

※ 본 양식은 출품접수기간(2023. 06. 12. ~ 07. 09.)에 제출하여 주시기 바랍니다.

사업계획서

※ 최종 제출 시, 파란색으로 표시된 가이드는 삭제 후 제출하시기 바랍니다.

※ 본 사업계획서는 최대 5장 내외로 작성하시기 바랍니다.

※ 도표, 이미지, 영상 등을 자유롭게 활용하여 작성하시기 바랍니다.

1. 아이디어 개요

1-가. 아이디어 명

Soldier_GPT

1-나. 적용 기술 분야

☒ 웹 서비스 ☒ 모바일 앱 ☐ 사물인터넷 ☐ (기타 :)

1-다. 아이디어 요약

국방 공공 데이터와 국가 중점 데이터를 활용하여 군 관련 인원에게 맞춤형 정보 서비스를 제공하려는 아이디어입니다. 핵심은 OpenAI에서 개발한 인공지능 모델인 Chat-GPT를 통해 대용량의 데이터를 처리하고 사용자 친화적인 인터페이스(User-friendly Interface)로 원하는 정보를 제공하는 것이 목표이며, 군 관련 인원들의 복지 향상에 기여하고자 합니다.

* 사업계획서에서 사용하는 '군 관련 인원' 은 '군인 및 군 가족'을 뜻함

2. 아이디어의 적절성

2-1. 활용 공공데이터

1. 저희 팀이 기획한 아이디어의 활용 데이터는 다음과 같습니다.

가. 특정 데이터에 1~2개에 국한 받지 않고, 가지고 있는 모든 데이터를 활용합니다.

나. 데이터 전처리 작업과 fine-tuning을 거침으로써, 데이터를 활용한 원하는 서비스를 얻습니다.

* OpenAI에서 제공하는 model은 대표적으로 'gpt-4'와 'gpt-3.5-turbo'가 있습니다.

2. 활용 가능한 예시입니다.

☞ 국방부_국군복지단 복지시설 상세현황(국방공공 데이터) + 군 복지시설 주변 관광지 정보(국가중점 데이터)

가. 위의 데이터를 활용하여 {질문 : 답변} 형식의 데이터로 전처리(preprocessing)한다면 본인이 살고 있는, 가고 싶은 지역의 복지시설 및 주변 관광지에 대해서 손쉽게 정보를 얻을 수 있습니다.

나. 질문: "강원도 00시 반경 3KM내에 어떤 군 복지시설이 있는지 알려주세요."

답변: "00 군병원, 00 복지마트, 00 군인회관이 있습니다."

위와 같은 질문과 답변 형식은 엑셀 데이터와 질문 패턴을 정해둔다면 통계적 기법(프로그램)을 통해 데이터 및 모델을 활용할 수 있습니다.

※ 본 양식은 출품접수기간(2023. 06. 12. ~ 07. 09.)에 제출하여 주시기 바랍니다.

2-2. 아이디어 기획 배경 및 활용 적절성

<아이디어 기획 배경>

1) AI 기술의 발전과 활용성

- 최근 몇 년 동안 AI 기술, 특히 자연어 처리(NLP) 기술의 발전이 두드러졌습니다. 그 중에서도 OpenAI의 CHAT-GPT 모델은 그 성능과 활용성이 탁월하여 다양한 분야에서 활용되고 있습니다. 이러한 배경 속에서, 최신 AI 기술을 활용하여 군 관련 인원에게 필요한 정보를 채팅 기반의 형태로 제공하는 새로운 서비스를 기획하게 되었습니다.

2) 데이터 활용의 중요성 인식

- 디지털화와 정보화 사회의 흐름 속에서, 공공 데이터를 효과적으로 활용하는 것이 중요하다는 인식이 강화되었습니다. 이에 따라, 국방 공공 데이터를 통해 군 관련 인원에게 정보를 제공하는 방법에 대해 고민하게 되었고, 이 과정에서 AI 기반의 서비스인 Chat-GPT를 활용하는 아이디어를 생각하게 되었습니다.

