NLP Development to Automate Proposal Writing for Nuclear Utility Clients

PURDUE UNIVERSITY | The Data Mine

BLUE WAVE AI LABS

INTRODUCTION

Blue Wave AI Labs builds actionable and ethical AI solutions that evaluate and mitigate risk, increase efficiency, and optimize performance for the nuclear energy and defense industries by transforming data into solutions. Blue Wave spends a considerable deal of time writing proposals to companies to lay out the terms of potential business together, which has incurred a fair amount of labor costs and lead time. Therefore, Blue Wave wants to automate the writing of proposals for their nuclear utility clients to reduce labor hours and improve lead time.

OBJECTIVE

Using open-source LangChain components, this project aims to develop a nuclear proposal chatbot that runs on private servers and does not share sensitive information with other entities on the Internet. Based on prompts and text-based interactions, the tool will generate and enhance proposals related to the nuclear energy domain using Natural Language Processing (NLP) for automation.

INTERACTION

One example of an interaction with the chatbot could be:

- User asks a question
- Chatbot provides an answer, or opts out of answering if the question is outside its domain of knowledge
- Question and answer are added to conversation history
- User is prompted to rate the chatbot's answer
- Thumbs up, thumbs down, or no rating
- User is prompted for reasoning/comment if they gave a thumbs down
- Rating is saved to database

RAG Search

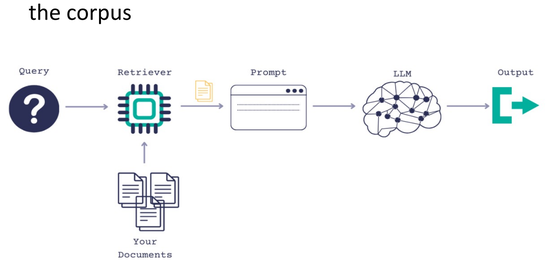
Retrieval augmented generation (RAG):
A method to retrieve information from large databases; used with a large language model (LLM). The trained LLM can generate a response to questions based on the results of the RAG search.
RAG (input question) --> RAG output
LLM (RAG output) --> final output

Works

- Used the 'all-MiniLM-L6-v2' and 'Zephyr' models to create chatbot
- Implemented output scoring using reinforcement learning with human feedback (RLHF)
- Accessed corpus for database used in RAG search

Methods

- Used Python libraries such as HuggingFace, Transformers, and LangChain
- To create our final model, we tested many different HuggingFace models on smaller scales, and then chose the most efficient one to build on top of, and train on the corpus




```
question = "What percent of the world's energy is generated by power plants?"
context = "As of 2011, about 15 percent of the world's electricity is generat

answer = question_answerer(question=question, context=context)
print(answer)
```

```
{'score': 0.3595345616340637, 'start': 18, 'end': 28, 'answer': '15 percent'}
```

Web Scraping

Web Scraping
The process of extracting content from a website.
Web Crawling
The process of discovering URLs on the website.
Nuclear Regulatory Commission (NRC)
Government agency that regulates use of nuclear material.

Works

- Scraped NRC website for all documents
- Stored all documents
- Scraped documents and extracted text
- Stored all data to add to corpus

Methods

- Used Python libraries such as Requests, PyPDF, BeautifulSoup
- Used the NRC API to access the documents

Document Title	Accession Number	Date Added	Document Date	Size
Federal Register Notice - Notice of Public Meeting - Licensing Support Network Advisory Review Panel - October 13, 1999	ML992520041	10/11/1999 08:33 PM EDT	09/09/1999	8.65 Kb

```
link = link_tag['href']
print(link)
```

```
https://adams.nrc.gov:443/wba/view?action=view&actionId=view&ids=[{%22documentId%22:%7B%22dataProviderId%22%3A%22ce_bp8os_repository%22%2C%22compound%22%3Afalse%2C%22properties%22%3A%7B%22%24os%22%3A%22ADAMSPublicOS%22%2C%22ce_object_id%22%3A%22%7B17B1ADFB-24C6-4721-8FB6-D36EE48C159A%7D%22%2C%22%24is_compound%22%3Afalse%7D%7D}]&mimeType=application/pdf&docTitle=Federal Register Notice - Notice of Public Meeting - Licensing Support Network Advisory Review Panel - October 13, 1999
```

```
title = link_tag.text
print(title)
```

```
Federal Register Notice - Notice of Public Meeting - Licensing Support Network Advisory Review Panel - October 13, 1999
```

```
print(response.headers['Content-Type'])
```

