



Department of Artificial Intelligence and Data Science

"Intelligent Chatbot for Government Education Schemes"

Supervisor
Mr S Suresh kumar
M.E.,(Ph.D) Professor

Nithish S 221801502

Problem Statement and Motivation

- How can we create a chatbot that can provide reliable and relevant information about various government schemes to the citizens, and help them identify their eligibility criteria for different programs, using an intuitive interface, engaging conversation, and diverse scenarios?
- Many citizens may find it challenging to navigate complex government websites and documents. A chatbot can simplify access to information. Automated information dissemination reduces the burden on government offices and call centers. A chatbot can provide tailored information based on user inputs, improving the relevance of the information about the government schemes.

Objectives

- Creating a chatbot for government education scheme information and eligibility requires a robust framework. It should integrate reliable data sources to ensure accuracy and relevance in its responses. The interface must be intuitive, offering clear navigation and engaging dialogue to guide users effectively. Diverse scenarios should be anticipated and included in its training to handle varied queries and edge cases adeptly.
- Regular updates are crucial to keep pace with changing eligibility criteria and new programs. Incorporating user feedback mechanisms will further refine the bot's capabilities, enhancing its ability to provide personalized and valuable assistance to citizens seeking government scheme information.

Abstract

The development of a chatbot designed to provide citizens with reliable information about government education schemes, utilizing advanced natural language processing techniques, including BERT (Bidirectional Encoder Representations from Transformers), Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs). BERT enhances the chatbot's ability to understand user queries by capturing context and intent, while RAG retrieves pertinent information from a dynamic database of educational programs, ensuring responses are both accurate and up-to-date. LLMs facilitate the generation of coherent, conversational replies, resulting in a user-friendly interface. By integrating these technologies, the chatbot aims to improve access to educational resources, promote informed decision-making, and ultimately enhance public service delivery in the education sector.

Introduction and Overview of the Project.

In an era where information is abundant yet often overwhelming, chatbots have emerged as effective tools for providing users with instant access to relevant data. Specifically, a chatbot focused on government education schemes can assist citizens in understanding various programs, eligibility criteria, application processes, and benefits. By utilizing advanced Natural Language Processing (NLP) techniques, such as BERT (Bidirectional Encoder Representations from Transformers), Large Language Models (LLM), and Retrieval-Augmented Generation (RAG), developers can create a sophisticated chatbot that delivers accurate and contextually relevant information.

Literature Survey

S.No	Author Name	Paper Title	Description	Journal	Volume/ Year
01	Jin K. Kim	ChatGPT and large language model (LLM) chatbots	Performed a scoping review of available literature to understand the current state of LLM use in medicine and to provide a guideline for future utilization in academia.	IEEE	VOLUME 19
02	Samuel Kernan Freire	Conversational Assistants in Knowledge-Intensive Contexts: An Evaluation of BERT versus Intent-based Systems	Conversational Assistants (CA) are increasingly supporting human workers in knowledge management. Traditionally, CAs respond in specific ways to predefined user intents and conversation patterns.	IEEE	2023
03	Sumit Kumar Dam	A Complete Survey on BERT AI Chatbots	Conversational agents, often referred to as AI chatbots, rely heavily on such data to train large language models (LLMs) and generate new content (knowledge) in response to user prompts.	IEEE	2022