* 채팅형 인공지능봇이 현실에도 존재하지만, 군 관련 정보를 얻기에는 매우 제한적임
(이유 : 군 관련 데이터, 즉 국방 공공 데이터 / 국가 중점 데이터의 활용도가 매우 낮음)

<활용 적절성>

1) 활용 빈도 및 비중

- Chat-GPT와 같은 AI 챗봇은 실시간으로 사용자의 질문에 응답하므로, 필요한 정보에 빠르게 접근할 수 있다는 장점이 있습니다. 국방 공공 데이터와 국가 중점 데이터가 국민들에게 개방되어 있지만 현실적으로 이에 대한 활용 빈도는 낮은 편입니다. 하지만, 이러한 국방 공공 데이터와 국가 중점 데이터를 처리하여 '검색과 채팅 기반의 서비스'를 제공한다면 사용자의 정보 접근성을 크게 향상시키며, 이에 따라 데이터 활용 빈도와 비중을 늘릴 수 있다고 생각합니다.

2) 가공 가능성

- OpenAI에서 제공하는 Chat-GPT모델은 Fine-Tuning이라는 과정을 통하여 새로운 정보를 주거나 원하는 형식의 답변이 나오도록 학습시킬 수 있습니다. 국방 공공 데이터 및 국가 중점 데이터는 주로 정형 데이터로 구성되어 있어, 자연어 처리 기술을 활용한 가공이 가능합니다. 데이터의 형태에 따라 적절한 전처리 과정을 거쳐 Fine-Tuning을 할 수 있으며, 이를 통해 군 관련 인원이 필요로 하는 정보를 제공하는 서비스를 구축할 수 있습니다. 채팅을 통하여 추가적인 요구사항과 행동 패턴을 학습하고 이해할 수 있기에 맞춤형 정보를 제공할 수 있습니다.

3) 적절성

- 이러한 서비스 제공은 군 관련 인원이 직면하는 정보 접근성 문제를 해결하는데 큰 도움이 될 것입니다. 특히, AI 기반의 서비스는 사용자 친화적인 인터페이스를 제공하며, 실시간 정보 제공이 가능하므로, 이를 활용한 아이디어는 매우 적절하다고 판단됩니다.

※ 본 양식은 출품접수기간(2023. 06. 12. ~ 07. 09.)에 제출하여 주시기 바랍니다.

3. 아이디어의 독창성

1) 폭넓은 데이터 활용

- 소수의 데이터셋을 활용하여 서비스 개발 및 아이디어를 기획하는 것이 그동안의 '데이터 활용 경진대회'에서의 흐름이었다면, 저희가 제시한 Chat GPT를 활용한 아이디어는 방대한 자료의 국방 공공 데이터 및 국가 중점 데이터를 {질문 : 답변} 형식의 데이터로 전처리 작업만 거친다면 위 데이터의 전반적인 내용을 종합적으로 활용하여 원하는 정보를 신속하고 정확하게 얻을 수 있다는 장점과 다른 아이디어와의 차별성이 있습니다.

2) 데이터 처리의 유연함 (flexibility)

- 데이터의 유형이나 내용에 국한 받지 않습니다. 즉, 저희가 제시한 아이디어는 데이터가 다양해짐에 따라 쉽게 확장이 가능합니다. 이런 유연성은 서비스가 성장하고 발전할 수 있는 능력을 가지게 합니다.
- * 국가에서 제공하는 객관적인 정보를 바탕으로 전처리 작업을 진행하기 때문에 군 관련 복지, 교통, 교육 등의 전반적인 내용에 대한 신뢰성이 보장됨

3) 맞춤형 정보 제공 및 사용자 중심의 서비스 제공

- 현실적으로 국방 공공 데이터 및 국가 중점 데이터를 활용하는 것이 일반적인 사람들에게는 제한적이라고 생각되는데, 이런 틀을 벗어나기 위해 '검색과 채팅 기반의 데이터 획득'이라는 아이디어를 제시하여 친근한 방식으로 정보를 제공할 수 있습니다.
- 군 관련 인원, 사용자 중심의 대화형 인터페이스를 활용하여, 사용자가 직접 질문을 통해 원하는 정보를 얻을 수 있게 합니다. 사용자의 질문을 잘 이해하고 만족시킬 수 있으며, 사용자 친화적인 서비스를 제공하는 데 있어 중요한 요소입니다.

