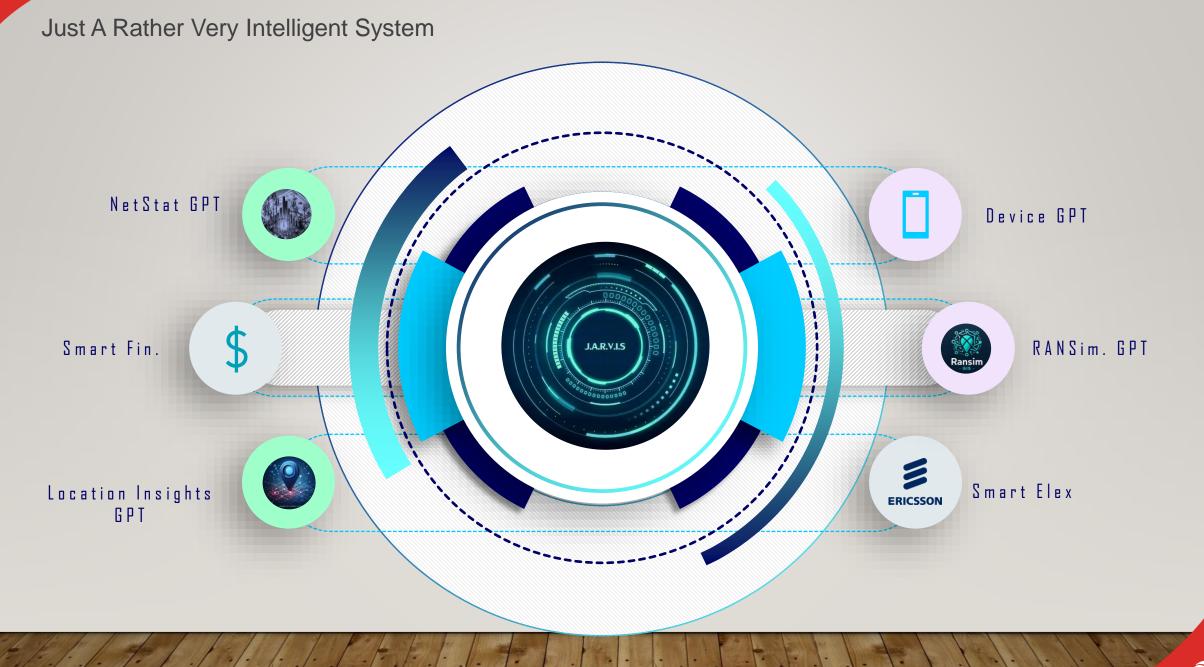
The J.A.R.V.I.S Initiative

RAN Engineering and Innovation Abhishek Paul

April, 2024





BUILDING THE CHATBOT

The Workflow

Data Conversion & Preprocessing

 Convert downloaded PDFs and XMLs from ELEX library to document objects.

Data Store

 Save the document objects into a Pickle file.

Build Pipeline

- Build an Q/A pipeline with the data store and downloaded pre-trained models from Hugging face.
- Model used for document search: multi-qa-mpnet-base-dotv l
- Model used for generative:
 Ilama-2-7B-chat-hf

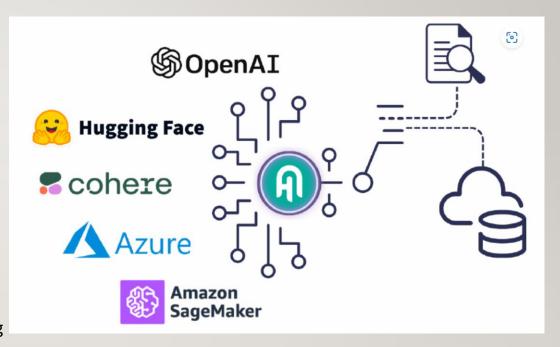
Deploy Web App

Deploy the Q/A
 pipeline as an web app
 using Django web
 server.



BUILDING THE CHATBOT

- Supporting Framework- Haystack
 - Why Choose Haystack?
 - Open-Source Efficiency: open-source, production-ready solution.
 - Model Exploration: Dive into advanced NLP models like BERT, RoBERTa, and LLMs (e.g., GPT-4, Llama-2).
 - Versatile Tasks: Simplify Information Extraction, Question Answering, and Semantic Document Search.
 - Unified Development: Streamlined development with integrated tools - Data Uploading, Preprocessing, Building Pipelines, etc.





