A Concurrent Intelligent Natural Language Understanding Model for an Automated Inquiry System

Gokul Sunilkumar, Srihari S, Steven Frederick Gilbert, Chitrakala S
Department of Computer Science and Engineering
College of Engineering Guindy, Anna University, Chennai, India

2022 IEEE World Conference on
Applied Intelligence and Computing (AIC 2022)

INTRODUCTION

- For any modern dialogue system, irrespective of it being conversational or not, understanding the language spoken and extracting information out of it before its formal analysis is a key component and that is where the domain of Spoken Language Understanding (SLU) begins.

- All the currently used models have some common issues like **poor semantic understanding** of human-like inputs and **less robust to ambiguities in speech**.

- We are proposing an approach wherein a multi-task model post following an **intelligent encoding strategy** attends to the important parts of the text to come up with the final classes of outputs that they belong to.

- Each class decides its **own output template** whose slots are then filled by **scraping information from associated websites** that host the required content.

- In all, we are targeting a model that takes in a speech query as input and understands its meaning to come up with a speech response to it in an efficient, and more importantly, accurate manner.

OBJECTIVES

To design and develop a multi-task model that uses a global representation of the input query for parallelly performing the first three tasks followed by the added application module:

- Dialogue Act Classification
- Intent Detection
- Slot Filling
- Query Response Retrieval

SAMPLE INPUT OUTPUT

INPUT	INTERMEDIARY OUTPUTS			FINAL OUTPUT
Input Query	Dialogue Act	Intent	Slot	Query Response
How many flights does Indigo have in business class?	Question	quantity	1) Indigo : B-airline 2) business:B-class_type 3) class:I-class_type	The required number of flights are 125.
Show me all the flights from Mumbai to New Delhi.	Command	flight	1) Mumbai :B-fromloc.city_name 2) New : B-toloc.city_name 3)Delhi : I-toloc.city_name	The required flights are: Jet Blue Airlines – JB2X34, Indigo Airlines - LF2Z43
I'm looking for ground transportation in Dallas.	Statement	ground_service	1) dallas:B-city_name	Cabs are available at Dallas

LITERATURE SURVEY

REFERENCE PAPER	METHODOLOGY	ISSUES
A Deep Multi-task Model for Dialogue Act Classification, Intent Detection and Slot Filling – Base Paper Firdaus, Mauajama, Hitesh Golchha, Asif Ekbal, and Pushpak Bhattacharyya. "A deep multi-task model for dialogue act classification, intent detection and slot filling." *Cognitive Computation* 13, Springer, no. 3 (2020)	**Multi task** Deep Learning Approach	Less incorporation of **semantic information**
AISE: Attending to Intent and Slots Explicitly for better spoken language understanding Yang, Peng, Dong Ji, Chengming Ai, and Bing Li. "AISE: Attending to Intent and Slots Explicitly for better spoken language understanding." *Knowledge-Based Systems* 211, Elsevier (2021)	Position-aware Multihead Masked Attention (**PMMAtt**)	Constrained by **reduced interactions** between slot and intents
Multi-turn intent determination and slot filling with neural networks and regular expressions Abro, Waheed Ahmed, Guilin Qi, Zafar Ali, Yansong Feng, and Muhammad Aamir. "Multi-turn intent determination and slot filling with neural networks and regular expressions." *Knowledge-Based Systems* 208, Elsevier (2020)	**Multi turn** Approach using Regular Expressions	Doesn't encourage **dialogue state tracking**

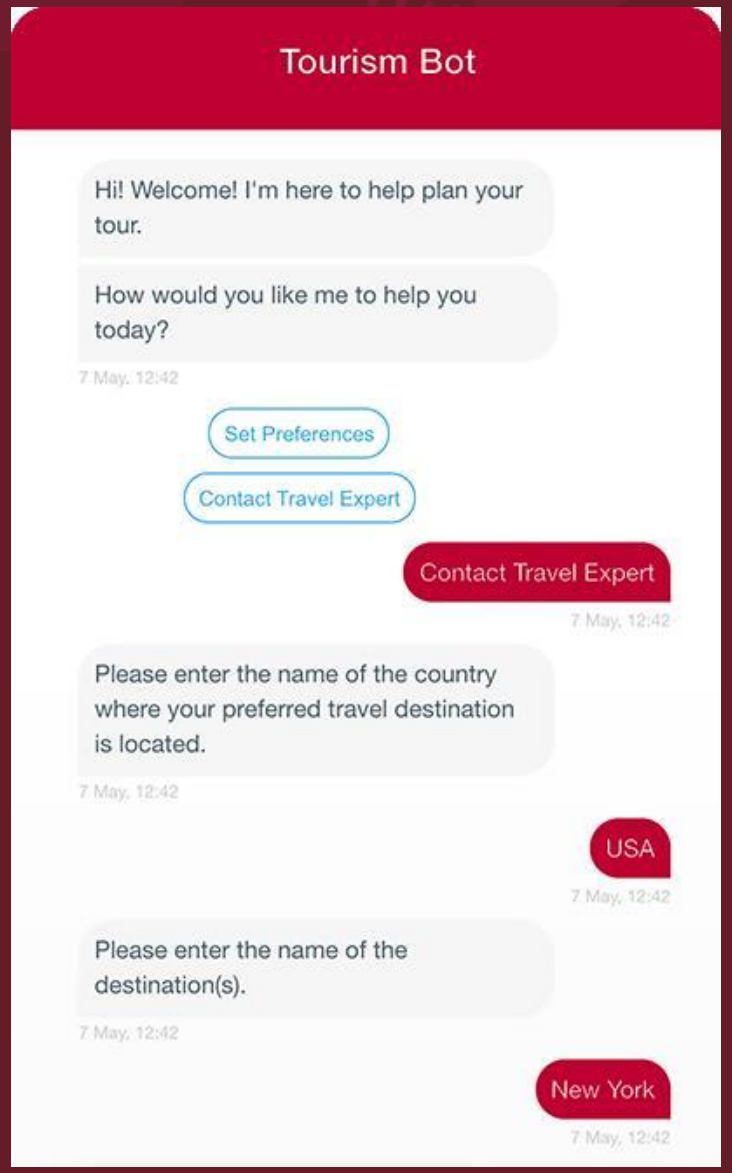
LITERATURE SURVEY

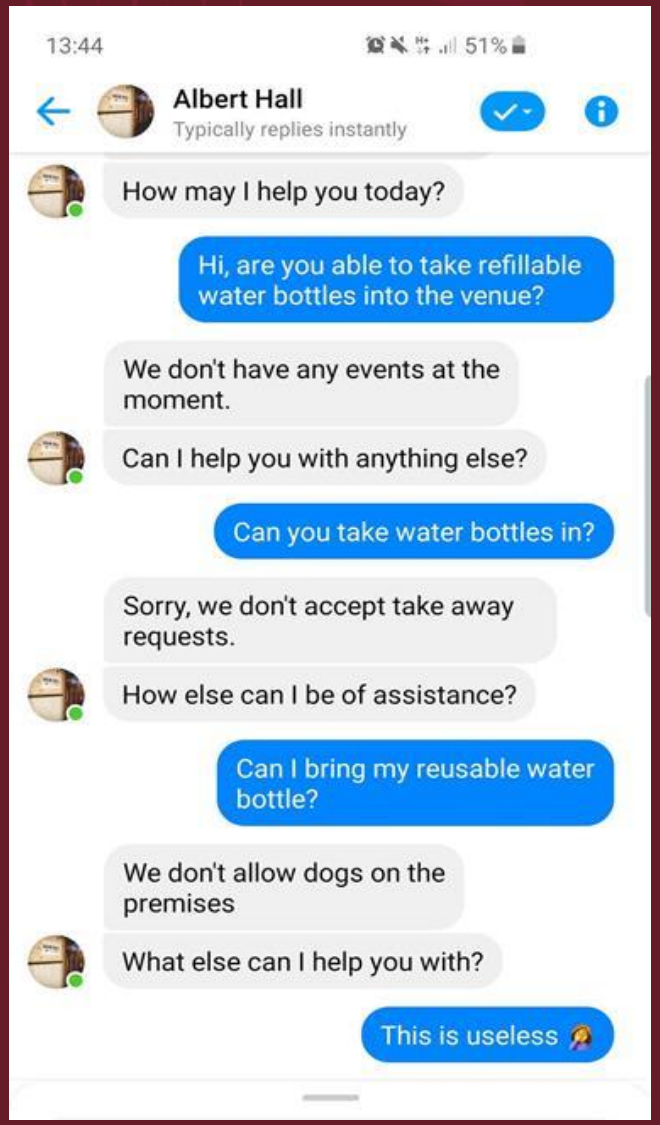
REFERENCE PAPER	METHODOLOGY	ISSUES
Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction Ni, Pin, Yuming Li, Gangmin Li, and Victor Chang. "Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction." *Neural Computing & Applications* 32, Springer, no. 20 (2020)	**Hybrid Based Joint** Model	Unable to perform well on **low data resource**
A Multi-Task Hierarchical Approach for Intent Detection and Slot Filling Firdaus, Mauajama, Ankit Kumar, Asif Ekbal, and Pushpak Bhattacharyya. "A multi-task hierarchical approach for intent detection and slot filling." *Knowledge-Based Systems* 183, Elsevier (2019)	**Hierarchical** Joint Dual task	**Lack of necessary semantic** information which can further draw out more meaning for better context.
Dual Learning for Semi-Supervised Natural Language Understanding Zhu, Su, Ruisheng Cao, and Kai Yu. "Dual learning for semi-supervised natural language understanding." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020)	Two Agent Semi-Supervised Learning	Limited by a single reward system

SUMMARY OF ISSUES

- A **lack of semantic information** can pose to be an issue when it comes to the system ability to draw out information to understand contextual information.
- **No interactions** between the intent and slot filling will yield a lower understanding of the input query which can be solved with a multi-task style system.
- Dialogue state tracking is essential to contextual understanding, without which the system's understanding of the input query is incomplete.
- Inability to **deal with ambiguities** in input utterance and certain misspellings

