 감사안함을 선택하면 추가되었던 감사 결재경로가 모두 ...	사전감사란 무엇입니까	최종결재자가 결재하기 전에 시행하는 감사	57	0
2	감사안함 감사를 받지 않습니다 감사안함을 선택하면 추가되었던 감사 결재경로가 모두 ...	사전감사를 선택하면 감사 결재경로는 어떻게 변경됩니까	감사 결재경로가 최종결재자 앞으로 이동합니다	94	0
3	감사안함 감사를 받지 않습니다 감사안함을 선택하면 추가되었던 감사 결재경로가 모두 ...	사후감사란 무엇입니까	최종결재자가 결제한 후에 시행하는 감사	124	0
4	감사안함 감사를 받지 않습니다 감사안함을 선택하면 추가되었던 감사 결재경로가 모두 ...	사후감사를 선택하면 감사 결재경로는 어떻게 변경됩니까	감사 결재경로가 최종결재자 이후로 이동합니다	160	0
...	...	...	...	...	...
1195	웹폴더 설정하기 목록 보기 유형을 설정합니다 미리보기 파일을 섹 네일로 표시합니다 이...	메일로 대용량파일을 첨부하여 전송했을 때 자동으로 저장될 폴 더를 어떻게 지정합니까	해당 대용량파일이 자동으로 저장될 폴더를 지정합니다	305	0
1196	웹폴더 설정하기 목록 보기 유형을 설정합니다 미리보기 파일을 섹 네일로 표시합니다 이...	폴더에 등록된 파일의 다운로드 가능 기간을 어떻게 설정합니까	폴더에 등록된 파일의 다운로드 가능 기간을 설정합니다	334	0
1197	웹폴더 설정하기 목록 보기 유형을 설정합니다 미리보기 파일을 섹 네일로 표시합니다 이...	파일이 자동으로 삭제되는 기간을 설정하면 어떤 파일들의 자동 삭제 조건을 설정할 수...	일반 첨부파일 대용량 첨부파일 휴지통 파일의 자동 삭제 조건을 각각 설정할 수 있습니다	418	0
1198	웹폴더는 언제 어디서든 파일을 다운로드하고 업로드할 수 있는 기 능입니다 조직의 전산...	웹폴더는 어떤 기능을 제공합니까	언제 어디서든 파일을 다운로드하고 업로드할 수 있는 기능입니 다	5	0
1199	웹폴더는 언제 어디서든 파일을 다운로드하고 업로드할 수 있는 기 능입니다 조직의 전산...	웹폴더를 사용하면 조직 구성원 간에 어떤 이점이 있습니까	조직 구성원 간에 쉽게 파일을 공유할 수 있습니다	59	0