SMARTELEX OVERVIEW

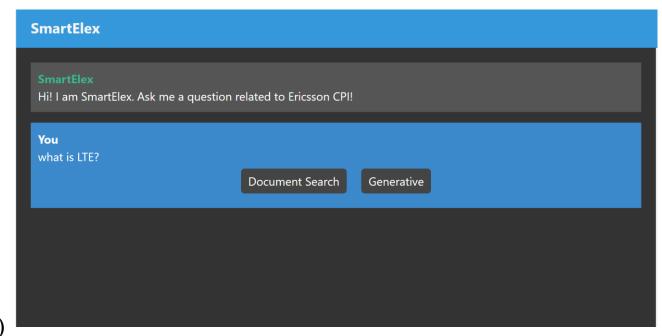
- Generates Answers for Ericsson CPI Questions
- Two Answering Options:

Document Search

- Utilizes Transformer Model
- Pros: Fast, Exact Text Extraction
- Cons: Doc Dependency

2. Generative

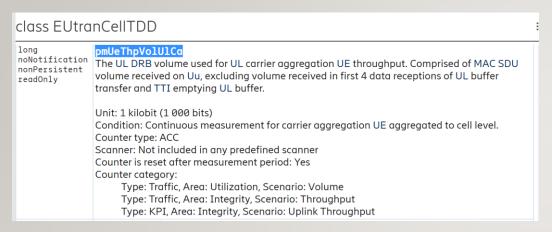
- Employs LLM (Large Language Model)
- Pros: Contextual Responses
- Cons: Slower, Requires GPU Power

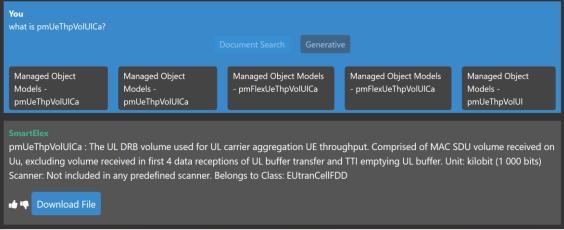




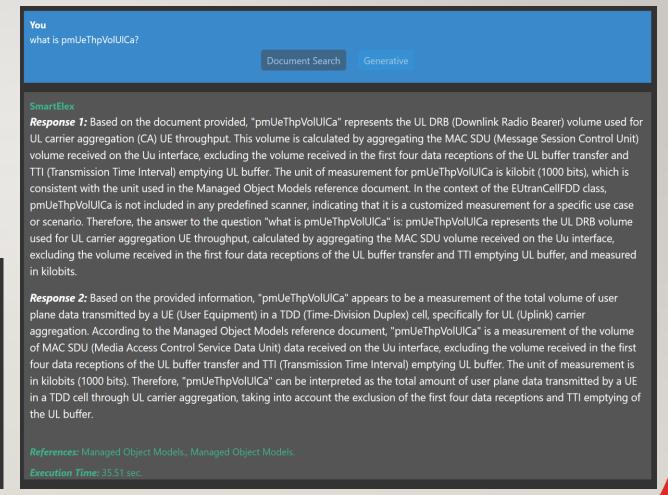
SMARTELEX

Demo





Document Search



Generative



3GPP Delay Recommendations

This section provides the GBR and non-GBR delay recommendations from 3GPP 23,203 as described in Table 10.

To derive the Packet Delay Budget (PDB) that applies to the radio interface, a delay of 20 ms for the delay between a PCEF and an RBS should be subtracted from an identified PDB. This is the average delay between the situation where the PCEF is located close to the RBS, approximately 10 ms, and the situation where the PCEF is located far from the RBS, for example in a roaming scenario with home-routed traffic. The one-way packet delay between Europe and the US west coast is approximately 50 ms. The average delay estimation takes into account that roaming is a less typical scenario. The expectation is that by subtracting this average delay of 20 ms from a given PDB, this leads to a desired end-to-end performance in most typical cases. The PDB defines an upper boundary. Actual packet delays, particularly for Guaranteed Bitrate (GBR) traffic, should typically be lower than the PDB specified for a QoS Class Identifier (QCI) as long as the UE has sufficient radio channel quality.