4. 아이디어의 구체성

1) 목표

- 방대한 국방 공공 데이터와 국가 중점 데이터를 효율적으로 활용하여 군 관련 인원들에게 쉽고 빠르게 원하는 정보를 제공하는 것입니다.

2) 기본 기능

- 질의응답(Q&A) : 사용자는 자유롭게 국방 관련 질문을 할 수 있습니다. AI는 질문을 분석하고 가장 적절한 응답을 데이터베이스에서 검색하여 제공합니다.
- 맞춤형 정보제공 : 사용자의 질문에 따라 맞춤형 정보를 제공합니다. 이러한 접근 방식은 사용자 만족도를 향상시키고 정보 접근성을 높이는 데 큰 장점이 됩니다.
- 데이터 검색 : 사용자는 특정 데이터를 직접 검색할 수 있습니다. 이는 국방 데이터의 접근 장벽을 낮춰주며 무엇보다 원하는 정보에 대한 빠른 검색이 가능하도록 돕습니다.
- 데이터 요약 : 복잡하고 방대한 데이터를 사용자가 이해하기 쉬운 형태로 요약해 주는데, 이는 데이터의 가공과 이해를 돕습니다.

3) 세부 기능(모델의 활용 측면)

- 자연어 처리 : AI는 사용자의 질문을 정확히 이해하고 적절한 응답을 제공하기 위해 자연어 처리 기술을 활용합니다. 이를 통해 사용자의 의도를 더 잘 이해하고, 그에 따른 정확한 응답을 제공할 수 있습니다.

※ 본 양식은 출품접수기간(2023. 06. 12.. ~ 07. 09.)에 제출하여 주시기 바랍니다.

- 인공지능 학습 : AI는 사용자의 질문과 피드백을 통해 지속적으로 학습하고 개선합니다. 이를 통해 시간이 지남에 따라 서비스의 품질이 계속해서 향상됩니다. 데이터는 계속 업데이트 되기에 추가된 데이터에 맞게 전처리, 학습만 해준다면 정보의 양과 질을 향상시킬 수 있습니다.

5. 아이디어의 가능성

5-1. 아이디어의 실현가능성

1) 데이터 수집과 전처리

- Chat-GPT 기반 AI 모델을 학습하기 위해서는 {질문 : 답변} 형식의 데이터가 필요합니다. 따라서 제안하는 서비스를 구현하려면 먼저 국방 관련 공공 데이터와 국가 중점 데이터를 수집하고 이를 전처리하여야 합니다. 데이터는 이미 공개되어 있으므로 누구나 접근할 수 있습니다. 데이터 전처리의 경우도 필요한 데이터 형식이 비교적 간단하므로 가능합니다.

* OpenAI에서 제공하는 서비스만 이용하더라도 원하는 형식의 정보를 가공, 획득할 수 있음

2) AI 모델 개발

- 사용자의 질문을 이해하고 적절한 응답을 생성하는 AI 모델이 필요합니다. 이 모델은 자연어 처리(NLP) 기술을 기반으로 해야 하며, 이는 AI가 사용자의 질문을 이해하고 적절한 응답을 찾는 데 중요한 역할을 합니다. GPT와 같은 미리 학습된 AI 모델을 사용하면 이 과정을 간소화할 수 있습니다.

* 한국에서는 네이버에서 한국어 기반의 GPT인 Clova라는 모델 및 서비스를 '23. 8월 경 오픈 예정

3) 예시

- 실제 OpenAI에서 제공하는 'gpt-3.5 turbo' 모델을 이용하여 {질문 : 답변} 형태로 진행 (참고 : 기본적인 데이터는 google에서 제공하지만, 국방 관련 데이터는 학습되지 않은 상태)