1200 rows × 5 columns

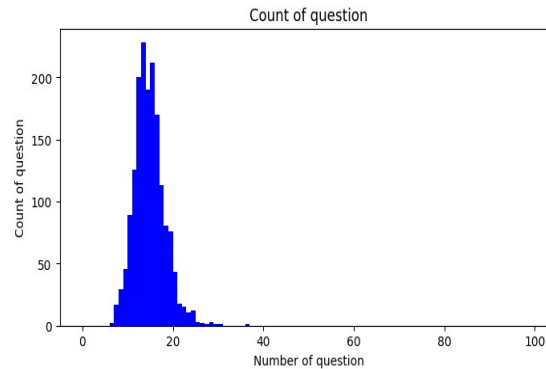
# 데이터셋 구축 및 모델 학습 평가

## 3. 데이터 분석 및 전처리

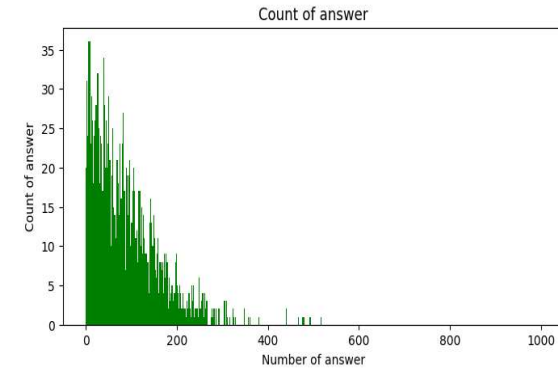
- 지문(Context) 길이 분석



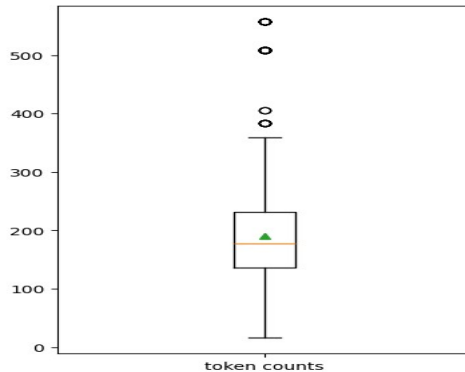
- 질문(Question) 길이 분석



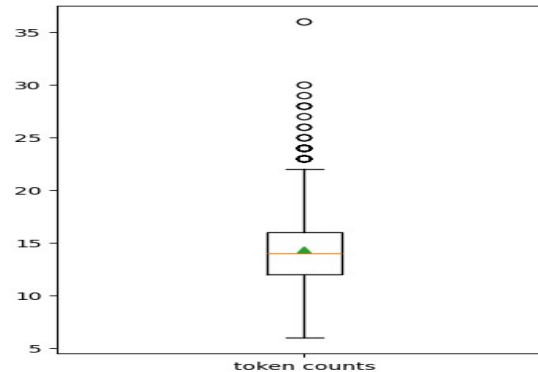
- 답변(Answer) 길이 분석



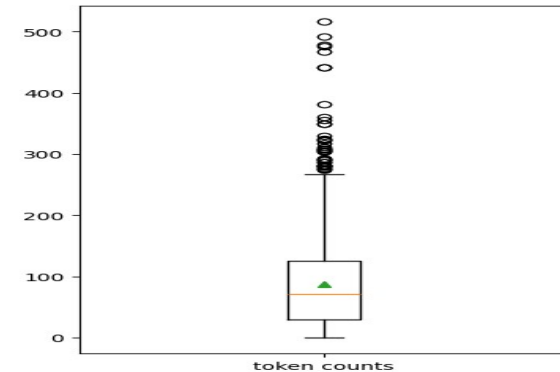
- 지문(Context) 토큰 길이



- 질문(Question) 토큰 길이



- 정답(Answer) 토큰 길이



# 데이터셋 구축 및 모델 학습 평가

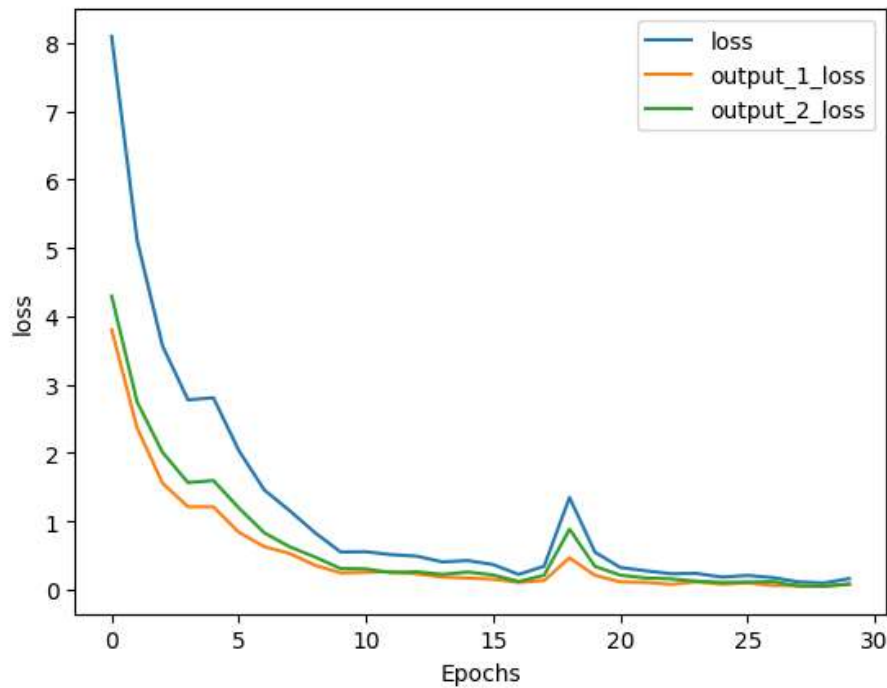
## 4. 환경 구성

- 컴퓨터 사양 (H/W)
  - ✓ CPU : Apple M3 Pro칩(11코어 CPU)
  - ✓ RAM : 36GB
  - ✓ GPU : 14코어 GPU, 16코어 Neural Engine
- 운영체제 및 소프트웨어
  - ✓ macOS Sequoia 15.0
  - ✓ Python
  - ✓ Visual Studio Code
- 실험방법
  - ✓ 데이터셋 8:2 로 분류
  - ✓ BERT 다국어 모델, KOBERT 모델 사용
  - ✓ 하이퍼파라미터 Adam, Epoch 30, Batch\_size 16, 학습률 0.01

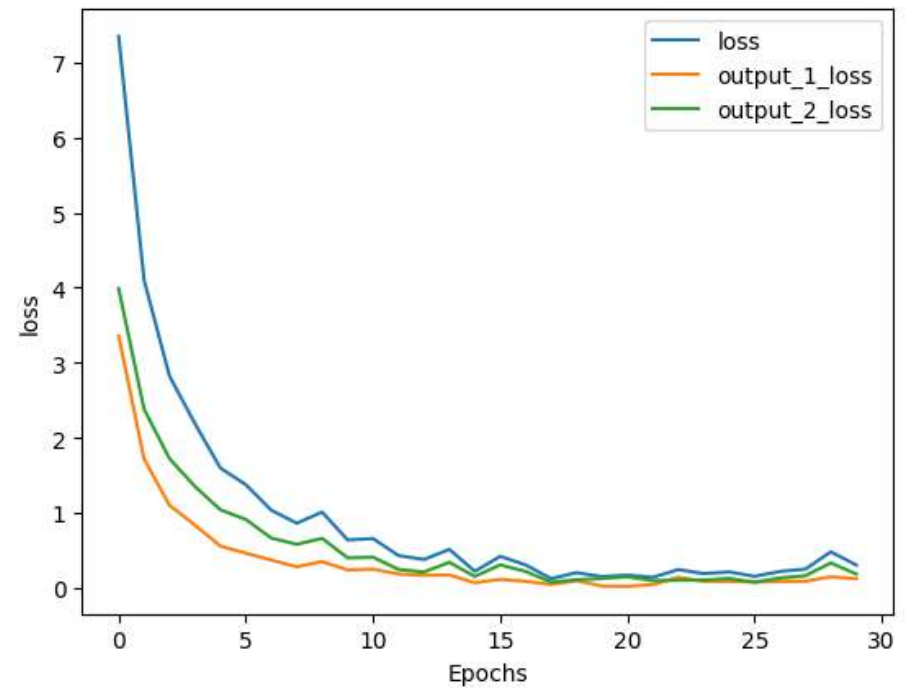
# 데이터셋 구축 및 모델 학습 평가

## 5. 모델별 학습 곡선

- BERT 다국어 모델



- KoBERT 모델





# 데이터셋 구축 및 모델 학습 평가

## 6. 모델별 평가 결과 비교

### • BERT 다국어 모델

```

=====
87% [██████████] | 26/30 [03:38<00:32, 8.16s/it]=====
model 26
EM : 40.112994350282484
F1 : 68.87817973622325

=====
Some weights of BertForQuestionAnswering were not initialized from the model checkpoint at bert-base-multilingual-cased ;
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
90% [██████████] | 27/30 [03:46<00:24, 8.11s/it]=====
model 27
EM : 36.72316384180791
F1 : 64.96005136260104

=====
Some weights of BertForQuestionAnswering were not initialized from the model checkpoint at bert-base-multilingual-cased ;
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
93% [██████████] | 28/30 [03:55<00:16, 8.26s/it]=====
model 28
EM : 34.463276836158194
F1 : 63.76207514889406

=====
Some weights of BertForQuestionAnswering were not initialized from the model checkpoint at bert-base-multilingual-cased ;
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
97% [██████████] | 29/30 [04:03<00:08, 8.38s/it]=====
model 29
EM : 38.983050847457626
F1 : 65.12959627495898

=====
Some weights of BertForQuestionAnswering were not initialized from the model checkpoint at bert-base-multilingual-cased ;
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
100% [██████████] | 30/30 [04:13<00:00, 8.43s/it]=====
model 30
EM : 38.983050847457626
F1 : 67.0186213717846