SUMMARY OF ISSUES

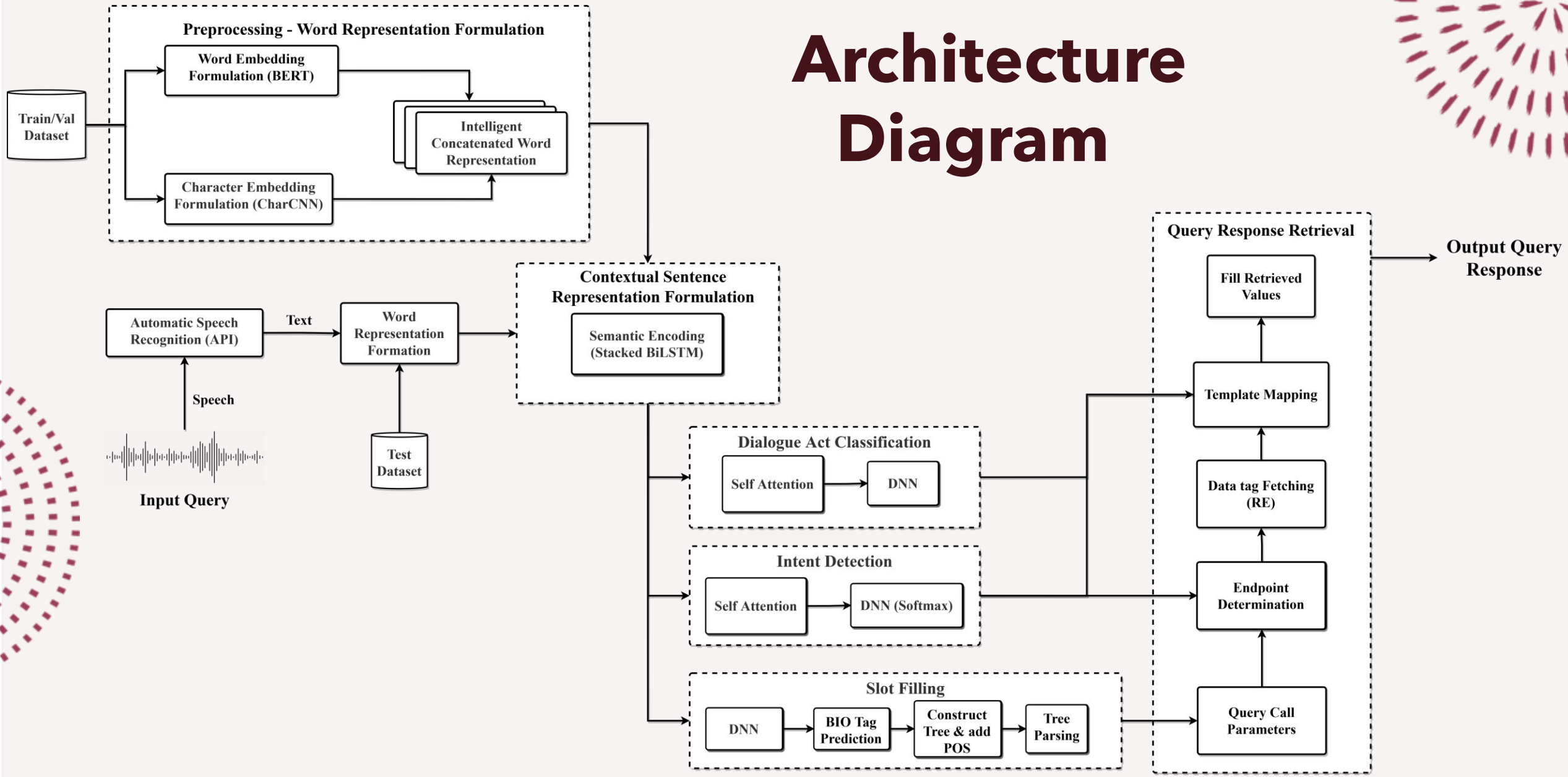




PROPOSED SYSTEM - CIDIS

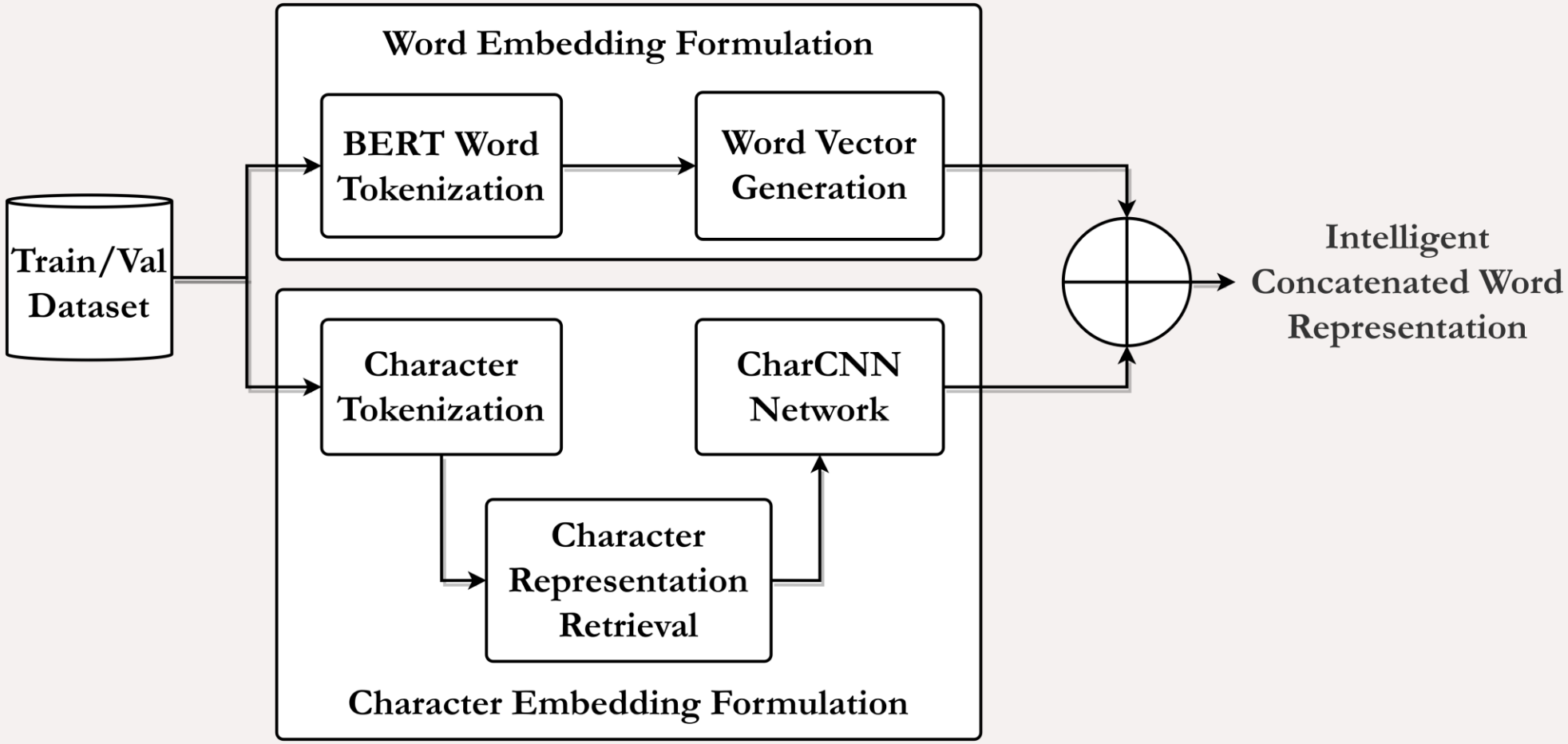
- In order to combat the previously stated issues, a **Concurrent Intelligent Model for Dialogue Act Classification, Intent Detection and Slot Filling (CIDIS)** is proposed.
- It employs an **Intelligent encoding strategy** which is capable of dealing with misspelled words and ambiguities in the text.
- This robust representation is fed to a **multi task model** to accomplish the tasks of Dialogue Act Classification, Intent detection and Slot filling.
- These predicted details are used in **forming a cogent response** upon fetching the required details using a scraping mechanism.

Architecture Diagram



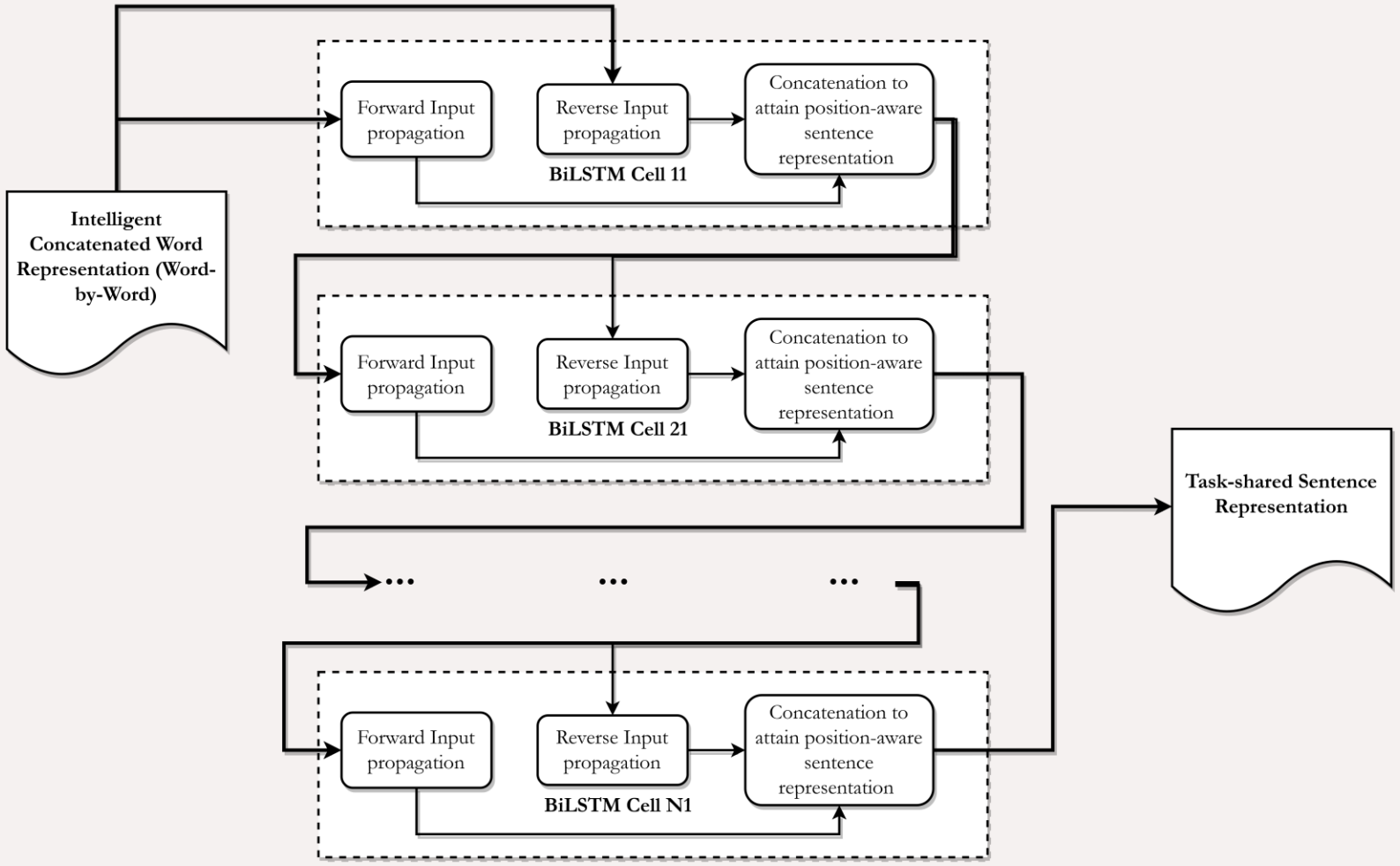
Module 1: Preprocessing - Word Representation Formulation