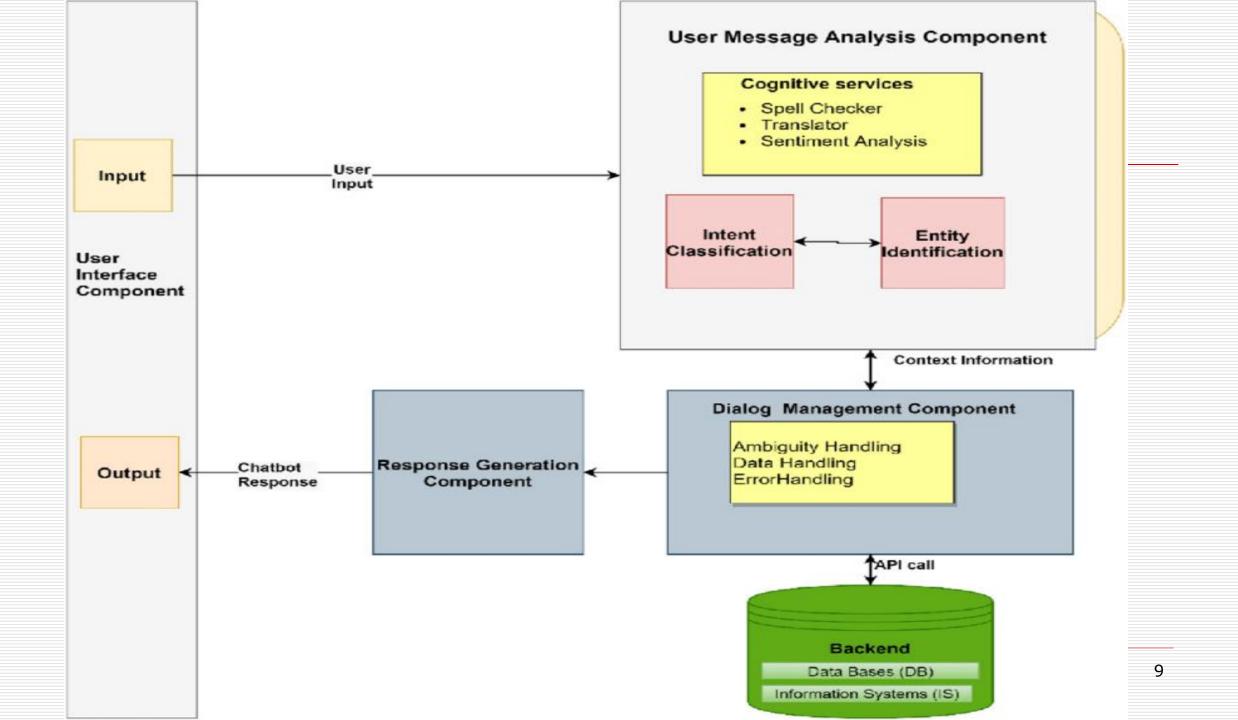
Third Review

Department of Artificial Intelligence and Data Science

S.No	Author Name	Paper Title	Description	Jorna I	Volume/ Year
04	Juan De Lara	Automating the Development of Task- oriented LLM-based Chatbots	Task-oriented chatbots are increasingly used to access all sorts of services – like booking a flight, or setting a medical appointment – through natural language conversation. There are many technologies for implementing task-oriented chatbots, including Dialogflow, Watson, and Rasa.	IEEE	2023
05	Babymol Kurian	GovInfohub: A Dynamic Government scheme Chatbot for informed Engagement and Accessibility	'GovInfoHub' presents a pioneering solution to provide citizens with real-time and accessible information regarding government schemes.	IEEE	2022

Proposed System

- ☐ Advanced Natural Language Processing Multi-Language Support:
 Implement support for multiple languages to make the chatbot accessible to a global audience.
- □ Context Switching: Allow users to change topics smoothly without losing track of the conversation flow.
- □ **Response Variation:** Implement multiple response options for the same query to make interactions feel more natural.
- ☐ **Feedback Collection:** Allow users to rate responses and provide feedback to improve the chatbot.



Algorithm

- 1. Collect, clean, and format data on government education schemes for training and testing.
- 2. Train or fine-tune a pre-trained BERT model to understand queries and provide relevant responses about schemes.
- 3. Set up a Django backend to handle API requests and responses from the BERT-based chatbot.
- 4. Develop the user-facing website using PHP for server-side logic and HTML/CSS for the interface.
- 5. Connect the trained BERT model via Django to handle chatbot queries and responses dynamically.
- 6. Embed the chatbot on the website, ensuring it communicates with the Django backend.
- 7. Perform unit, integration, and user testing to ensure proper chatbot and website functionality.
- 8. Deploy the website and backend on a web server, ensuring the chatbot is live for user interaction.
- 9. Continuously monitor, update the chatbot model, and maintain the website for new schemes or enhancements.

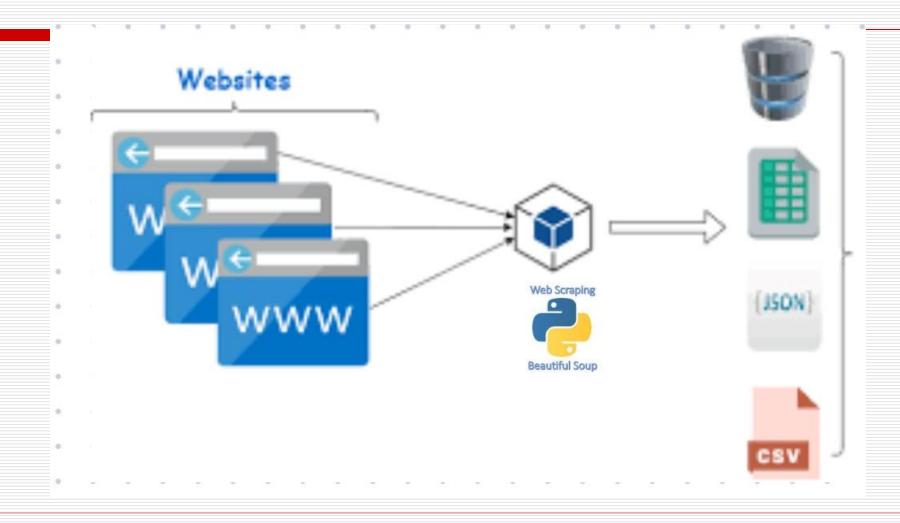
List of modules

- 1. Data collection
- 2. Data pre-processing
- 3. BRET model
- 4. Frontend Module
- 5. Integrating backend and frontend

Data collection

- □ 1. Install Required Libraries: Install `beautifulsoup4`, `requests`, and `pandas` using pip.
- □ 2. Identify Target Website: Select a website hosting the education schemes data.
- □ 3. Fetch the Web Page: Use the `requests` library to fetch the webpage content.
- 4. Parse the Web Page: Parse the fetched HTML content with BeautifulSoup.
- □ 5. Identify Data Elements: Inspect the HTML structure to locate relevant tags and classes.
- ☐ 6. Extract Relevant Data: Use BeautifulSoup methods to extract scheme details like name, eligibility, and benefits.
- ☐ 7. Clean and Preprocess Data: Clean the data by removing unnecessary characters or tags.