```

### • KoBERT 모델

```

=====
Evaluating Models: 92% [██████████] | 46/50 [04:13<00:21, 5.39s/it]=====
Model 26
EM : 48.5
F1 : 64.5400196075614

=====
Evaluating Models: 94% [██████████] | 47/50 [04:18<00:16, 5.39s/it]=====
Model 27
EM : 47.5
F1 : 61.632530913551385

=====
Evaluating Models: 96% [██████████] | 48/50 [04:24<00:10, 5.43s/it]=====
Model 28
EM : 49.5
F1 : 65.37780964655032

=====
Evaluating Models: 98% [██████████] | 49/50 [04:29<00:05, 5.42s/it]=====
Model 29
EM : 50.0
F1 : 67.0351272581068

=====
Evaluating Models: 100% [██████████] | 50/50 [04:34<00:00, 5.50s/it]=====
Model 30
EM : 44.5
F1 : 61.8386852693726
=====

```

# 데이터셋 구축 및 모델 학습 평가

## 7. 모델별 추론 결과 비교

### • BERT 다국어 모델

지문 : 선택 결재하기 두 개 이상의 문서를 선택하여 결재할 경우, 선택한 문서를 순서대로 창을 띄워 원하는 문서를 간편하게 결재할 수 있습니다. 1. 결재 메뉴에서 결재할 문서에서 결재할 문서를 선택하여 체크하세요. 2. 체크  
질문 : 참조문서는 처리구분에 어떻게 [UNK]?  
정답 : [UNK] 참조 [UNK] 로 표시  
예측 : [UNK] 참조 [UNK] 로 표시되고 [ 전자결재 ] 창에 참조확인 버튼이 나타납니다.

지문 : 메일 일반설정하기 일반 설정 메뉴에서는 메일 수신자의 별칭, 메일의 보관 기간, 메일 읽기 화면, 보낸 편지 저장 여부, 답장과 관련된 설정, 본인제외 설정 등 메일과 관련된 설정을 할 수 있습니다. 1. 메일의 설정  
질문 : 메일의 보관 기간이 경과된 메일은 어떻게 [UNK]?  
정답 : 자동으로 [UNK].  
예측 : 자동으로 [UNK].

지문 : 협조 기안문서의 안전이 업무와 관련이 있을 경우 해당 안전을 승인하는 행위입니다. 협조자가 두 명 이상일 경우, 순차적으로 결재 승인이 이루어지며 협조자가 문서를 반려할 경우 기안자에게 문서가 [UNK]. 병렬협조 자  
질문 : 협조란 무엇을 의미합니까?  
정답 : 기안문서의 안전이 업무와 관련이 있을 경우 해당 안전을 승인하는 행위  
예측 : 기안문서의 안전이 업무와 관련이 있을 경우 해당 안전을 승인하는 행위

지문 : 버튼설명 달기 : 창을 닫습니다. 결재정보 : 문서의 기록물철 설정 및 결재선을 지정합니다. 공람지정 : 결재선에 포함되어 있지 않지만, 문서가 접수되면 해당 문서를 열람할 사용자를 지정합니다. 지정 : 문서를 처리할  
질문 : 일괄접수는 무엇입니까?  
정답 : 여러 개의 문서를 선택하여 열람한 후, 설정 사항을 모든 접수 문서에 동일하게 적용하여 접수합니다.  
예측 : 여러 개의 문서를 선택하여 열람한 후, 설정 사항을 모든 접수 문서에 동일하게 적용하여 접수합니다.

지문 : 일정 간편 등록하기 일정 등록에 필요한 필수 정보만 입력하여 간편하게 일정을 등록할 수 있습니다. 1. 주간 보기 또는 월간 보기인 경우, 달력에서 예약할 날짜를 클릭하세요. 일간 보기인 경우, 예약할 시간을 클릭해서  
질문 : 달력을 다른 사용자와 공유하면 공유한 달력은 어디에 [UNK]?  
정답 : 공유자의 공유받은 달력에 [UNK].  
예측 : 공유자의 공유받은 달력에 [UNK].