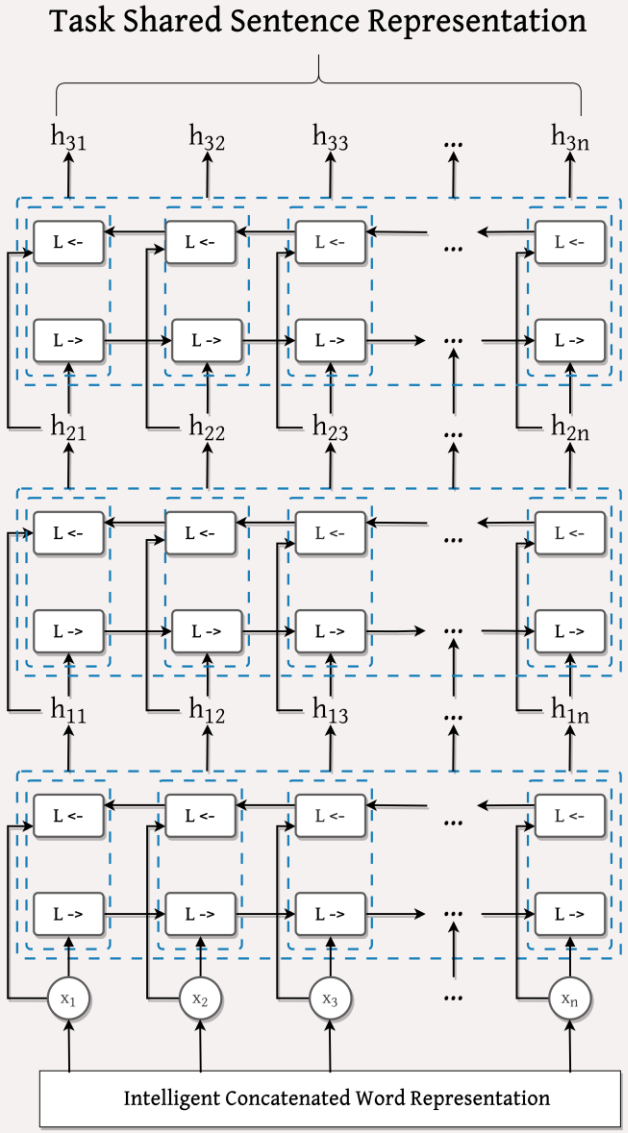
- **Input :** Text from user's utterance
- **Output :** Concatenated Word Representation



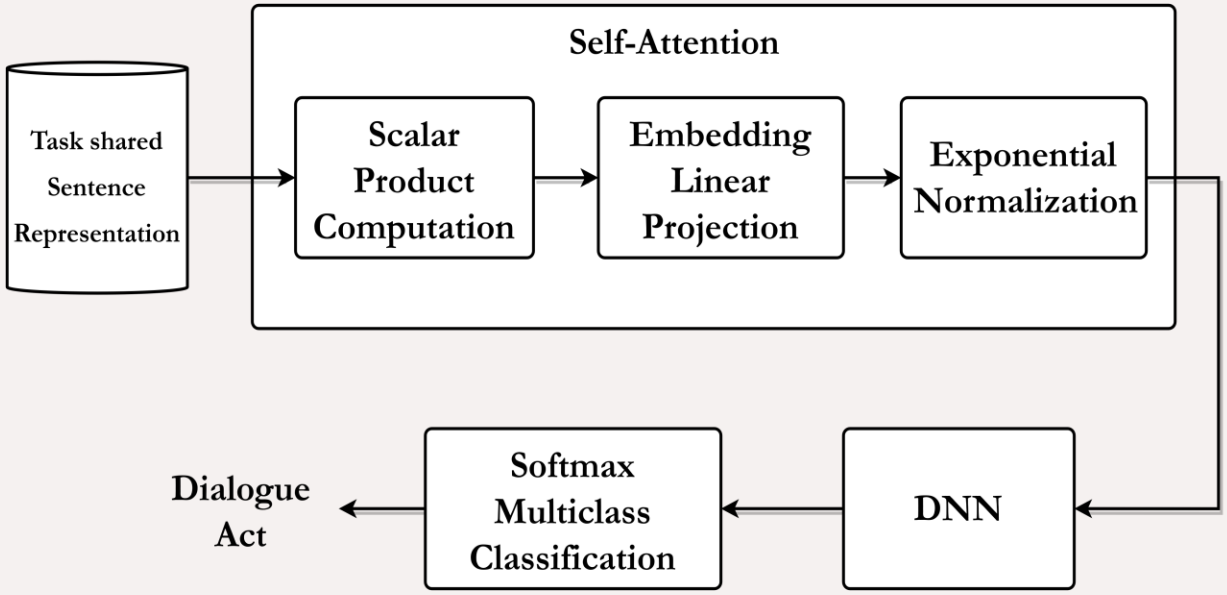
Module 2: Contextual Sentence Representation Formulation

- **Input :** Intelligent Concatenated Word Representation
- **Output :** Task Shared Sentence Representation





Module 3: Dialogue Act Classification

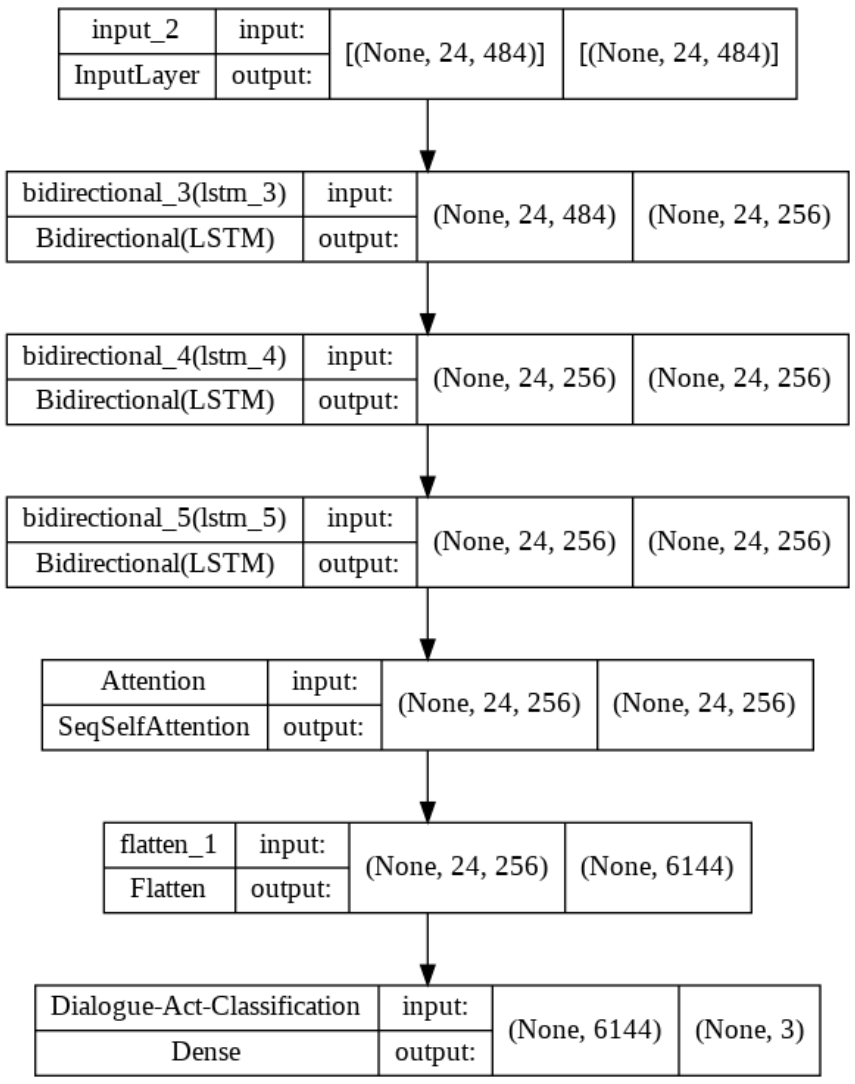


Input : Task Shared Sentence Representation

Output : Dialogue act of the utterance

```
def classifyDialogueActs(sentenceRepresentation):

  sentenceRepresentation = selfAttention(sentenceRepresentation)
  extractedFeatures = DNN(sentenceRepresentation)
  extractedFeatures = affineTransformation(extractedFeatures)
  predictedDialogueAct = Softmax(extractedFeatures)

  return predictedDialogueAct
```