8. Convert Data to a Structured Format: Organize the data into a Pandas DataFrame.
9. Save the Dataset: Save the structured data as a CSV or Excel file.
10. Handle Pagination (if applicable): Automate scraping for multiple pages if the data
spans several pages.



4	A	В	С		D	E	F	G	Н	I	J	K
1	Scheme Name	Launch Date	Objective	Eligibility	Criteria		Source					
2	Chief Minister's Breakfast Scheme	September 15, 2022	Provides free breakfas	Students of g	overnment schoo All enrol	ed students are eligible withou	t (TNPSC Current Affairs)ps://www.ti	npscthervupe	ettagam.co	m/articles-	detail/rece	nt-tamil
3	Pudhumai Penn Scheme	September 5, 2022	Supports female stude	Girls studying	in government schools fro	om Class 6 to 12 who join any h	Recent Government Schemes.					
4	Illam Thedi Kalvi	October 27, 2021	Focuses on after-scho	Students fror	n grades 1 to 8 in Available	to all students needing extra e	Official Scheme Details.					
5	Makkalai Thedi Kalvi	N���(TNPSC Current Af	fairs) Door-to-door edu	Rural student	ts who missed cla Focus or	children in rural areas with lov	Government Scheme Overview.					
6	Chief Minister���s Merit Scholarship	2021	Scholarships for top-pe	Government	school students f High per	ormance in board exams.	(TNPSC Current Affairs)(https://ww	w.tnpscther	/upettagan	.com/artic	les-detail/re	ecent-ta
7	Post-Matric Scholarship for SC/ST/OBC	Ongoing	Provides financial assis	stance for SC,	ST, and OBC students to p	ursue studies beyond Class						
8	Eligibility: Students belonging to S���(TNP	PSC Current Affairs)iteria**:	Must be enrolled in red	Students belo	onging to S���(TNPSC (Current Affairs)iteria**: Must be	Social Welfare Programs.					
9	Mudhalvarin Mugavari Scheme	Feb-24	Integrated helpline for	Open to all re	esidents of Tamil No speci	fic restrictions; available to any	Official Government Portal.					
10	(TNPSC Current Affairs)Launch Date*: 2011, c	ongoing	Provides free laptops t	Higher secon	dary students in ¿Students	must be in Classes 11 and						
11	Source: Education Schemes.						Education Schemes.					
		2001, ongoing					(TNPSC Current Affairs)t](https://w	ww.tnpscthe	rvupettaga	m.com/art	icles-detail/	recent-
13	Girl Child Protection Scheme	2012	Financial assistance to	Families with	one���(TNPSC Curren	t Affairs)teria**: Family must n	Tamil Nadu Welfare Schemes.					
14	Educational Assistance for Children of Differ	1982, ongoing	Offers nutritious meals	All governme	nt school studen Enrolled	in government schools from Cl	E[Social Welfare Programs](https://v	www.tnpscth	ervupettag	gam.com/a	rticles-deta	il/recent
15	Source: Education Department Portal.						Education Department Portal.					
16	Vocational Education Scheme	2022	Integrates vocational t	Students in h	igher secondary s Must be	enrolled (TNPSC Current Affairs	Vocational Education Initiatives.					
_	Educational Loan Scheme for Overseas Studi		Provides financial supp	ort for studer	nts pursuing highe Must me	et financial need and academic	Education Finance Programs.					
18	Thalikku���(TNPSC Current Affairs)Schen	2010	Provides financial supp	_	-							
	Eligibility: Girls from economically disadvanta				-		t [S���(TNPSC Current Affairs)ttp:	s://www.tnp	scthervupe	ettagam.co	m/articles-c	letail/re
20	E.V.R. Maniammaiyar Memorial Free Educat	2012	Offers free education	Students fron	n backward class Must bel	o���(TNPSC Current Affairs	Tamil Nadu Welfare Programs.					
	National Means-cum-Merit Scholarship (NM	2008	Financial aid for disadv	Students in g	overnment schools from C	lasses 8 to						
22	Criteria: Based on merit and financial need.					merit and financial need.	Education Programs.					
	Free Uniform Scheme		Provides free uniforms	Government	school students from Clas							
24	Criteria: All students in government schools q					nts in government schools qual						
25	Periyar EVR Nagammai Free Education Scher	2020	Free education for eco	Students fror	n backward class Must me	et the economic need threshol	Government Schemes.					
	Chief Minister's Uzhavar Scholarship	2022	Scholarships for childre	Farmers' chil	dren pursuing edt Must be	enrolled in educational institut	i Social Welfare Programs.					
27												
	dataset +					: (·					Þ

