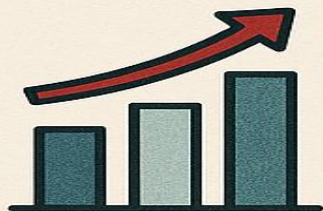
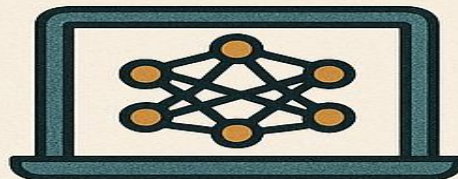


DATATHON



orderfox

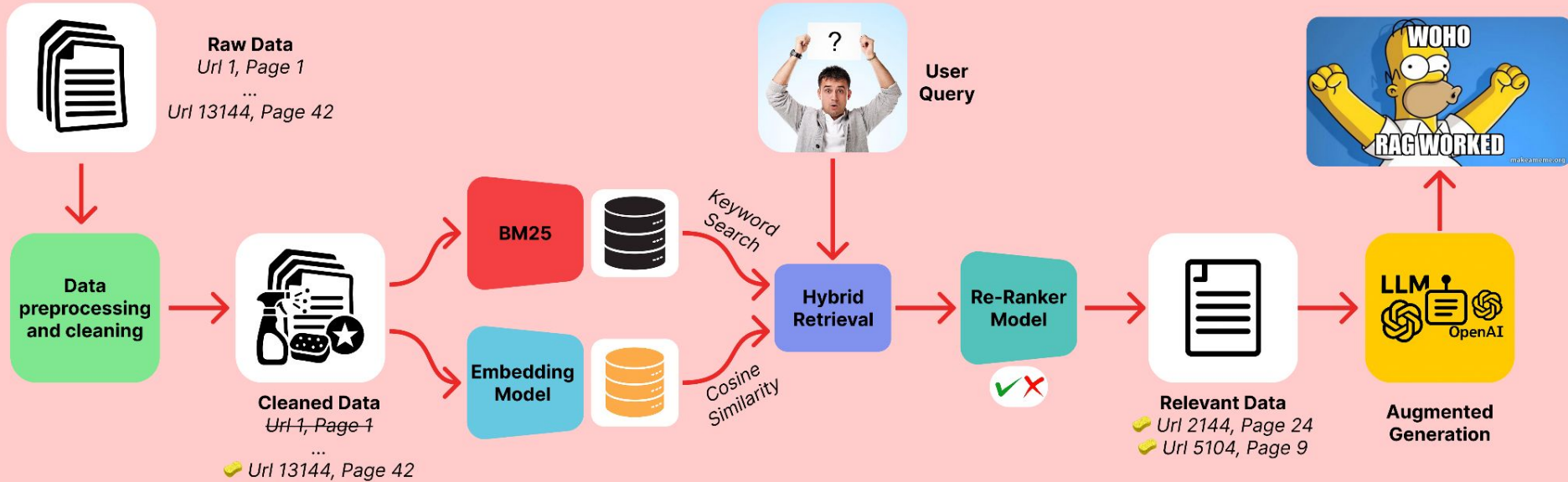


**RAGE AGAINST
THE MACHINE
LEARNING**



RAG(E) Pipeline

RAG(E Against the Machine Learning)



Data Cleaning and Preprocessing

- ❑ Only keep alphanumeric characters
- ❑ Remove stopwords
- ❑ Lowercase
- ❑ Lemmatization
- ❑ Remove large garbage pages
- ❑ Chunked each page into overlapping windows