지문 : 보류 문서의 처리를 미룹니다. 회송 부재 대리 결재문서일 경우 활성화되는 버튼입니다. 안전이 중요하여 원결재자가 확인해야 할 경우, 원결재자에게 문서를 전달합니다. 이전문서 여러 개의 문서를 선택하여 열람한 경우,  
질문 : 문서가 수정된 경우 변경 내역을 확인하려면 어떤 기능을 사용합니까?  
정답 : 이력관리  
예측 : 이력관리

# 데이터셋 구축 및 모델 학습 평가

## 7. 모델별 추론 결과 비교

### • KoBERT 모델

지문 : 결재 처리 방법 결재 기안문서의 안전에 대한 의사 결정 권한이 있는 결재자가 그 의사를 결정하는 행위입니다 검토 중간결재자의 결재 행위입니다 전결 조직의 장으로부터 그 권한에 속하는 사무의 일부를 위임받은 자가 일정한  
질문 : 검토는 어떤 행위입니까  
정답 : 중간결재자의 결재 행위  
예측 : 중간결재자의 결재 행위

지문 : 키워드 메일 모아보기 키워드 메일은 자주 찾는 검색 조건을 저장하여 필요한 메일을 빠르게 찾을 수 있는 기능입니다 키워드 메일은 20개까지 설정할 수 있습니다 메일 메뉴에서 키워드 메일에서 확인할 키워드를 클릭하세요  
질문 : 키워드 메일을 설정하려면 어디에서 추가 아이콘을 클릭해야 합니까  
정답 : 키워드 메일 메뉴 우측의 추가 아이콘  
예측 : 키워드 메일 메뉴 우측의 추가 아이콘

지문 : 메일 예약하기 메일을 즉시 보내지 않고 원하는 시간에 발송되도록 설정할 수 있습니다 예약한 메일은 예약 편지함에 보관됩니다 1 메일 메뉴에서 메일쓰기 버튼을 클릭하세요 2 메일 내용을 작성하세요 3 메일쓰기 창 하단의  
질문 : 특정일에 메일을 반복하여 보내려면 어떤 버튼을 클릭해야 합니까  
정답 : 반복예약 버튼  
예측 : 반복예약 버튼

지문 : 투표 참여기간 변경하기 진행 중인 투표의 참여 기간을 변경할 수 있습니다 1 설문 메뉴에서 진행중인 투표에서 참여기간을 변경할 투표 제목을 클릭하세요 2 투표 상세 화면에서 참여기간변경 버튼을 클릭하세요 3 종료일을  
질문 : 투표 종료는 누가 할 수 있습니까  
정답 : 투표 작성자 또는 투표 삭제 권한이 있는 사용자  
예측 : 투표 작성자 또는 투표 삭제 권한이 있는 사용자

지문 : 게시물 삭제하기 본인이 등록한 게시물에 한해 게시물을 삭제할 수 있습니다 게시물 목록에서 삭제할 게시물에 체크하고 삭제 버튼을 클릭하세요 게시물 본문 상단에 있는 삭제 버튼을 클릭하세요 게시물 임시 저장하기 게시물  
질문 : 게시물을 삭제하려면 게시물 목록에서 무엇을 클릭해야 합니까  
정답 : 삭제 버튼  
예측 : 삭제할 게시물에 체크하고 삭제 버튼

지문 : 발송대기 문서 열람 시 기안서식에 연결된 시행문 서식이 있으면 해당 시행문 서식으로 자동 변환된 문서가 열리고 기안서식에 연결된 시행문 서식이 없으면 변환할 시행문 서식을 선택하는 창이 나타납니다 1 결재 메뉴에서  
질문 : 기안서식에 연결된 시행문 서식이 없으면 어떤 창이 나타납니까  
정답 : 변환할 시행문 서식을 선택하는 창  
예측 : 변환할 시행문 서식을 선택하는 창

