

생성형 AI 기반의 재난안전 대응 챗봇 구축에 관한 연구

안진수*, 신승미

*핸디소프트

*anjinsu96@handysoft.co.kr, ohori@handysoft.co.kr

Research on building disaster response chatbots powered by generative AI

An Jin Su*, Shin Seung Mi

*Handy Soft.

요 약

재난 대응 매뉴얼은 빠르게 변화하는 재난 상황의 특성상 모든 상황에 대해 대응책을 제시하기 어려운 문제가 있다. 본 연구에서는 생성형 AI 기술, 특히 RAG(Retrieval-Augmented Generation) 방식의 검색과 생성 접근법을 이용하여 정적인 매뉴얼 한계를 해결할 수 있는 방안을 제시한다. 생성형 AI 기술을 이용해 재난안전 대응 챗봇을 구현하였고 부족한 학습데이터는 RAG 방식을 이용해 연구를 진행하였다. 생성형 AI 모델은 OpenAI 에서 공개한 ChatGPT 3.5-turbo 사전 학습 모델을 활용하였으며, 학습데이터는 행정안전부에서 제공하는 49 건의 재난 안전 매뉴얼을 수집하여 RAG 방식으로 학습하였다. 챗봇의 질의 능력을 확인하기 위해 10 종의 재난 상황에 대한 질의를 제시하고 재난 대응 답변을 확인하였다. 확인 결과 재난 안전 매뉴얼에 존재하는 답변에 대해 코사인 유사도를 비교하고 답을 도출하는 것을 확인하였다. 본 연구를 재난 대응시에 활용이 가능할 것으로 보인다. 하지만, 재난 안전 분야는 생명과 직결된 문제이기 때문에 향후 환각 현상과 편향 문제를 최소화하는 연구가 필요하다.

I. 서 론

재난 대응 매뉴얼은 위기 상황에서 대응자에게 지침을 제공하는 필수 문서이다. 하지만 이러한 매뉴얼은 급격하게 변화하는 재난 상황의 특성을 따라잡는 데 어려움이 많아 유용성에 한계가 있다.

인공 지능, 특히 생성형 AI 분야의 발전은 기존 재난 대응 매뉴얼의 단점을 해결할 수 있는 방안을 제시한다. 재난 대응의 분야에서 생성형 AI 는 대응자가 중요한 정보에 접근하고 활용하는 방식을 혁신적으로 바꿀 수 있는 잠재력을 가지고 있다.

기존 재난 대응 매뉴얼의 한계는 다음과 같다. 첫째, 재난 대응 매뉴얼의 업데이트는 주기적으로 이루어져야 하지만, 업데이트가 되지 않아 매뉴얼에 포함된 정보가 오래되거나 관련성이 떨어지는 상황이 발생한다. 정확하고 시기적절한 지침에 의존하는 대응자에게 어려움을 초래한다. 둘째, 기존 매뉴얼은 종종 새롭거나

빠르게 진화하는 재난 상황에 적용할 수 있는 대응책을 제공하는데 어려움을 가진다. 이러한 매뉴얼의 정적인 구조는 예측할 수 없는 재난 상황의 특성을 설명하지 못할 수 있다.

생성형 AI 기술, 특히 RAG(Retrieval-Augmented Generation)[1] 방식은 이러한 문제에 새로운 해결책을 제시한다. RAG 는 검색과 생성에 특화된 모델을 각각 구축하고 두가지의 장점을 결합하여 역동적이고 반응성이 뛰어난 시스템을 구축한다. RAG 는 검색 매커니즘을 활용하여 재난 대응 매뉴얼과 같은 문서에 대해 유사도를 검색하기 때문에 생성된 콘텐츠가 올바른지 확인할 수 있다. 동시에 RAG 의 생성 기능을 통해 새로운 정보를 통합하고 맞춤형 지침을 제공하여 정적인 매뉴얼의 한계를 해결할 수 있다.

본 연구에서는 생성형 AI 기술을 이용해 재난안전 대응을 위한 챗봇 모델을 구현하였고 부족한 학습데이터는 RAG 방식을 이용해 연구를 진행하였다.

이를 통해 재난안전 대응을 위한 챗봇의 활용 가능성을 확인하고자 한다.

II. 본론

2017 년에 공개된 Transformer[2] 모델의 등장 이후 NLP 분야에서는 PLM(Pre-trained Language Model)과 LLM(Large Language Model)의 발전이 이루어졌다. PLM 은 다량의 텍스트 데이터를 사전에 학습하여 언어에 대한 깊은 이해를 얻은 후, 이를 Fine-Tuning 하여 특정 NLP 작업에 활용하는 모델이다. LLM 은 PLM 보다 큰 규모의 모델로, 수억 개에서 수천억 개에 이르는 많은 파라미터를 갖추며, 이는 더 깊은 언어적 표현과 더 복잡한 문맥 이해 능력을 제공한다. 이와 같은 LLM 의 능력을 이용하여 재난안전 분야에 활용하고자 한다. 하지만 LLM 을 구축하는 것은 많은 데이터의 학습을 필요로 하기 때문에 많은 GPU 리소스와 데이터가 있어야 한다. 이 문제를 해결하기 위해 OpenAI 에서 공개한 ChatGPT 3.5-turbo[3] 사전 학습 모델을 활용하고 재난안전 매뉴얼 데이터를 이용하여 챗봇을 구현한다. ChatGPT 3.5-turbo 모델과 같은 대형 사전 학습 모델을 Fine-Tuning 하려면 고성능의 GPU 와 메모리 및 리소스가 필요하다. 해당 문제를 해결하기 위해 Fine-Tuning 이 필요하지 않은 RAG 방식을 이용한다. RAG 방식은 사용자의 질의가 들어올 때 관련된 정보를 벡터 데이터베이스에서 검색하여 해당 정보가 담긴 문서들을 프롬프트에 담아 LLM 에 전달하는 방식이다. RAG 방식은 이미 저장되어 있는 문서를 기반으로 답을 도출하기 때문에 LLM 의 대표적인 문제인 환각[4] 현상을 감소시킬 수 있다. 본 연구에서는 RAG 방식을 지원하는 LlamaIndex[5] 라이브러리를 이용하였다. [그림 1]은 LlamaIndex의 작동방식을 보여준다.