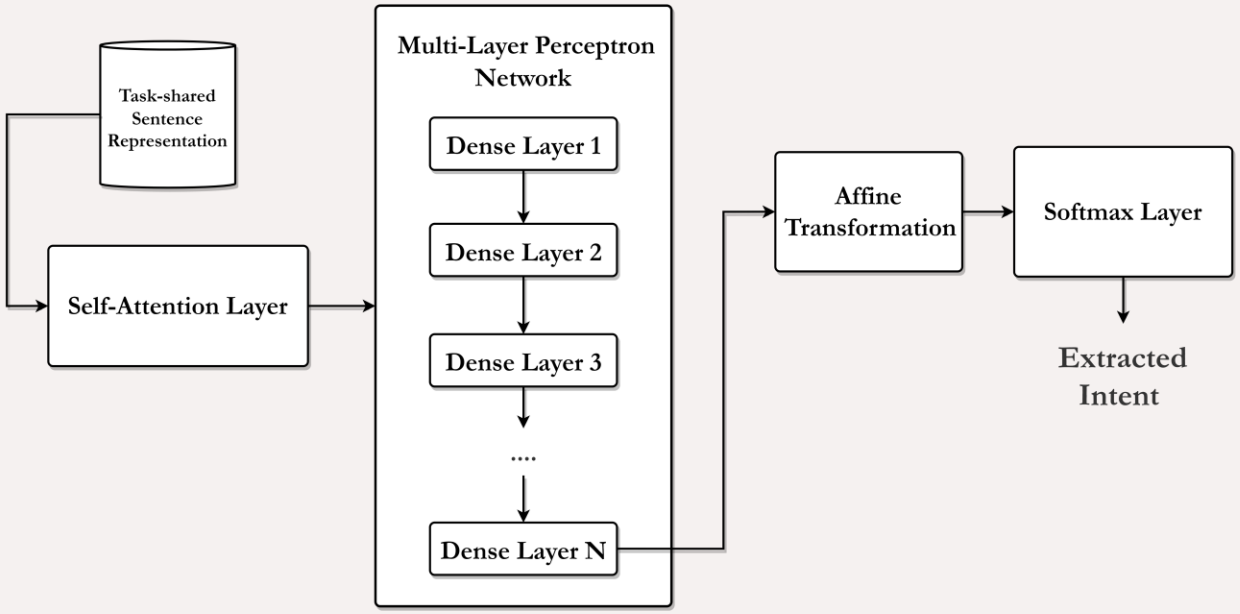
DIALOGUE ACT CLASSIFICATION



Input : Task Shared Sentence Representation
Output : Dialogue act of the utterance

```
Input Query  list flights from newark to houston
Expected DAC:  command
['Predicted DAC: command']
**********************************************************************
Input Query  please list flights from milwaukee to philadelphia
Expected DAC:  command
['Predicted DAC: command']
**********************************************************************
Input Query  what is the lowest fare for a flight from washington dc to boston
Expected DAC:  question
['Predicted DAC: question']
**********************************************************************
Input Query  list airlines flying from seattle to salt lake city
Expected DAC:  command
['Predicted DAC: command']
**********************************************************************
```

Module 4: Intent Detection

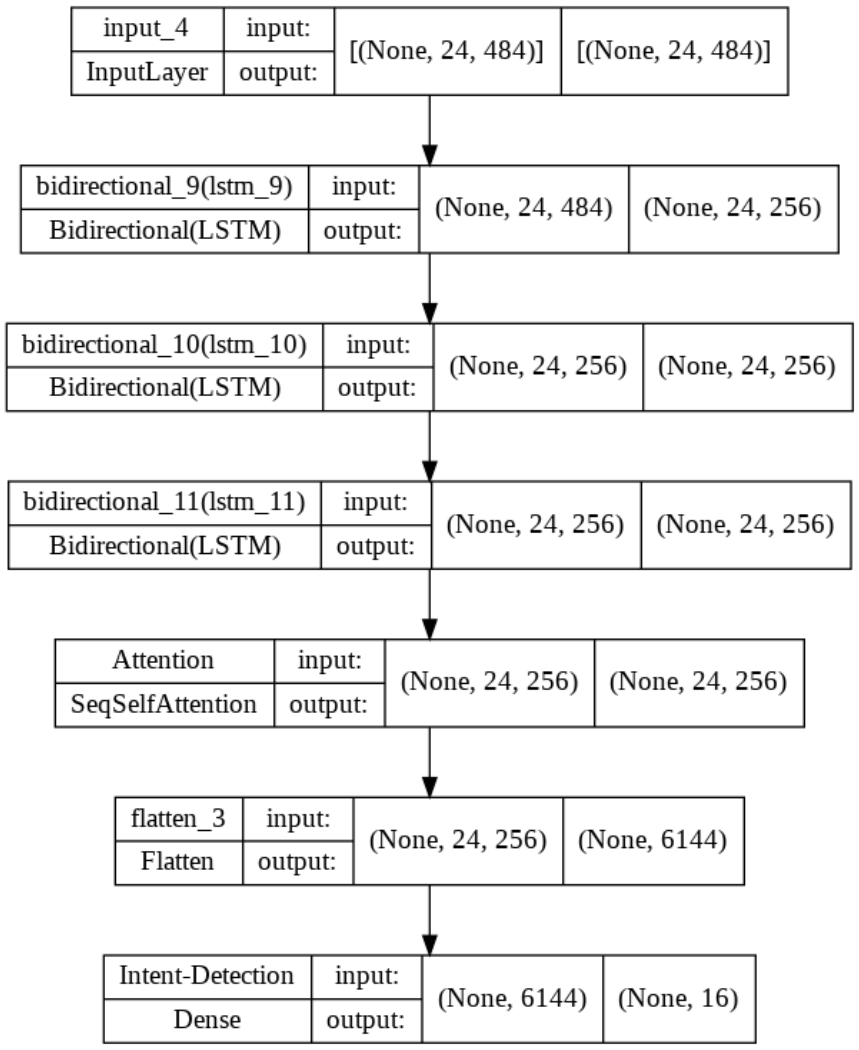


Input : Task Shared Sentence Representation

Output : Intent of the utterance

```
def detectIntents(sentenceRepresentation):

  sentenceRepresentation = selfAttention(sentenceRepresentation)
  extractedFeatures = DNN(sentenceRepresentation)
  extractedFeatures = affineTransformation(extractedFeatures)
  predictedIntent = Softmax(extractedFeatures)

  return predictedIntent
```

INTENT DETECTION

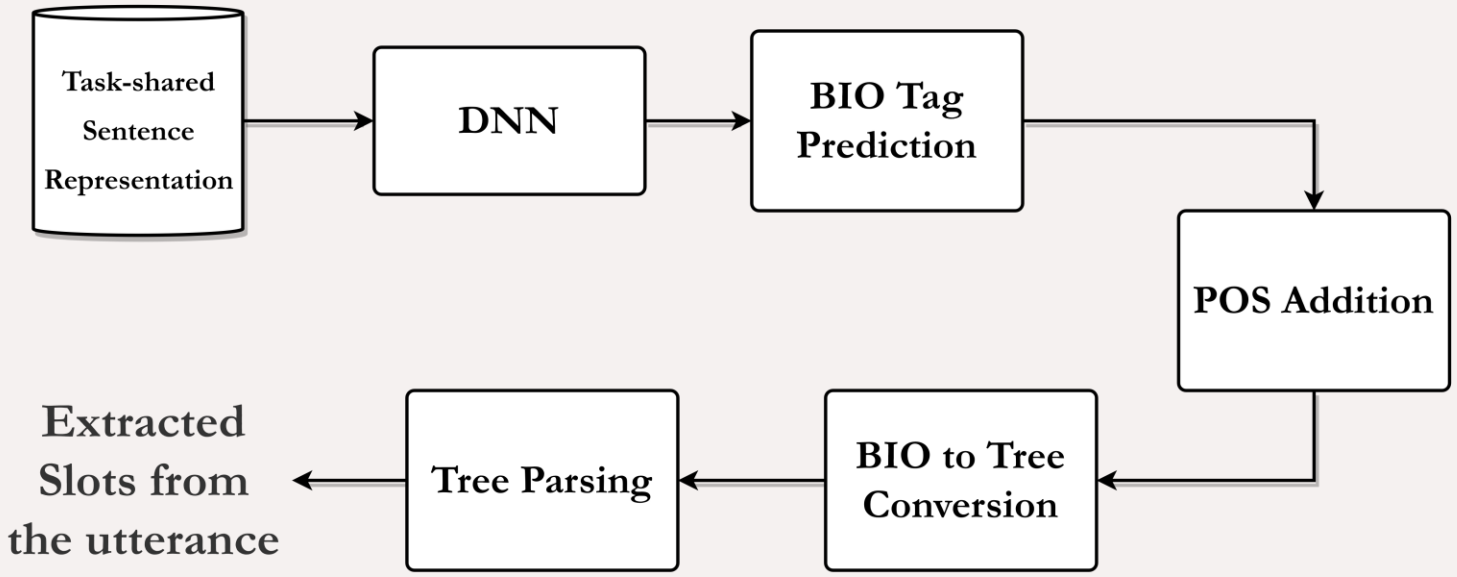


Input : Task Shared Sentence Representation

Output : Intent of the utterance

```
Input Query  what does hp stand for
Expected Intent:  atis_abbreviation
['Predicted Intent: atis_abbreviation']
**********************************************************************
Input Query  how far is toronto international from downtown
Expected Intent:  atis_distance
['Predicted Intent: atis_distance']
**********************************************************************
Input Query  what airline is dl
Expected Intent:  atis_airline
['Predicted Intent: atis_airline']
**********************************************************************
Input Query  list a flight on american airlines from toronto to san diego
Expected Intent:  atis_flight
['Predicted Intent: atis_flight']
**********************************************************************
```

Module 5: Slot Filling

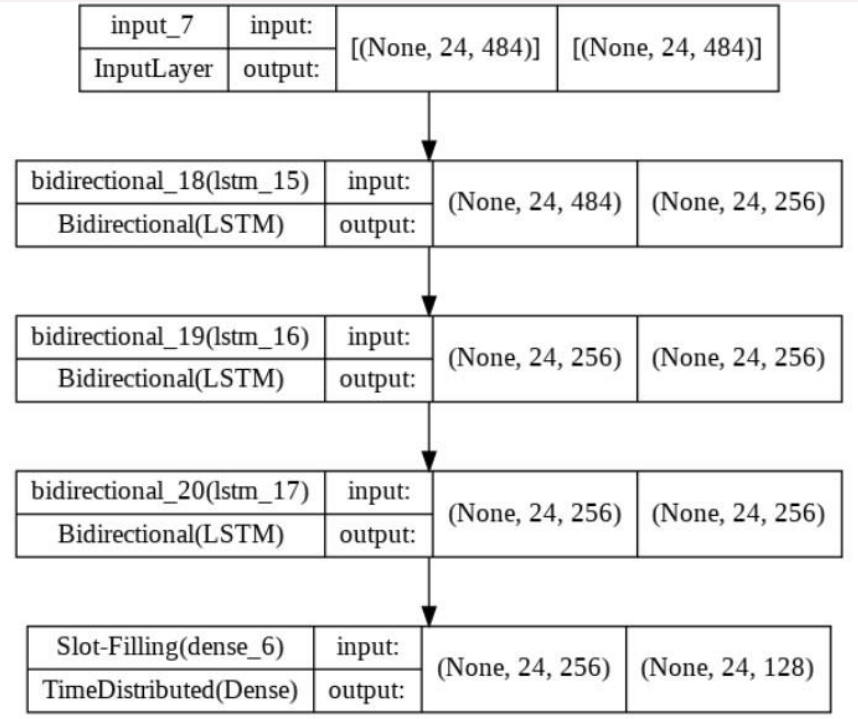


Input : Task Shared Sentence Representation

Output : Extracted slots from the utterance

```
def fillSlots(sentenceRepresentation):
    extractedFeatures = DNN(sentenceRepresentation)
    predictedBIOtags = Softmax(extractedFeatures)

    for tag in predictedBIOtags:
        posTags = pos_tags(tag)


    slotTree = BIOtoTree(predictedBIOtags,posTags)
    extractedSlots = parse(slotTree)

    return extractedSlots
```