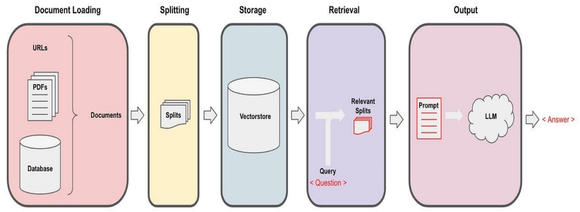
```
application/pdf
```

```
with open(f"{title}.pdf", 'wb') as file:
    file.write(response.content)
    print(f"Downloaded: {title}.pdf")
```

```
Downloaded: Federal Register Notice - Notice of Public Meeting - Licensing Support Network Advisory Review Panel - October 13, 1999.pdf
```

How They Work Together

- Web scrape to continuously update the corpus
- Process the scraped data to be usable for the model
- Train the chatbot on the corpus
- Ask a question to the chatbot
- RAG search looks through the corpus to find a range of possible answers
- Chatbot takes the RAG output and figures out the best answer for the user



Results

- Fine-tuned the model to take less than 30 seconds to answer questions
- Expanded the corpus to further its nuclear domain knowledge, increasing accuracy of answers
- Chatbot saves conversation history, so it can better expand on previous answers

```
%%time
# Ask a question
question = "How is nuclear energy created?"
answer = conversation_buf(question)['response']
```

```
CPU times: user 7.73 s, sys: 61.7 ms, total: 7.79 s
Wall time: 10.3 s
```

```
record_score(question, answer, model_name)
```

```
Please provide a rating for the answer provided by chatbot
    'Question' : How is nuclear energy created?
    'Answer'   : The following is a friendly conversation between a human and an AI. The AI is talkative and provides lots of specific details from its context. If the AI does not know the answer to a question, it truthfully says it does not know.

Current conversation:

Human: How is nuclear energy created?
AI: Nuclear energy is generated through a process called nuclear fission. In this process, the nucleus of an atom is split into smaller particles, releasing a tremendous amount of energy in the process. This process occurs inside a nuclear reactor, which contains fuel rods made up of enriched uranium or plutonium. The fuel rods are heated by control rods, which regulate the rate at which the reaction occurs. The heat produced by the fission is then used to generate steam, which drives turbines that produce electricity. The remaining waste products from the fission process are highly radioactive and must be carefully stored and disposed of to prevent environmental harm.

Please rate the answer:
        for          👎 enter         : 0
        for          👍 enter         : 1
        for exit without rating enter : e
```

Conclusions

Implementing Web Scraping and RAG Search leads to a cohesive final product which includes a fine-tuned machine learning model that is capable of answering nuclear-related questions for Blue Wave clients. The model allows many people to be serviced at one time, and lessens the load of cost, time, and resources.

Future Work

RAG Search: 1) Optimizing chatbot to provide higher quality answers by further training of the LLM through RLHF. 2) Extracting data more efficiently by optimizing the algorithm. 3) Working on building a user-friendly GUI.

Web Scraping: 1) Fine-tune the model by constantly updating the web scraping function. 2) Web scraping more websites to give the model more data.

Acknowledgments

Our sincerest gratitude to Maria Berardi, Rizki Oktavian, Phil McCalley and everyone at Blue Wave AI Labs for their guidance and support in this project. We would also like to extend our utmost thanks to all the Data Mine staff that helped to make this project happen.

CONTRIBUTORS: Sharanya Devgon, Enes Erickan, Rohan Gaddam, Sara Huang, Keisuke Nakamura, Elina Shah, Pranav Tallapaka