Indexing Strategies

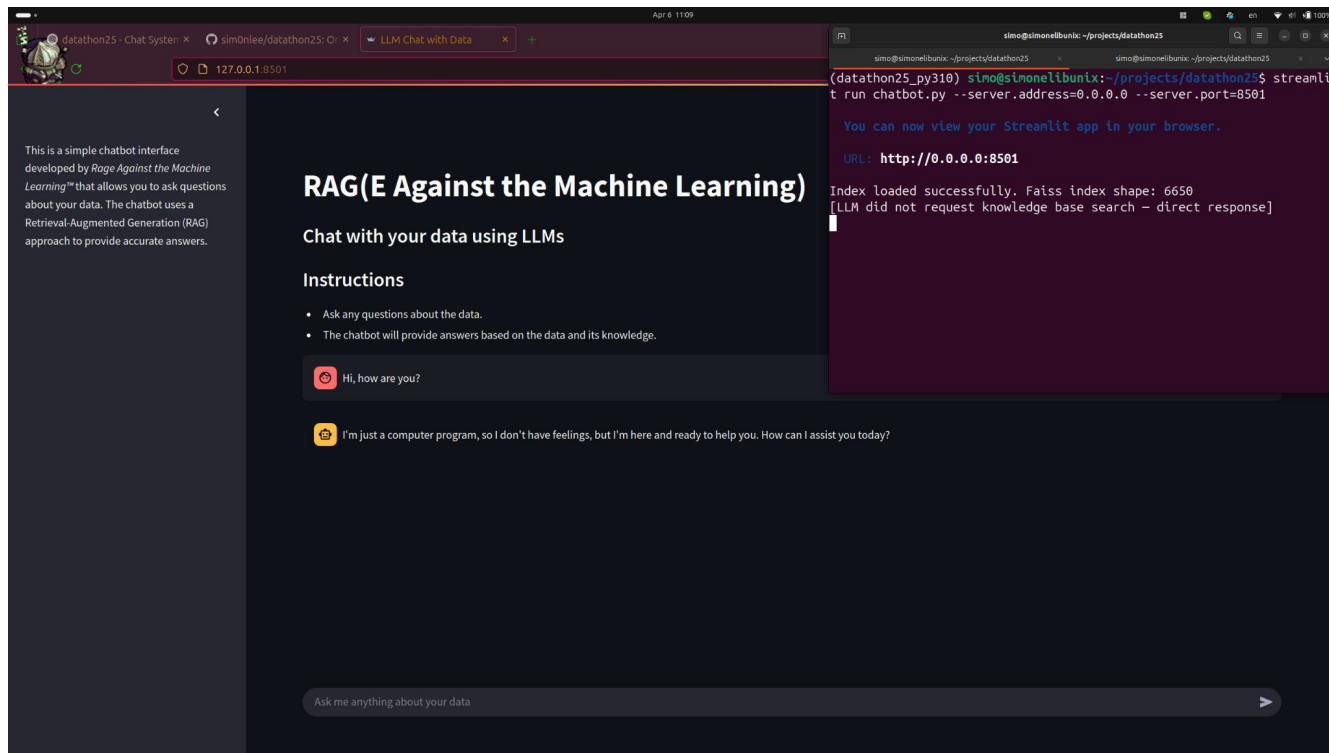
1. Embeddings Indexing with FAISS

- LLM-embeddings-based approach for efficient similarity search and retrieval using FAISS indexing.

2. BM25 Indexing

- BM25 uses traditional term-frequency-based scoring for fast and effective lexical retrieval.

Retrieval and Generation (1/3)



The screenshot displays a web browser window with a chatbot interface on the left and a terminal window on the right.

Chatbot Interface:

- Header:** "datathon25 - Chat System" and "simonlee/datathon25: O".
- URL:** "127.0.0.1:8501".
- Tab:** "LLM Chat with Data".
- Intro Text:** "This is a simple chatbot interface developed by *Rage Against the Machine Learning*™ that allows you to ask questions about your data. The chatbot uses a Retrieval-Augmented Generation (RAG) approach to provide accurate answers."
- Section:** "RAG(E Against the Machine Learning)"
- Subtitle:** "Chat with your data using LLMs"
- Instructions:**
 - Ask any questions about the data.
 - The chatbot will provide answers based on the data and its knowledge.
- Chat History:**
 - User: "Hi, how are you?"
 - Bot: "I'm just a computer program, so I don't have feelings, but I'm here and ready to help you. How can I assist you today?"
- Input Field:** "Ask me anything about your data" with a send button.