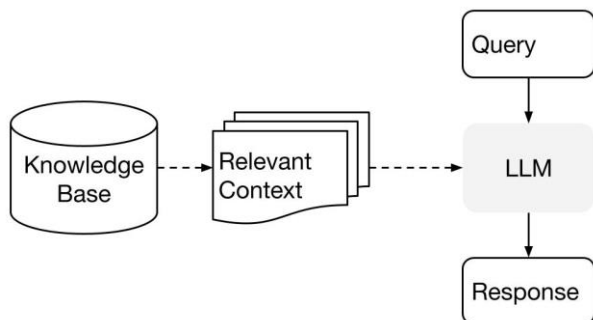


그림 1. LlamaIndex Retrieval Augmented Generation

학습데이터로는 행정안전부에서 제공하는 49 건의 재난안전 매뉴얼[6]과 경상북도 소방본부 국민행동요령 웹 데이터[7]를 수집하였고, 데이터 전처리는 LlamaIndex 라이브러리의 데이터 커넥터 기능을 활용하여 문서를

로드하고 저장하였다. RAG 데이터 검색을 위한 임베딩 작업은 OpenAI 의 API(text-embedding-ada-002)[8]를 이용하였다.

질의 능력을 평가하기 위해 10 종의 재난에 대한 질의를 제시하고 답변을 확인하였다. 재난 안전 매뉴얼에 존재하는 답에 대해 LlamaIndex 를 이용해 코사인 유사도를 비교하고 답을 도출하는 것을 확인할 수 있다. [표 1]

표 1. 태풍 대응 정보 질문에 대한 결과 예시 일부

| Q. 태풍 발생시 어떻게 행동해야 하나요? |
|---|
| <p>A. 태풍 발생시에는 다음과 같은 행동을 취해야 합니다:</p> <ol style="list-style-type: none"> 1. 라디오나 TV를 통해 태풍의 진로와 도달시간을 숙지합니다. 2. 가정의 하수구나 집주변의 배수구를 점검하고 막힌 곳을 뚫어야 합니다. 3. 침수나 산사태가 일어날 위험이 있는 지역에 거주하는 주민은 대피장소와 비상연락방법을 미리 알아둡니다. <p>...중략</p> |

III. 결론

본 연구에서는 생성형 AI 모델을 활용하여 재난안전 매뉴얼을 학습한 재난안전 대응 챗봇을 구현하여 활용 가능성을 확인해 보았다. 질의 능력을 평가한 결과, 질의의 유사도에 따라서 재난 안전 매뉴얼에 존재하는 답변을 추론하는 것을 확인해 볼 수 있었다. 본 연구를 통해 정적인 재난 대응 매뉴얼의 한계를 극복할 수 있는 방안을 제시하였으며 재난 대응 시에 활용이 가능할 것으로 기대된다. 하지만, 재난 안전 분야는 생명과 직결된 문제이기 때문에 무엇보다 정확한 정보를 추론하는 것이 중요하다. 이를 위해 생성형 AI 모델의 문제인 환각 현상과 편향 문제를 최소화하는 연구를 지속할 예정이다.

ACKNOWLEDGMENT

본 연구는 행정안전부 생활안전 예방서비스 기술개발사업의 연구비 지원(2019-MOIS34-001(RS-2019-ND618021))에 의해 수행되었습니다.

참 고 문 헌

- [1] LEWIS, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems*, 33: 9459-9474, 2020.
- [2] VASWANI, Ashish, et al. "Attention is all you need," *Advances in neural information processing systems*, 30, 2017.
- [3] OpenAI. "Introducing ChatGPT," 2022, (<https://openai.com/blog/chatgpt>).
- [4] Maynez, Joshua, et al. "On faithfulness and factuality in abstractive summarization." *arXiv preprint arXiv:2005.00661*, 2020.
- [5] Liu, Jerry. "LlamaIndex" 2022, (https://github.com/jerryliu/llama_index).
- [6] Ministry of the Interior and Safety. "National Disaster Safety Portal Resource Center", (https://www.safekorea.go.kr/idsiSFK/neo/sfk/cs/csc/bbs_conf.jsp?menuSeq=593&bbs_no=9).
- [7] Gyeongbuk Fire Service Headquarters. "Emergency Response Guidelines", (<https://www.gb119.go.kr/disaster/disaster.do>).
- [8] Ryan Greene, et al. "new-and-improved-embedding-model" 2022, (<https://openai.com/blog/new-and-improved-embedding-model>).