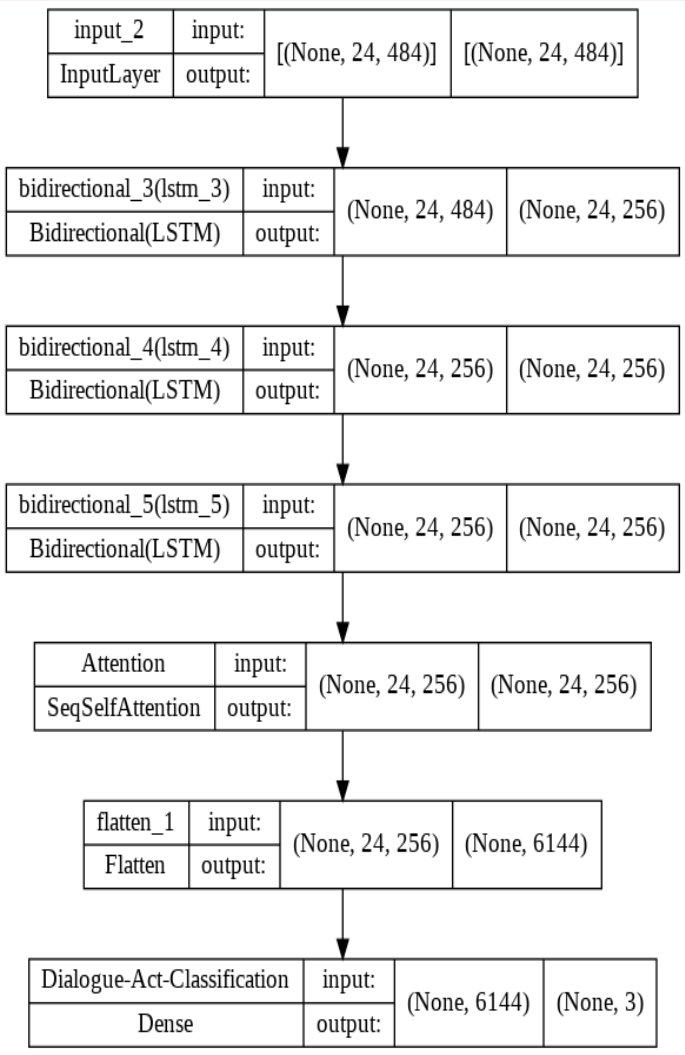
SLOT FILLING

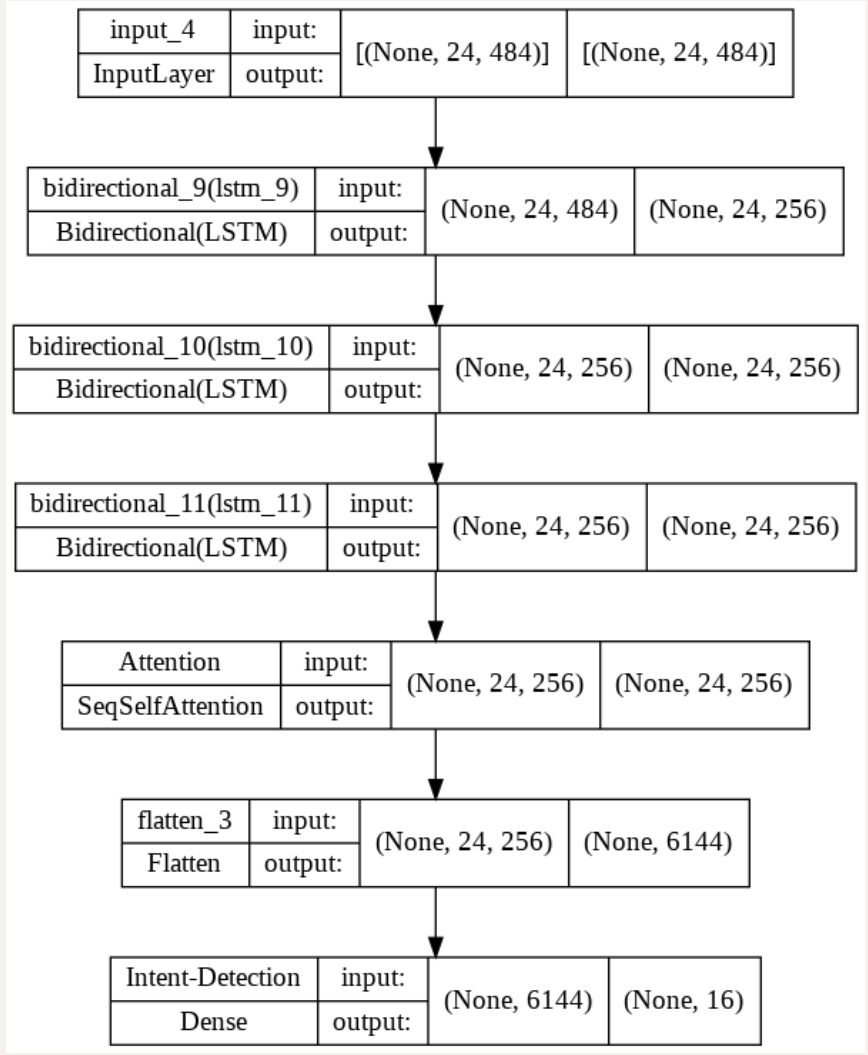


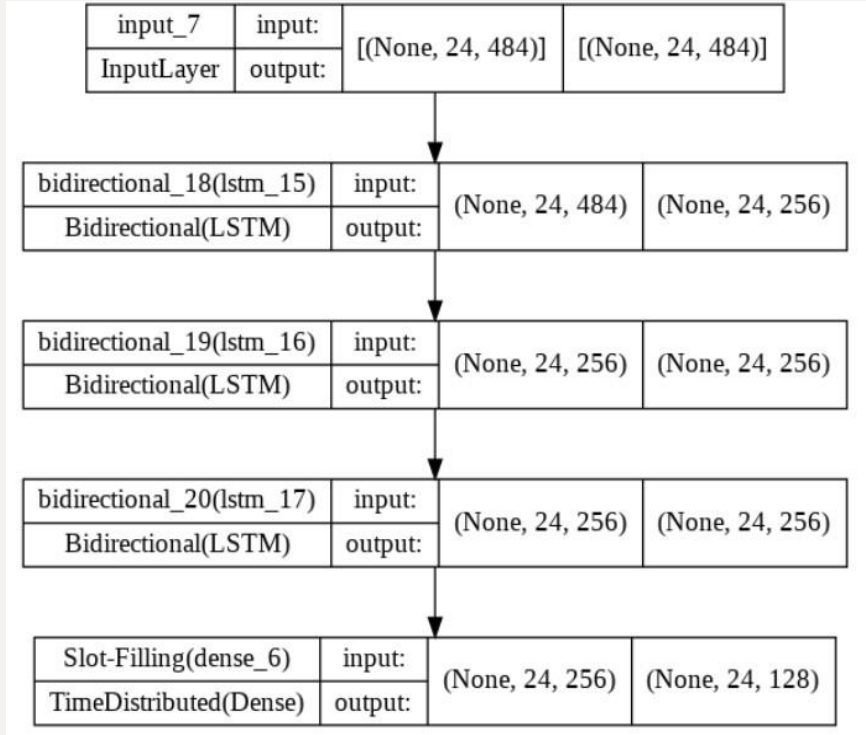
- **Input :** Task Shared Sentence Representation
- **Output :** Extracted slots from the utterance

```
Input Query  what does the restriction ap58 mean

Expected Slots:  O O O O B-restriction_code O

Predicted Slots:  ['O', 'O', 'O', 'O', 'B-restriction_code', 'O']
**************************************************************************
Input Query  please find a flight from las vegas to michigan

Expected Slots:  O O O O O B-fromloc.city_name I-fromloc.city_name O B-toloc.state_name

Predicted Slots:  ['O', 'O', 'O', 'O', 'O', 'B-fromloc.city_name', 'I-fromloc.city_name', 'O', 'B-toloc.city_name']
**************************************************************************
Input Query  list airlines flying from seattle to salt lake city

Expected Slots:  O O O O B-fromloc.city_name O B-toloc.city_name I-toloc.city_name I-toloc.city_name

Predicted Slots:  ['O', 'O', 'O', 'O', 'B-fromloc.city_name', 'O', 'B-toloc.city_name', 'I-toloc.city_name', 'I-toloc.city_name']
**************************************************************************
```