Terminal Window:

- Path:** "simo@simonleebunick: ~/projects/datathon25"
- Command:** `(datathon25_py310) simo@simonleebunick:~/projects/datathon25$ streamlit run chatbot.py --server.address=0.0.0.0 --server.port=8501`
- Output:**
 - "You can now view your Streamlit app in your browser."
 - "URL: <http://0.0.0.0:8501>"
 - "Index loaded successfully. Faiss index shape: 6650"
 - "[LLM did not request knowledge base search - direct response]"

Retrieval and Generation (2/3)

The image shows a Streamlit chatbot interface on the left and a terminal window on the right. The chatbot interface has a dark theme and includes a description of its Retrieval-Augmented Generation (RAG) approach. It features an 'Instructions' section with a list of capabilities and a chat input area. The terminal window shows the command to run the chatbot and its output, which includes the URL to access the chatbot and a detailed log of the RAG process, including the search for 'Farwest Corrosion Company overview' and the retrieval of relevant documents.

This is a simple chatbot interface developed by *Rage Against the Machine Learning™* that allows you to ask questions about your data. The chatbot uses a Retrieval-Augmented Generation (RAG) approach to provide accurate answers.

Instructions

- Ask any questions about the data.
- The chatbot will provide answers based on the data and its knowledge.

What do you know about farwest corrosion company?

Farwest Corrosion Control Company is a leader and industry pioneer in cathodic protection and corrosion control services. They offer a comprehensive range of services and products related to cathodic protection and corrosion control, including engineering, installation, and maintenance. The company serves as a full-line distributor and manufacturer of corrosion prevention products, such as coatings, isolation equipment, and test instruments. They are also known for custom fabricating cathodic protection systems and related specialty products.

Established originally in Seattle, Washington, Farwest has expanded over the years to numerous locations across the United States, including a headquarters in Downey, California. They have made several acquisitions to enhance their service offerings and market reach, such as the purchase of Seaboard Supply Company and Pacific Corrosion Research.

Farwest emphasizes integrity, customer satisfaction, and quality service. They operate with a mission of finding continuous improvement in their industry while maintaining a solid foundation of business stability and dependability.

Ask me anything about your data

```
simo@simonellibunx: ~/projects/dataathon25
simo@simonellibunx:~/projects/dataathon25$ streamlit run chatbot.py --server.address=0.0.0.0 --server.port=8501

You can now view your Streamlit app in your browser.

URL: http://0.0.0.0:8501

Index loaded successfully. Faiss index shape: 6650
[LLM tool call]
[ChatCompletionMessageToolCall(id='call_RQ6Q6tQlLycRGHhB2FPvZN', function=Function(arguments={'query': 'Farwest Corrosion Company overview'}, name='do_rag'), type='function')]
[LLM requested database search - executing RAG]
[RAG query (LLM-generated)] Farwest Corrosion Company overview
[RAG user prompt] [QUESTION]
What do you know about farwest corrosion company?

Use ONLY the following documents to answer the question:

Document 1: farwest corrosion control company u farwest corrosion control company in
dustry pioneer leader comprehensive cathodic protection corrosion control service re
lated product cathodic protection corrosion control service include product engineer
ing installation farwest leading full line distributor manufacturer corrosion preven
tion product include coating isolation equipment test inspection tool instrumentatio
n addition firm custom fabricator cathodic protection cp anode junction bond box pow
er supply test station test equipment variety specialty product cathodic protection
corrosion control engineering service include system design solution cathodic protec
tion field survey system testing monitoring troubleshooting technical consultation p
roject oversight product design installation service include cathodic protection sys
tem installation system abandonment grounding ac mitigation system system maintenanc
e repair project planning installation project use drilling construction equipment ol
wined

Document 2: ed product service solution mission find way improvement leader industry
maintaining solid foundation business stability dependability vision known helping
people always driven find solution value integrity customer commitment quality outpu
t care positive attitude promise provide right solution cathodic protection corrosio
n control challenge integrity attention customer need quality second none introducti
on farwest corrosion control leader cathodic protection corrosion control history fa
rwest corrosion control corporation established seattle wa gordon rankin accepts pos
ition wayne broyles engineering houston tx gordon accepts offer broyles run farwest
```


Retrieval and Generation (3/3)

The image shows a Streamlit chatbot interface on the left and its terminal output on the right.