USER	Where is the nearest place of 'PX' from yongsan station?
ASSISTANT	Yongsan Station is a major transportation hub in Seoul, South Korea. The nearest PX to Yongsan Station would likely be the Yongsan Army Garrison PX, which is located on the Yongsan Army Garrison installation itself. To access the PX, you would need to enter the military base. Please note that access to military installations, including PXs, is typically restricted to military personnel, their dependents, and other authorized individuals. If you have specific access or eligibility questions, I would recommend reaching out to the Yongsan Army Garrison for further information.
USER	Do you know how much the PX sells "Dr. G cream"?
ASSISTANT	As an AI language model, I don't have real-time access to specific product information or pricing for PX stores, including the availability or cost of "Dr. G cream." PX stores tend to have a wide variety of products, including skincare items, but the specific brands and prices can vary. It is best to visit the PX or check their online store or catalog for the most up-to-date information on the availability and pricing of "Dr. G cream" or any other specific product you are interested in.

* 용산역 주변의 'PX' 정보는 답변하였지만, 'PX'의 구체적인 상품정보(가격)는 답변하지 못함

* 국방 관련 데이터를 학습시키면 군 관련 인원들이 원하는 정보를 채팅형식으로 신속, 편리하게 얻을 수 있음

※ 본 양식은 출품접수기간(2023. 06. 12.. ~ 07. 09.)에 제출하여 주시기 바랍니다.

- 다음은 OpenAI의 내부 서버를 이용해서 국방 공공 데이터 중 '**Px 인기상품**'의 일부를 이용하여 'Curie'라는 모델을 학습하였고, Python 언어를 사용해서 실행한 결과를 아래에 제시하였음

🔗 학습한 결과제시

: 학습과정은 #1 ~ #4에 나와있으며, 학습과정의 코드는 공유하지 않고, 데이터 및 결과만 제시하였음

국방 데이터 경진대회

✓ # 1. 필요한 라이브러리 불러오기 ...

✓ # 2. API key ...

✓ # 3. OpenAI에서 제공하는 모델로 학습 : 'curie' 모델 이용 ...

✓ # 4. 실제 학습하기 ...

Outputs are collapsed ...

```
# 5. 학습한 모델이 잘 실행되는지 확인
""" Inference """
fine_tuned_model = openai.FineTune.retrieve(id=fine_tuner.id).fine_tuned_model
example_prompt = 'Do you have "355ml Terra beer" at the military mart "PX"?
example_response = openai.Completion.create(
    model=fine_tuned_model,
    prompt=example_prompt,
    max_tokens=20,
    temperature=0.1,
)
response = example_response.choices[0]['text'].replace("\n", "")
```

```
# 6. 결과 확인
print('질문 : ', f'{example_prompt}', '\n답변 : ', f'{response}')
```

✓ 0.7s

질문 : Do you have "355ml Terra beer" at the military mart "PX"?
답변 : Yes, we have 355ml Terra beer at the military mart "PX".

🔗 사용한 데이터

- * 데이터를 모두 사용하기에 비용, 시간이 제한적이어서 일부 데이터만 사용함
- * { 'prompt' : 질문, 'completion' : 답변 } 형식으로 데이터를 전처리함

contest_train_english.jsonl X contest_train.jsonl

C:\Users> sarah > Desktop > 대학원 3학기 > chatgpt > data > contest_train_english.jsonl

```
1 {"prompt": "Is PX a military mart?->", "completion": "Yes, it's a military welfare mart.\n"}
2 {"prompt": "What does PX mean?->", "completion": "It's called PX, a military mart that soldiers can use.\n"}
3 {"prompt": "Do you have beer among PX products?->", "completion": "Yes, there is.\n"}
4 {"prompt": "What kind of beer do you have on the px?->", "completion": "There are terra, cass cans, etc.\n"}
5 {"prompt": "Do you have Cass beer in PX?->", "completion": "We have 355ml of Cass beer (can) and another product.\n"}
6 {"prompt": "Do you have Terra beer at the military mart?->", "completion": "We have 355ml of Terra Can and another product.\n"}
7 {"prompt": "Do you sell can beer on PX?->", "completion": "Yes, we sell can beer.\n"}
```