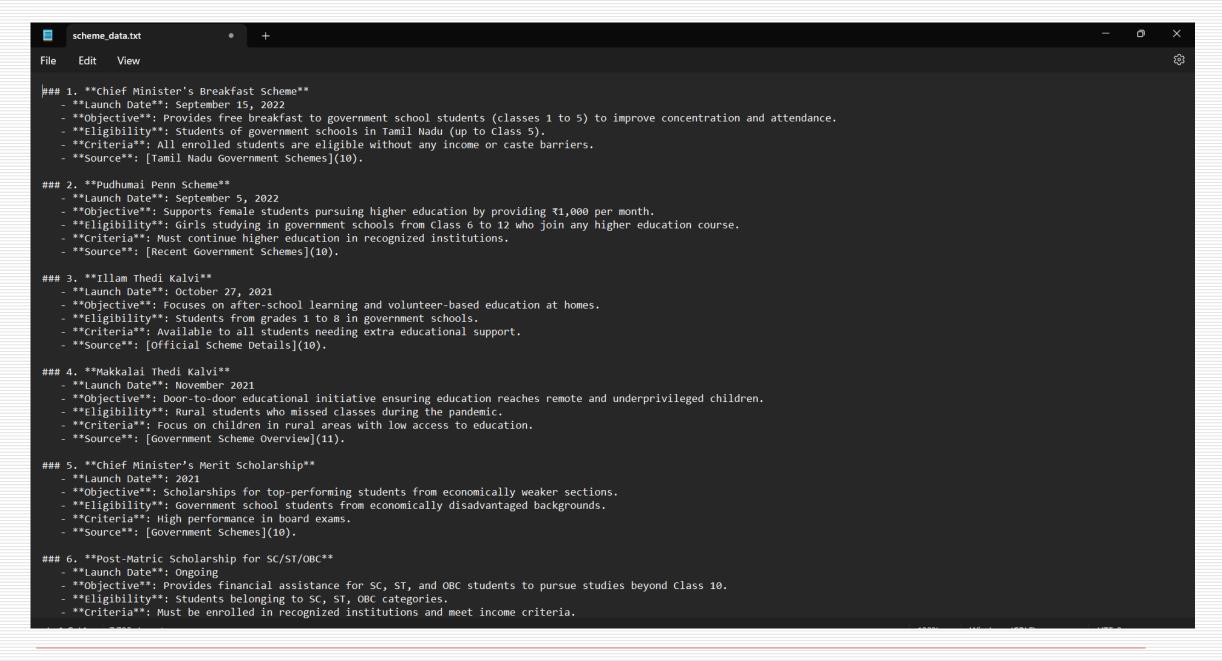
Third Review

Accessibility: Unavailable

 \blacksquare

Data pre-processing

- 1. Lowercasing
- 2. Removing Punctuation
- 3. Tokenization
- 4. Removing Stop Words
- 5. Lemmatization / Stemming
- 6. Removing Numbers
- 7. Removing Extra Whitespaces
- 8. Handling Special Characters
- 9. Spelling Correction
- 10. Handling Negations



BERT Model

- 1. **Tokenization**: The input question and context are broken into smaller units called tokens.
- 2. Input Representation: Each token is converted into a dense vector or embedding for numerical processing.
- **3. Positional Encoding**: Positional embeddings are added to the tokens to help BERT understand the order of words.
- **4. BERT Architecture**: The model consists of multiple Transformer layers that process tokens and build context-aware representations.
- 5. Self-Attention Mechanism: Each token pays attention to other tokens in the sequence to capture word relationships in both directions.
- **6. Contextual Encoding**: BERT layers refine token meanings based on the context, enhancing understanding as tokens pass through the model.
- **7. Question Understanding**: The [CLS] token at the start of the sequence helps BERT comprehend the question.

BERT Model

- **8. Start and End Token Prediction**: BERT predicts the tokens marking the beginning and end of the answer span.
- **9. Answer Extraction**: The model selects the most likely answer from the context based on start and end token predictions.
- **10. Final Output**: BERT returns the extracted answer from the context as the response to the question.

Tokenization:

1. Basic Tokenization

Split the text into words or subwords using BERT's tokenizer, which supports **WordPiece** or **Subword Tokenization**.

Formula:

Ttokens=BERT_Tokenizer(T)

2. Add Special TokensAdd special tokens to mark the beginning and end of the input sequence:

[CLS]: Indicates the start of a sequence (used for classification or context embedding).

[SEP]: Separates different sentences or marks the end of a single sentence.

Formula:

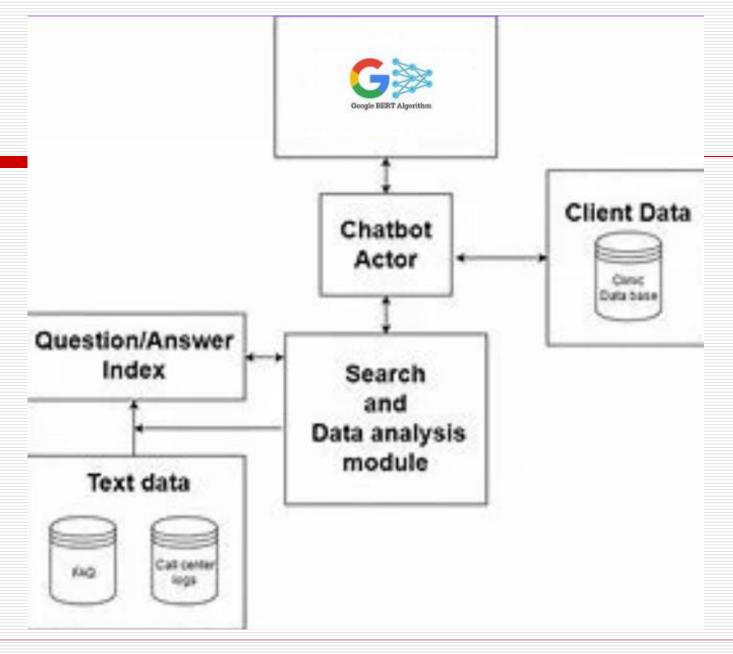
Ttokens=[[CLS]]+Ttokens+[[SEP]]

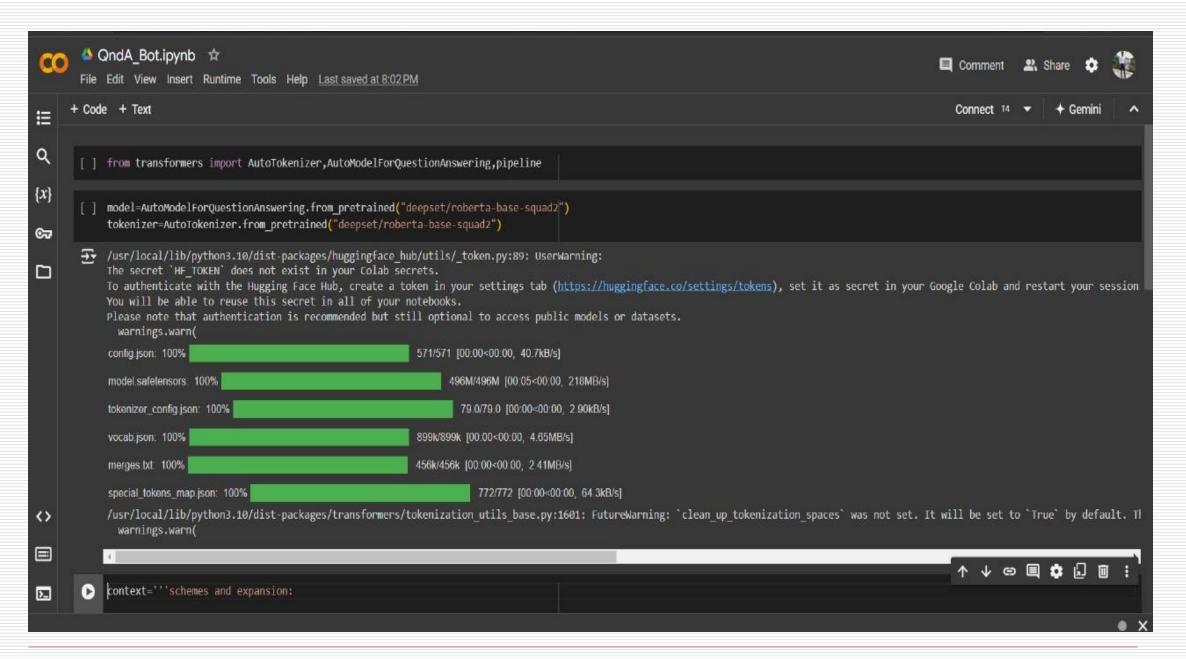
3. Convert Tokens to IDs

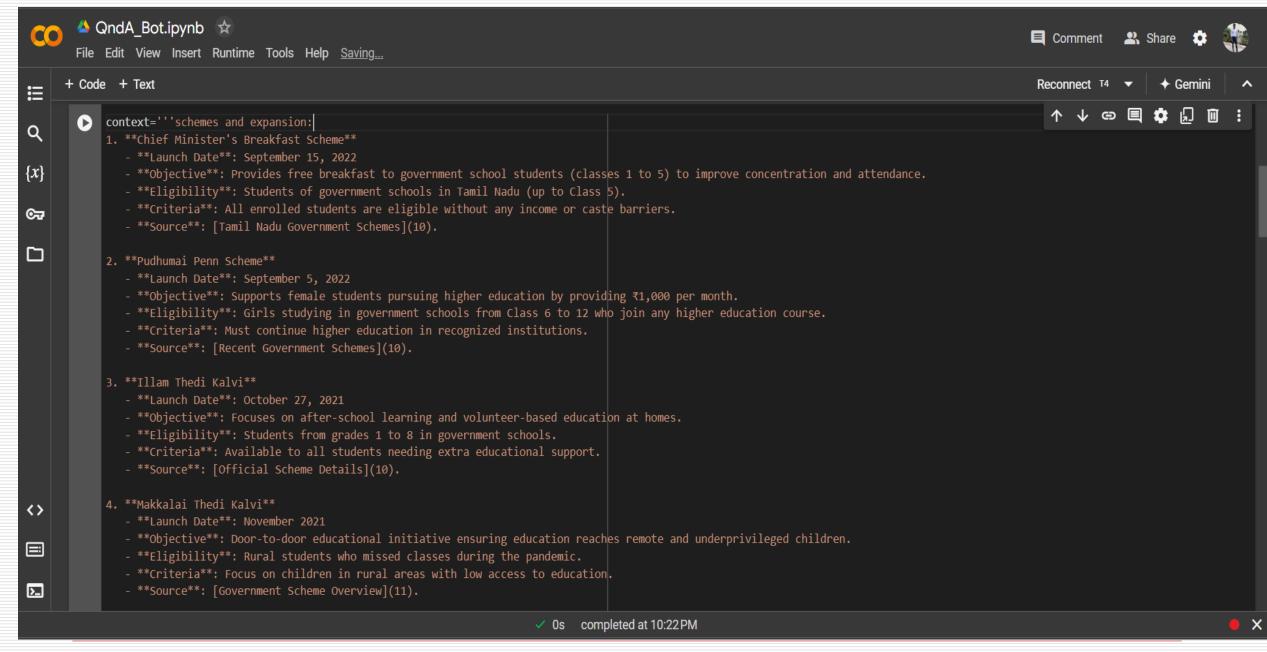
Convert each token into its corresponding ID from the BERT vocabulary.