# 데이터셋 구축 및 모델 학습 평가

## 7. 모델별 추론 결과 비교

### • KoBERT 모델

지문 : 참조업무는 업무 내용과 진행상황을 확인만 할 수 있습니다 참조업무는 상태 아이콘에 참조 아이콘으로 표시됩니다 1 메뉴 바에서 업무넷 메뉴에서 처리할 업무를 클릭하세요 2 처리할 업무의 제목을 클릭하세요 3 업무의 내;  
질문 : 참조업무는 무엇을 할 수 있습니까  
정답 : 업무 내용과 진행상황을 확인만 할 수 있습니다  
예측 : 업무 내용과 진행상황을 확인만 할 수 있습니다

지문 : 개인 보관함 관리하기 개인 보관함은 여러 결재함이나 문서함에 등록된 문서를 저장할 수 있는 문서함입니다 사용자의 업무 및 편의에 맞게 개인 보관함을 추가하여 필요한 문서를 하나의 문서함에서 모아볼 수 있습니다 개인  
질문 : 개인 보관함은 무엇을 저장할 수 있습니까  
정답 : 여러 결재함이나 문서함에 등록된 문서를 저장할 수 있는 문서함입니다  
예측 : 여러 결재함이나 문서함에 등록된 문서를 저장할 수 있는 문서함

지문 : 토론 등록하기 토론 주제를 등록하여 특정 사건에 대한 다른 사람들의 의견을 들을 수 있습니다 등록된 토론 주제는 즉시 공개되지 않고 동호회 관리자의 승인을 받은 후에 공개됩니다 1동호회 메뉴에서 동호회 홈 화면 상단  
질문 : 등록된 토론 주제가 공개되기 전에 어떤 절차가 필요합니까  
정답 : 동호회 관리자의 승인을 받은 후에 공개됩니다  
예측 : 동호회 관리자의 승인을 받은 후에 공개됩니다

지문 : 수신자 지정하기 수신자 탭에서는 문서를 수신할 부서 또는 업무담당자를 선택합니다 수신함을 보유한 부서만 수신부서로 선택할 수 있습니다 업무담당자는 수신함이 없는 부서의 사용자도 설정할 수 있습니다 1 결재정보 창에  
질문 : 업무담당자는 어떤 부서의 사용자도 설정할 수 있습니까  
정답 : 수신함이 없는 부서의 사용자도 설정할 수 있습니다  
예측 : 수신함이 없는 부서의 사용자

지문 : 메일 예약하기 메일을 즉시 보내지 않고 원하는 시간에 발송되도록 설정할 수 있습니다 예약한 메일은 예약 편지함에 보관됩니다 1 메일 메뉴에서 메일쓰기 버튼을 클릭하세요 2 메일 내용을 작성하세요 3 메일쓰기 창 하단  
질문 : 메일을 원하는 시간에 발송되도록 설정하려면 무엇을 체크해야 합니까  
정답 : 예약발송 영역  
예측 : 예약발송 영역

지문 : 문서 정보 확인하기 문서의 기본정보 결재경로 열람기록 댓글 등을 확인할 수 있습니다 분할보기 상태일 경우 문서 제목을 클릭하세요 화면 우측에 문서 정보가 나타납니다 목록보기 상태일 경우 제목 우측에 있는 문서정보 0  
질문 : 결재경로 탭에서 확인할 수 있는 정보는 무엇입니까  
정답 : 문서의 결재자와 처리시간 검토 의견 등을 확인할 수 있습니다  
예측 : 문서의 결재자와 처리시간 검토 의견 등을 확인할 수 있습니다 의견 아이콘

# 결론 및 향후 연구 과제

## 1. 결론

- BERT 다국어 모델로 질의응답 예측 답변시 [UNK](unknown token) 로 빈발 발생으로 인해 예측 답변의 정확성이 저하되는 문제로 한국어 질의 응답에 적합하지 않음
- KoBERT를 통해 한국어의 문법과 어휘를 효과적으로 반영하여 질의 응답 품질을 향상시켰으며, 한국어 자연어 처리에서 언어 특화 모델의 중요성을 강조하고 KoBERT 활용을 권장될 수 있음을 확인

## 2. 향후 연구 과제

- KoBERT를 통해 한국어 질의응답 정확도를 향상시키는 연구가 필요
- Ko-Sentence-BERT-SKTBERT 모델을 활용하여 지문을 학습시켜서 질문만 입력하여 답변을 받을 수 있도록 연구 필요



# 2024년 지능화 파일럿 프로젝트 프로젝트 계획서

## 감사합니다