The rate of non-congestion-related packet loss that may occur between an RBS and a PCEF should be regarded to be negligible. A Packet Error Loss Rate (PELR) value specified for a standardized QCI therefore applies solely to the radio interface between the UE and the RBS.

Table 10 Delay Recommendations from 3GPP 23.203

QCI	Resource Type	Priority	PDB	PELR	Example Services
1	GBR	2	100 ms	10-2	Conversational Voice
2		4	150 ms	10-3	Conversational Video (Live Streaming)
3		3	50 ms	10-3	Real Time Gaming
4		5	300 ms	10-6	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10-6	IMS Signaling
6		6	300 ms	10-6	Video (Buffered Streaming) TCP-based (for example, WWW, e-mail, chat, FTP, P2P file sharing, progressive video, and so on)
7		7	100 ms	10-3	Voice, Video (Live Streaming) Interactive Gaming
8		8	300 ms	10-6	Video (Buffered Streaming) TCP-based
9		9			(for example, WWW, e-mail, chat, FTP, P2P file sharing, progressive video, and so on)

A QCI of 1-5 or 7 is typically associated with an operator-controlled service, that is, a service where the Service Data Flow (SDF) aggregate UL or DL packet

Source Document

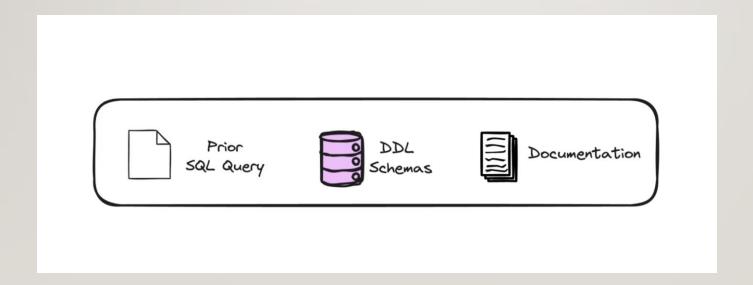
the GBR and non-GBR delay recommendations from 3GPP According to the provided documents, the GBR and non-GBR delay recommendations from 3GPP are as follows: Table 10 Delay Recommendations from 3GPP 23.203 Resource Type Priority PDB (ms) PELR **GBR** 2 100 ms 10-2 Non-GBR 4 10-3 150 ms Non-GBR 3 10-3 50 ms Non-GBR 5 10-6 300 ms Non-GBR 10-6 100 ms Non-GBR 6 300 ms 10-6 **GBR** 10-3 100 ms Non-GBR 100 ms 10-3 **Table 4 E2E QoS Guidelines** Resource Type Priority Level PDB (ms) PER GBR High 80 ms 1% Non-GBR Medium 150 ms 5% Note: The PDB values are one-way delays, and the PELR values represent the maximum allowed packet error rate.

Query Search



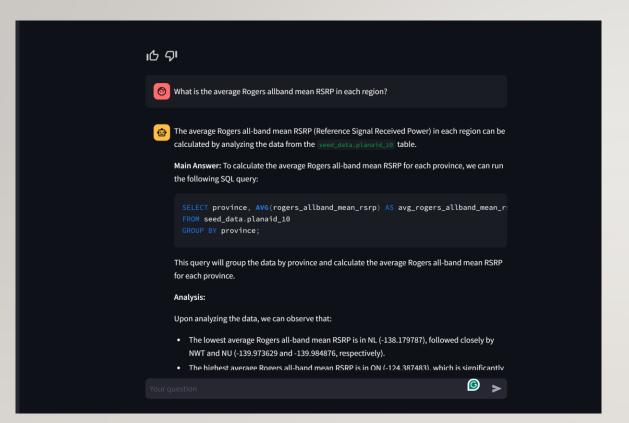
LOCATIONINSGHTS GPT

- Fed it the database schema for our LocationInsights database
- Utilized Vanna for training and LammaIndex as the base model
- Users can ask it a question and it will generate the sql query and return the answer





DEMO





Additional Features

- References:
 - Document search shows top 5 documents along with section names.
 - Generative displays the documents used as references.
- Voting and feedback:
 - Collect user feedback through voting.
 - Use the feedback to train and refine the model.
- Downloadable files:
 - Conveniently download files.
 - Easily access documents of interest.