contest_val_english.jsonl X contest_val.jsonl

C:\Users> sarah > Desktop > 대학원 3학기 > chatgpt > data > contest_val_english.jsonl

```
1 "ompt": "Is PX a mart for soldiers?->", "completion": "Yes, it's a mart used by soldiers.\n"}
2 "ompt": "Is PX available to the public?->", "completion": "No, it's possible for soldiers and military families who are certified as soldiers.\n"}
3 "ompt": "Do you sell beer on PX?->", "completion": "Yes, there is.\n"}
4 "ompt": "What beer do you have among the PX products?->", "completion": "Terra and Cass exist.\n"}
5 "ompt": "Do you have Cass beer?->", "completion": "We have 355ml of Cass beer.\n"}
6 "ompt": "Do you have Terra beer at the military mart?->", "completion": "We have 355ml of Terra Can.\n"}
7 "ompt": "Can you tell me the price of beer?->", "completion": "We can't tell you the price.\n"}
```

사용한 데이터 : 국방데이터 'PX 인기상품 정보'

※ 본 양식은 출품접수기간(2023. 06. 12. ~ 07. 09.)에 제출하여 주시기 바랍니다.

- 활용 가능한 데이터 예시

구분	데 이 터		활용
1	국방 공공	국방부_영외 군인아파트 현황	군인아파트와 주변 민간 아파트를 비교하여 군인 가족의 전입(예정)시 선택지 폭을 늘릴 수 있음
	국가 중점	국토교통부_아파트 전월세 자료	
2	국방 공공	군 병원 정보	해당지역 인근의 민간병원뿐만 아니라 군 병원 의료체계 서비스를 받을 수 있음 (주소, 전화번호, 교통편 제공)
	국가 중점	건강보험심사평가원_전국 병의원 및 약국 현황	

* 저희 팀이 제시한 아이디어는 위의 예시를 제외하고도 모든 데이터를 활용 가능하다는 장점이 있음

<결론>

아이디어를 실제 서비스로 구현하는 것은 충분히 가능합니다. 필요한 데이터와 기술이 이미 현존합니다. 실제 구현을 위해서는 데이터를 전처리 및 학습(fine-tuning)할 수 있는 대용량 서버, 기술력이 필요합니다. 이러한 자원을 확보하는 것이 서비스의 성공을 위한 핵심입니다.

5-2. 사회적 가치 창출

1) 정보에 대한 접근성 향상과 이에 따른 삶의 만족도 증가

- 국방 공공 데이터 및 국가 중점 데이터를 사용자 친화적인 형태로 제공하면, 정보에 대한 접근성이 크게 향상됩니다. 이를 통해 자신들에게 필요한 정보를 더 쉽게 찾을 수 있게 되어 사회적 가치를 높일 수 있습니다.
- 예시) 군 복무에 대한 정보 : '입대 전 / 중 / 후' 교통(00부대를 가기 위한 버스나 지하철 노선), 숙박(00부대 주변의 호텔 및 모텔), 식당 · 관광지 정보를 얻을 수 있음.

* 군부대 주변 정보는 적어서 현지 주민들을 통한 정보를 얻는 경우도 많음

2) 공공 데이터 활용도 향상

- 이 서비스는 공공 데이터의 활용도를 높이는 데 크게 기여 할 수 있습니다. 이용 빈도가 낮은 많은 데이터가 활용되어 유의미한 정보 및 장치 활용되었을 때 가져올 수 있는 유의미한 분석도 가능합니다.

3) 사회적 연결성 강화

- 군 관련 인원들이 이 서비스를 통해 군인들의 상황, 복지 혜택, 군 생활에 대한 정보 등을 쉽게 접할 수 있게 된다면, 군인과 사회, 그리고 가족 간의 이해관계가 강화되어 사회적 연결성을 더욱 강화할 수 있습니다.

6. 기타

- 공공데이터포털에 있는 『국방부_국방통계연보』와 같은 파일 데이터의 경우 엑셀 파일이 아닌 한글 파일로 되어있어 데이터 분석을 하기에 불편함이 있는데, 한글 파일 내에 있는 정보의 경우 엑셀로 저장하여 공공 데이터로 공유될 수 있다면 데이터 활용 측면에서 보다 편리할 것이라 생각합니다.