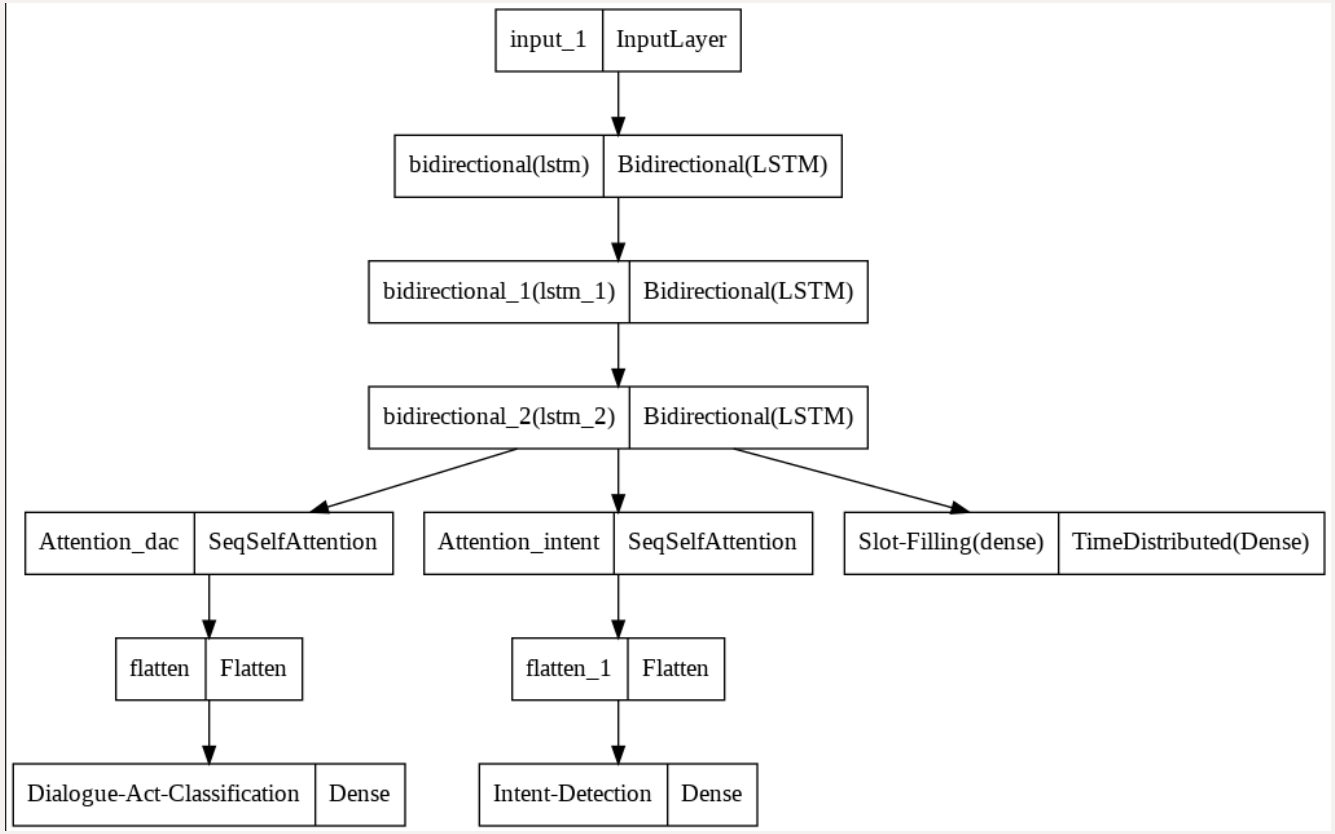
INDIVIDUAL MODELS







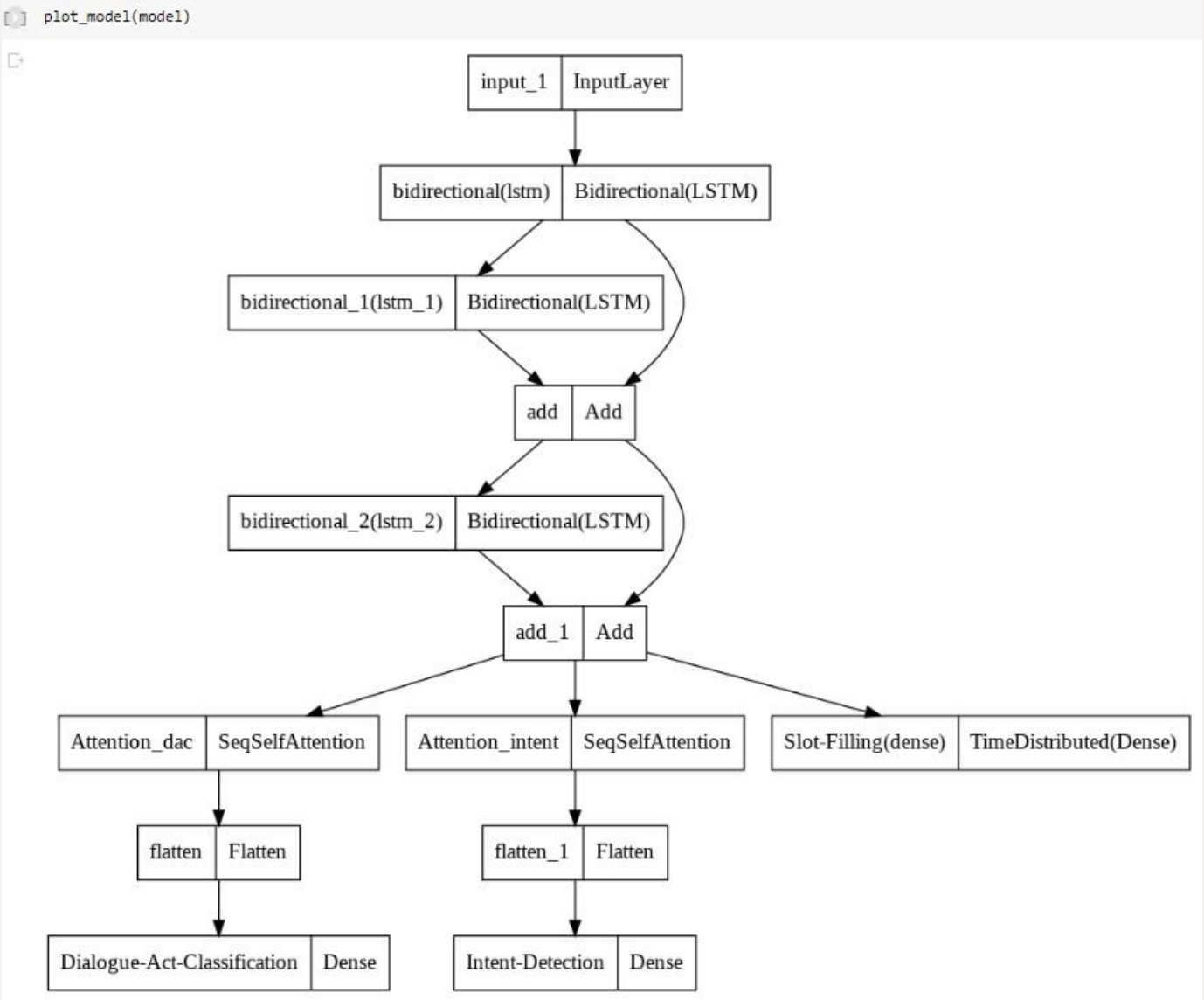
MULTI-TASK MODEL




```
Enter query: show me all meals on flights from atlanta to washington
Input Query: show me all meals on flights from atlanta to washington
100%|██████████| 1/1 [00:01<00:00,  1.72s/it]
Predicted Dialogue Act :  ['command']
Predicted Intent :  ['atis_meal']
Predicted Slots :  ['O', 'O', 'O', 'B-meal', 'O', 'O', 'O', 'B-fromloc.city_name', 'O', 'B-toloc.city_name']
```