Chatbot Interface:

- Header: "This is a simple chatbot interface developed by *Rope Against the Machine Learning*™ that allows you to ask questions about your data. The chatbot uses a Retrieval-Augmented Generation (RAG) approach to provide accurate answers."
- Question: "What services does a spectrum engineer provide? Arrange results in a table"
- Answer: "Based on the documents provided, the services offered by a Spectrum Engineer are summarized in the table below:"
- Table:

Service Category	Description
Mechanical Engineering	Includes HVAC, plumbing, and sustainable design to achieve energy efficiency goals.
Electrical Engineering	Encompasses reliable power systems, medium voltage distribution, and fault current analysis.
Plumbing Engineering	Designs systems with a focus on sustainability, water conservation, air protection, and waste elimination.
Fire Protection Engineering	Designs fire protection systems including fire sprinklers, fire alarms, fire suppression, and smoke management.
Audio Visual	Design systems for conveying messages and data in an understandable and operable manner.
Structured Cabling	Critical for data and voice communication systems.
Acoustical Design	Provides design services for acoustic environments.
Lighting Design	Offers lighting design services, potentially including energy-efficient solutions.
Theatre Design	Involves design services for theatre infrastructure.
Security Consulting	Offers consulting services for security system design.

These services are aimed at providing robust, integrated engineering solutions tailored to client needs across various sectors.

Input field: "Ask me anything about your data"

Terminal Output:

```
simo@simonellibunx: ~/projects/datathon25
(simonellibunx:~/projects/datathon25$ streamlit run chatbot.py --server.address=0.0.0.0 --server.port=8501)

You can now view your Streamlit app in your browser.
URL: http://0.0.0.0:8501

Index loaded successfully. Faiss index shape: 6650
[LLM tool call]
[ChatCompletionMessageToolCall(id='call_U8lYqrhFKRWSENY7vG9x8acr', function=Function(arguments=({'query':'spectrum engineer services'}, name='do_rag'), type='function'))]
[LLM requested database search - executing RAG]
[RAG query (LLM-generated)] spectrum engineer services
[RAG user prompt] [QUESTION]
What services does a spectrum engineer provide? Arrange results in a table

Use ONLY the following documents to answer the question:

Document 1: service spectrum engineer skip content spectrum engineer professional engineering firm home service project team article career contact linkedin page open new window instagram page open new window facebook page open new window twitter page open new window connect engineer home service project team article career contact service mechanical engineering spectrum engineer provides host mechanical engineering service including hvac plumbing sustainable design mechanical engineer understand energy efficiency goal poised assist achieving view project electrical engineering spectrum engineer provides comprehensive electrical engineering service including reliable practical system medium voltage distribution fault current analysis spectrum project call robust practical solution help view project plumbing engineering spectrum plumbing engineer design system whole building mind implementing sustainable system conserve water use protect air breathe eliminate waste safe manner view project fire protection goal

Document 2: spectrum engineer skip content spectrum engineer professional engineering firm home service project team article career contact linkedin page open new window instagram page open new window facebook page open new window twitter page open new window connect engineer home service project team article career contact spectrum firm profile spectrum engineer multi discipline consulting engineering firm office satellite city utah phoenix arizona boise idaho baltimore maryland spectrum provides centered service throughout united state various location around world offering electrical mechanical acoustical fire protection lighting engineering technology lighting theatre design spectrum provides sustainable design exemplifying attitude caring future generations
```