Formula:

Tids=Token_to_ID(Ttokens)







```
18. Free Uniform Scheme
Launch Date: 2011
Objective: Provides free uniforms and footwear to reduce economic barriers for students.
Eligibility: Government school students from Class 1 to 8.
Criteria: All students in government schools qualify.
Source: Social Assistance Programs.
```

```
[13] ask=pipeline('question-answering',model=model,tokenizer=tokenizer)
result=ask(question="which scheme provide uniform and footwear?",context=context)

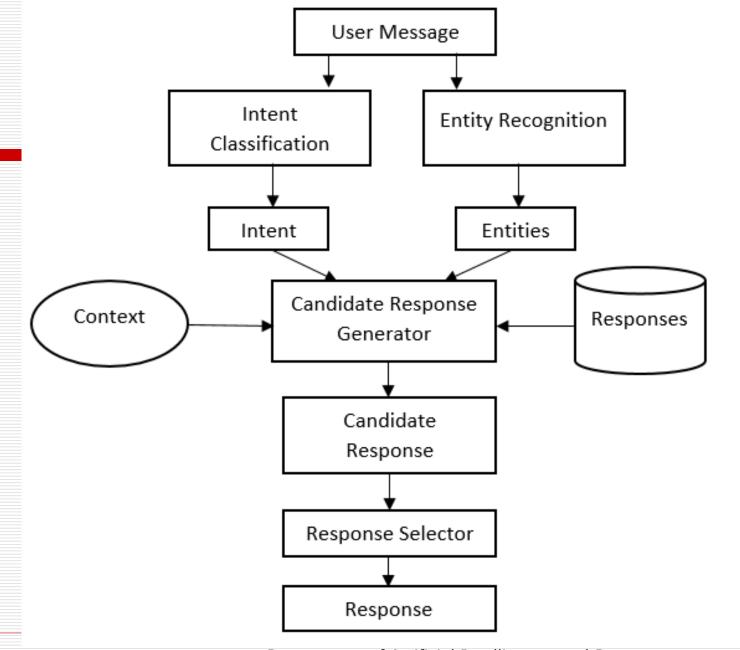
☐ Hardware accelerator e.g. GPU is available in the environment, but no 'device' argument is passed to the 'Pipeline' object. Model will be on CPU.

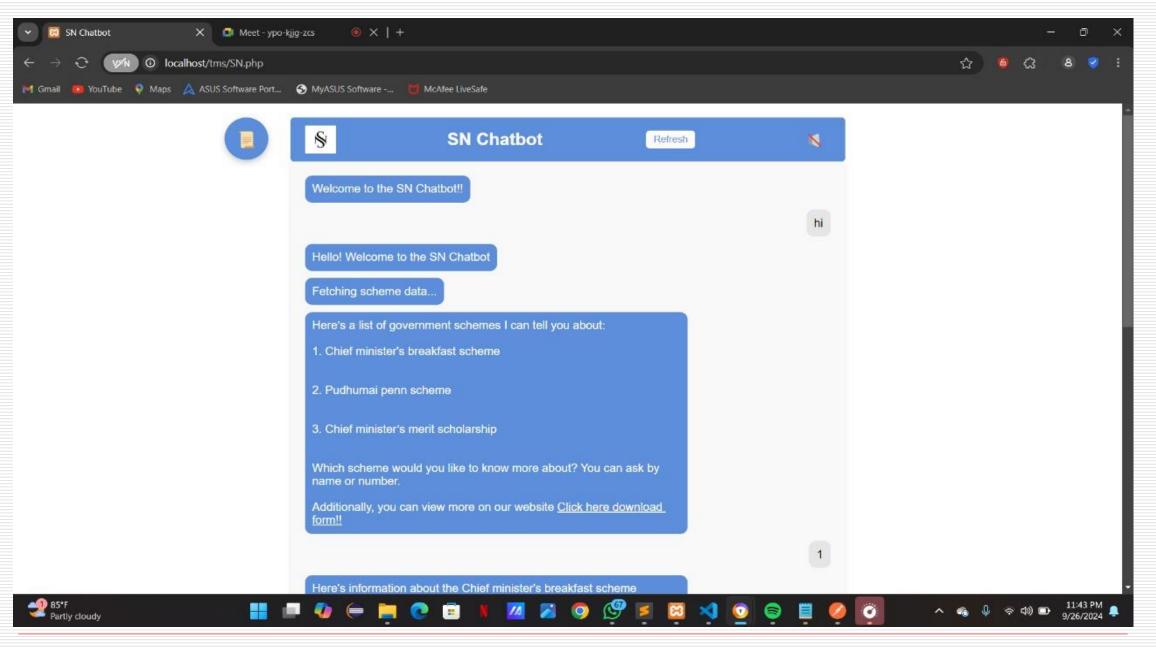
[14]
result['answer']

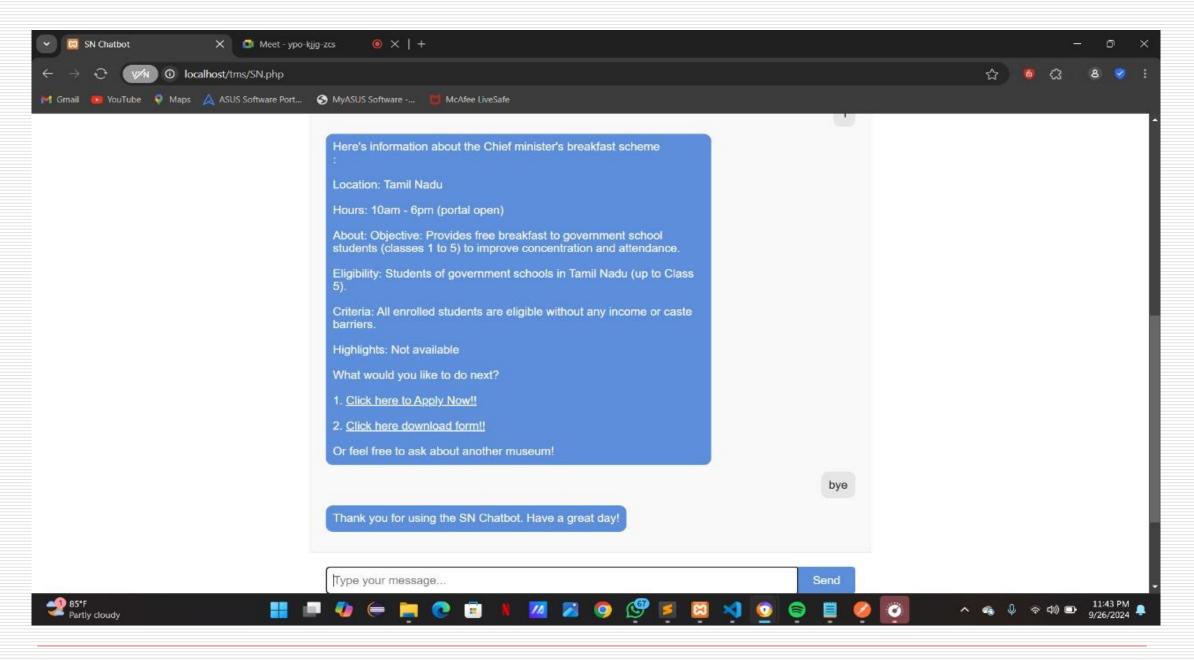
☐ 'Free Uniform Scheme'
```