MULTI-TASK MODEL – RESIDUAL CONNECTIONS

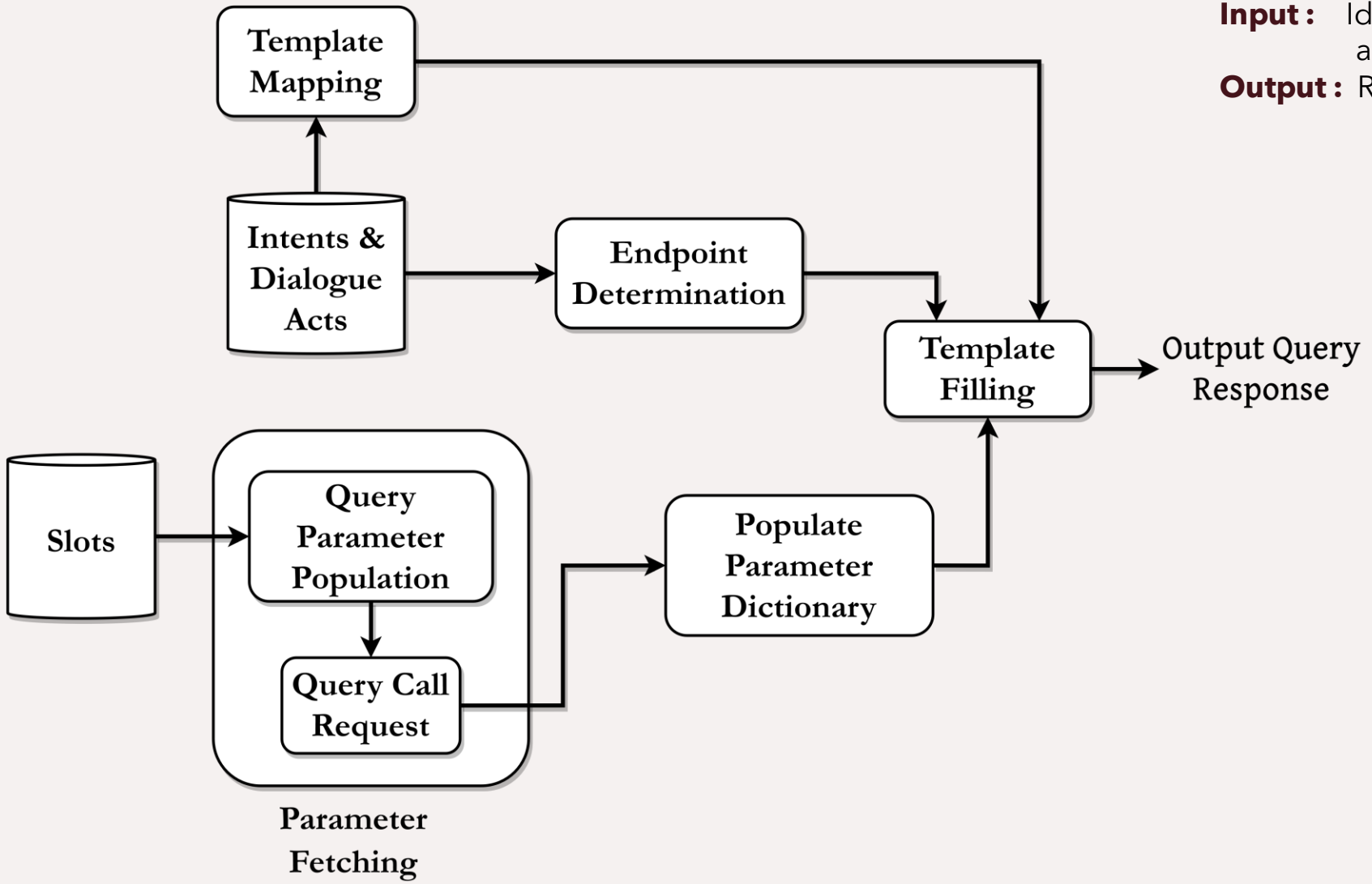
```
plot_model(model)
```



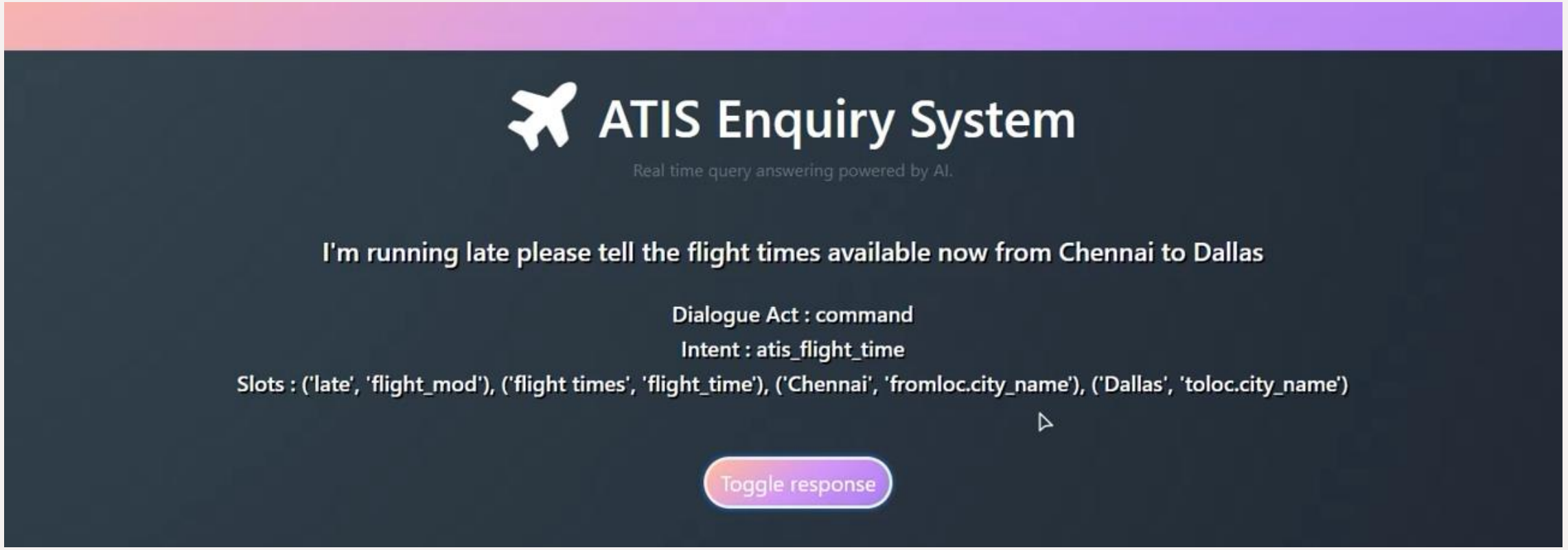
Module 6: Query Response Retrieval

Input : Identified Intents, Dialogue Acts and Slots from the user's utterance

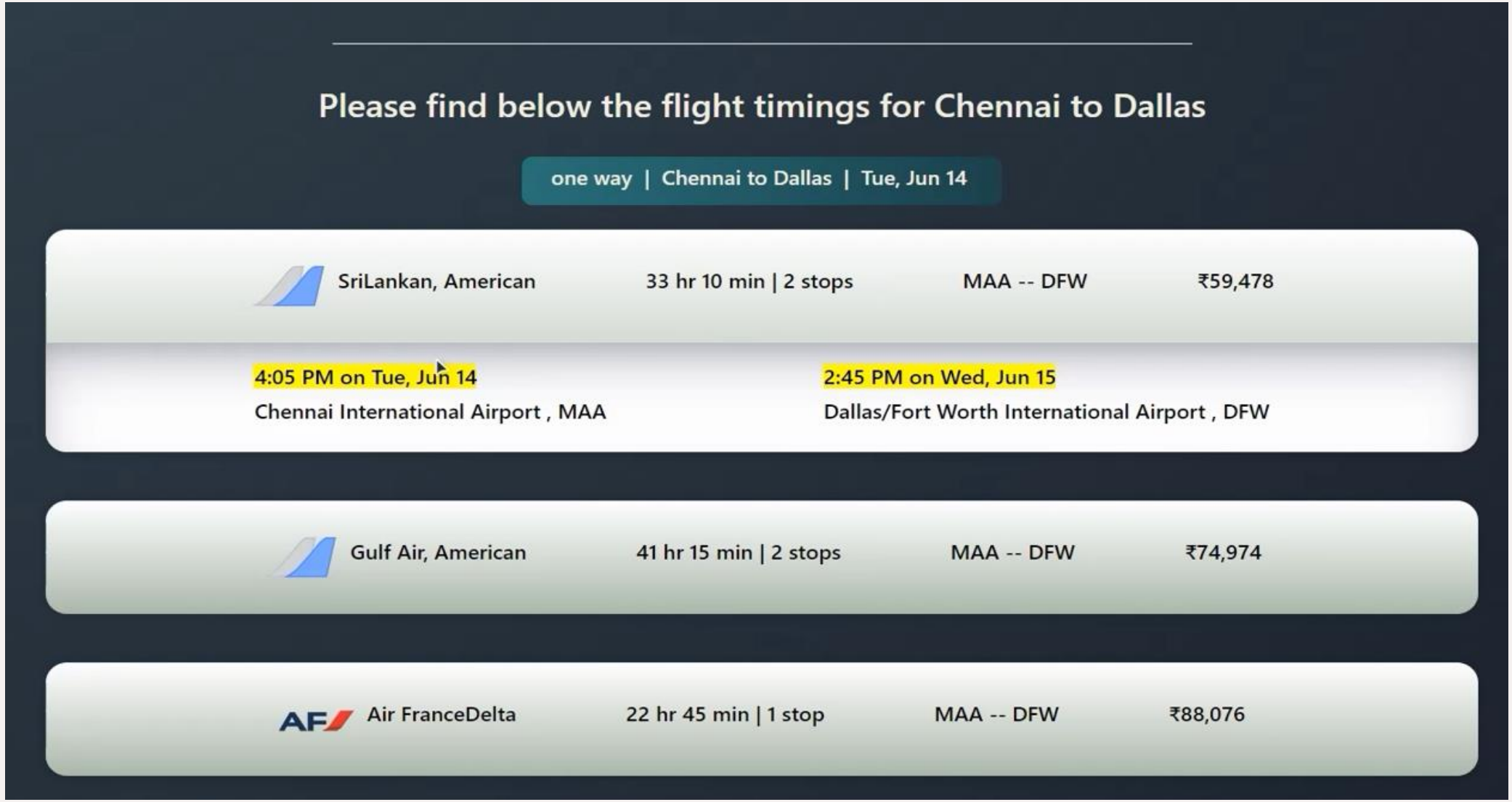
Output : Response to the user's query