Frontend Design

- 1. Plan Your Chatbot Design: Define chatbot features, layout, and functionality.
- 2. Set Up the Project Directory: Create a structured directory for frontend and backend files.
- 3. Create the Frontend (HTML): Design the chatbot interface using HTML.
- 4. Style the Frontend (CSS): Use CSS to style the chatbot layout and make it visually appealing.
- 5. Add Client-Side Logic (JavaScript): Write scripts to handle user input and display bot responses.
- 6. Set Up the Node.js Server: Build a backend server using Node.js and Express.
- 7. Install Dependencies: Install required packages like `express` and `body-parser`.
- 8. Test the Website Locally: Run the server and verify the chatbot functionality in the browser.
- 9. Improve Chatbot Logic: Integrate AI models or APIs to enhance chatbot responses.
- 10. Deploy the Website: Host the website on platforms like Heroku, Vercel, or AWS.







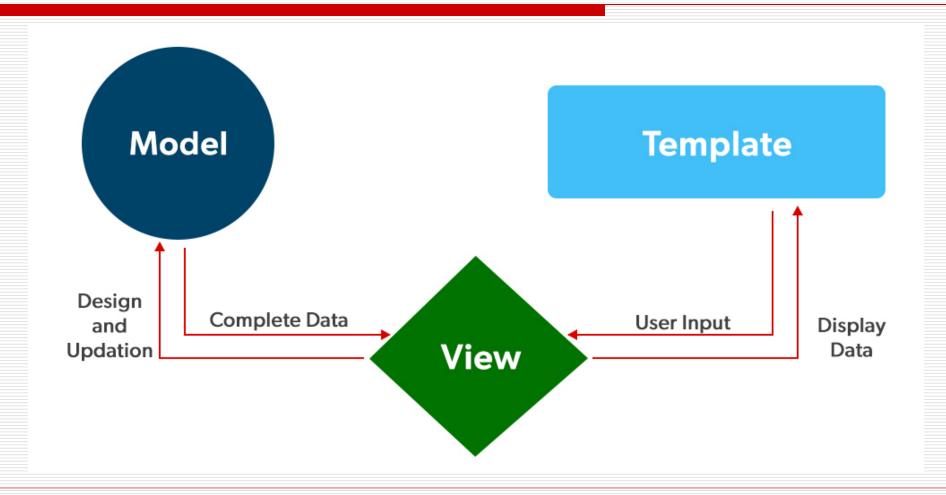
Integrating backend and frontend

- **1.Install Django**: Set up Django by installing it via pip (pip install django) and create a new project using django-admin startproject.
- **2.Create an App**: Inside the project, create an app using python manage.py startapp [app_name] where the backend logic will reside.
- **3.Define Models**: In the app's models.py, define the data models that represent the database schema.
- **4.Migrate Database**: Run python manage.py makemigrations and python manage.py migrate to create the necessary database tables based on models.
- **5.Create Views**: In views.py, define the logic that handles requests and processes backend operations like handling forms, interacting with models, and returning responses.
- **6.Set up URLs**: Map views to URLs by configuring urls.py to route frontend requests to the correct backend views.
- **7.Create HTML Templates**: Design frontend HTML templates that will render the data returned from Django views using Django's templating engine.

Integrating backend and frontend

- **8.Connect Views to Templates**: In views, return HttpResponse or render() functions that combine the HTML templates with backend data and pass it to the user.
- **9.Handle Static Files**: Serve CSS, JavaScript, and images by configuring STATIC_URL in settings.py and creating a static directory for frontend assets.
- **10.Run Server**: Start the Django development server with python manage.py runserver to connect the backend and frontend, allowing user interaction via the web browser.

DFD



Comparisons & Analysis

S.No.	Title Of The Paper	Methodology Used	Limitation(s)	Advantage(s)	Accuracy (In %)
1	A Conversational Agent for Learning: Development of an Educational Chatbot Using BERT	BERT, Intent, Entity Recognition, Response Generation.	Limited dataset size may affect the model's generalization. The complexity of user queries is not fully handled by the current model.	Improved understanding of user intent due to BERT's contextual embeddings. Enhanced user engagement through interactive conversations.	80
2	Enhancing Student Learning through AI: An Intelligent Chatbot Using LLMs	LLM	Dependency on training data; biases in the dataset can reflect in responses.	High-quality response generation due to LLM capabilities. Flexibility in handling a wide range of educational topics.	85
3	Integrating RAG for Enhanced Educational Chatbots: A Case Study	RAG(Retrieval augmented generation)	The performance may vary based on the quality of retrieved documents. Increased complexity in implementation and maintenance.	Combines real-time data retrieval with generative responses, improving accuracy Can handle more diverse queries with factual support.	90
4	Building a Context- Aware Educational Chatbot Using BERT and LLM	LLM(Large language model), BERT(Bidirectional Encoder Representations from Transformers)	High computational requirements for real-time processing. Limited understanding of domain-specific jargon.	Context-aware conversations leading to improved user satisfaction. Ability to adapt responses based on user interactions.	78
5	High Accuracy Phishing Detection Based on Convolutional Neural Networks	Convolutional Neural Network, Random Forest, Naive Bayes, SVM, J48.	As more features are extracted in this system, the model becomes more complex.	The model achieved high accuracy and had a faster implementation.	97.3

References.

- □ "LoRA: Low-Rank Adaptation of Large Language Models "Jacob Devlin, Ming-Wei Chang, IEEE Human Language Technologies, vol 11, Date: June 2022, Pages:133-144.
- "Towards Intelligent Public Services: An AI-based Chatbot for the Swedish Social Insurance Agency", Anders Henriksson, IEEE Access, vol: 7, Date: 2023, Pages: 56-65.
- □ "LLM for SoC Security: A Paradigm Shift", Gianluca Klopfenstein, IEEE Transactions on data,vol:10, Pages: 32-43.
- □ "LLM for SoC Security: A Paradigm Shift", Yingqiang Ge, Wenyue Hua, IEEE Access, Date: 2023, Pages:11-20

Conclusion

In conclusion, developing a chatbot for education schemes using BERT, Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs) offers a powerful framework for providing accurate, contextually relevant information to users. By leveraging BERT's capabilities for understanding user queries and generating embeddings, combined with RAG's ability to retrieve pertinent documents and LLMs' proficiency in generating coherent responses, the chatbot can effectively address diverse inquiries related to educational schemes. This integrated approach not only enhances user experience through natural and engaging interactions but also ensures that the information provided is up-to-date and precise, ultimately empowering citizens with valuable knowledge about government education initiatives.

Thank You