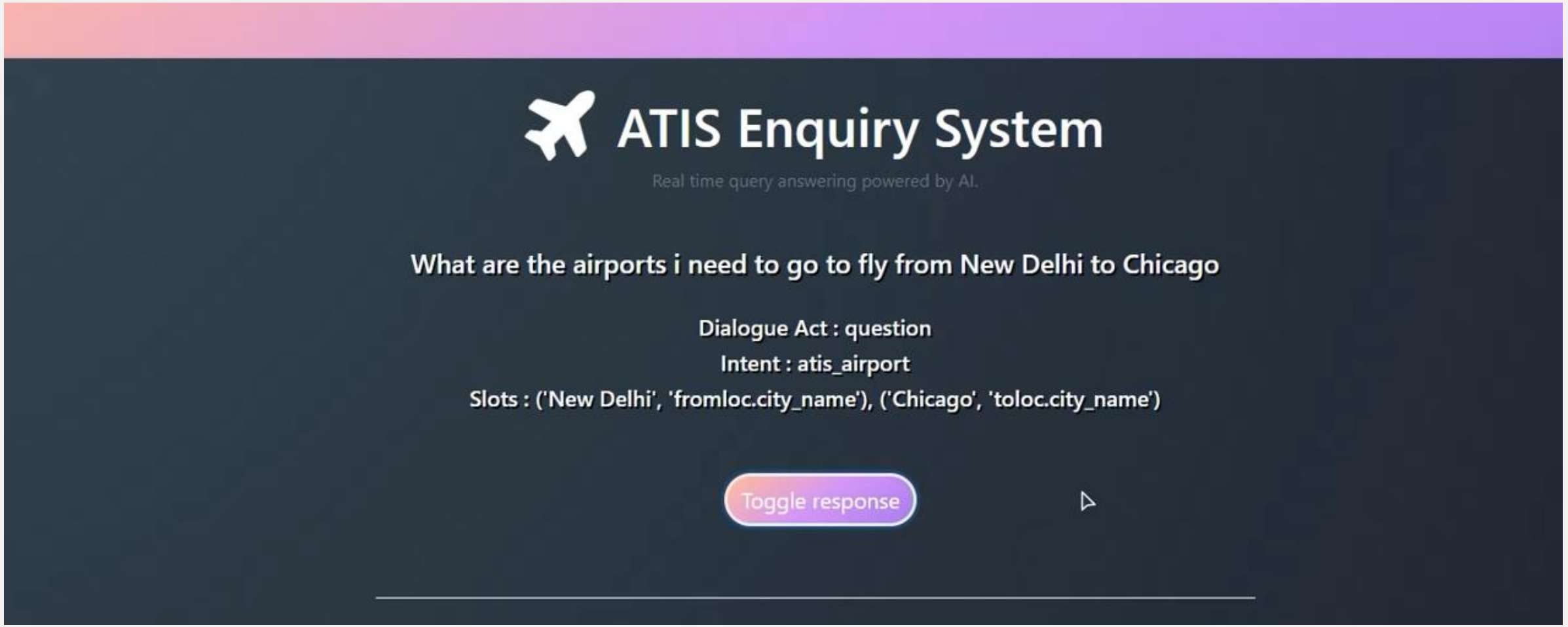
Module 6: Query Response Retrieval



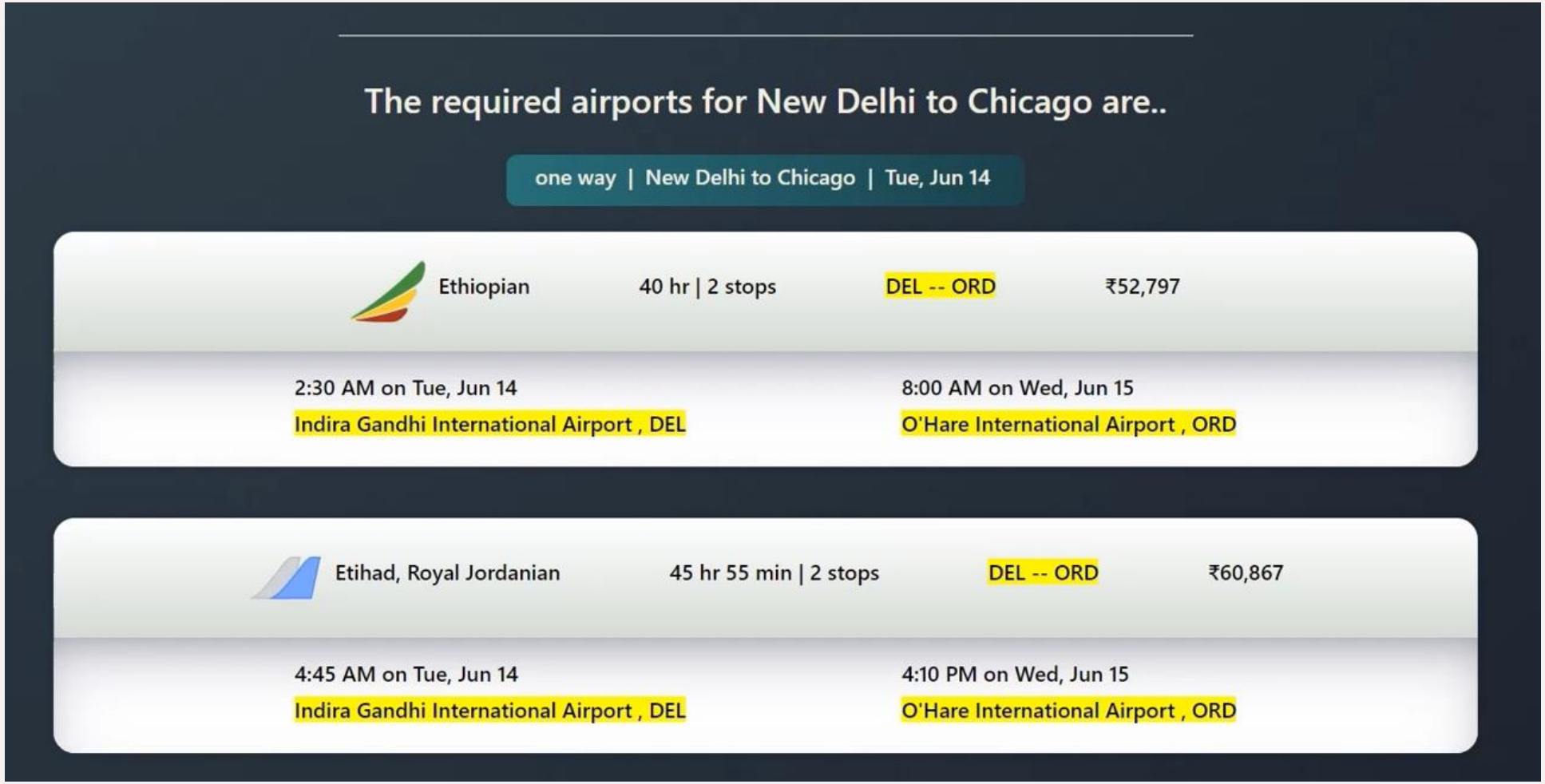
Module 6: Query Response Retrieval



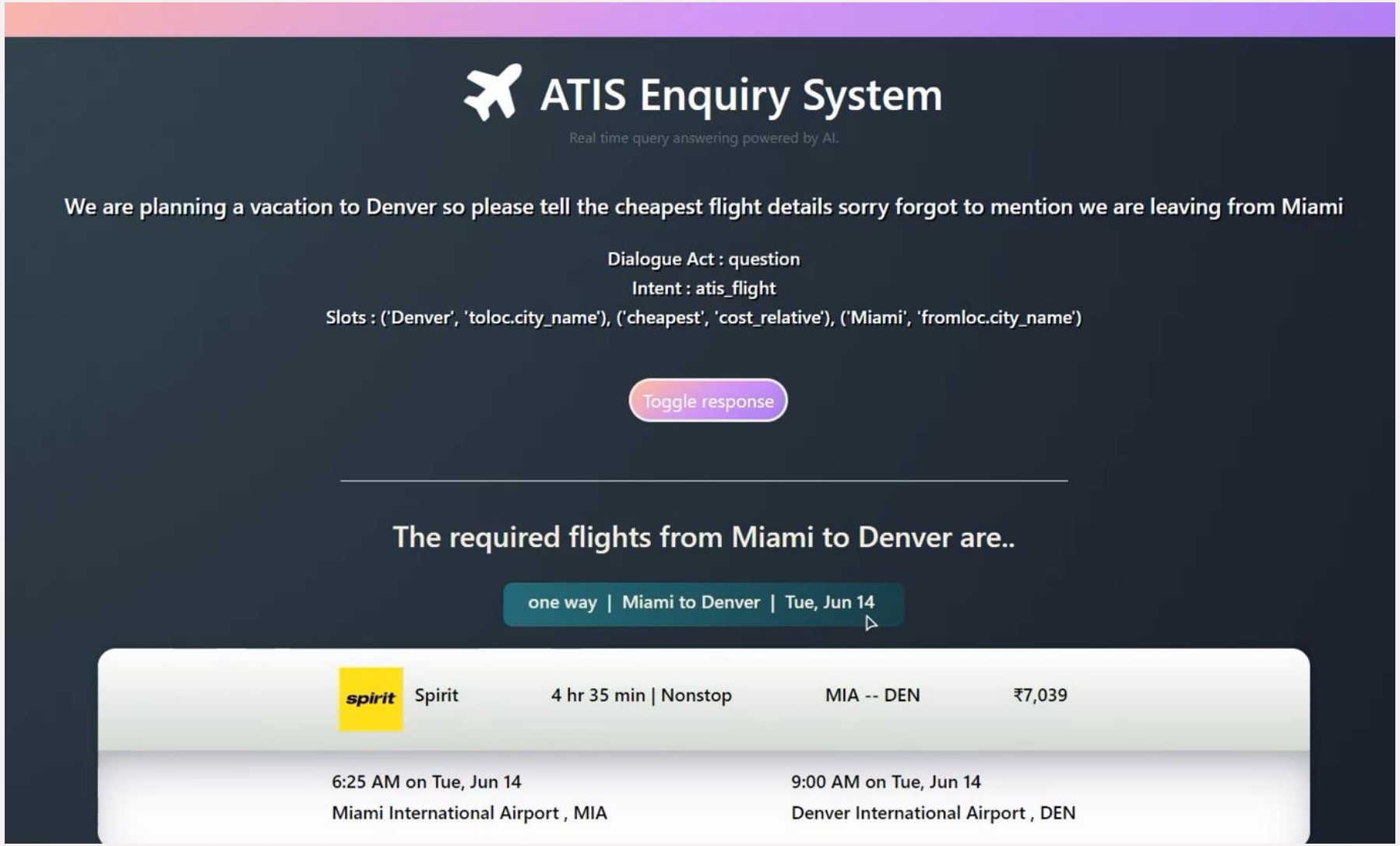
Module 6: Query Response Retrieval



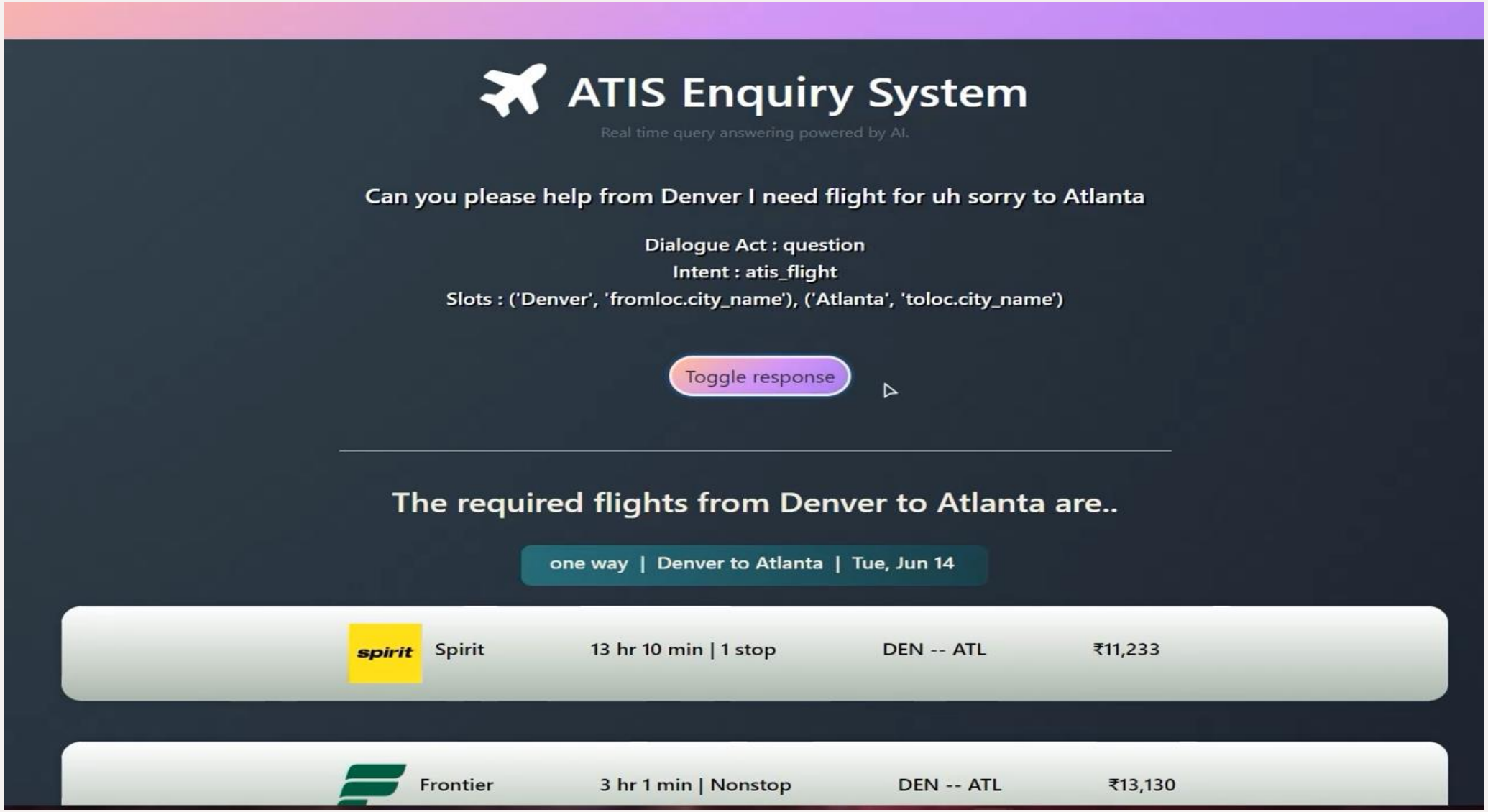
Module 6: Query Response Retrieval



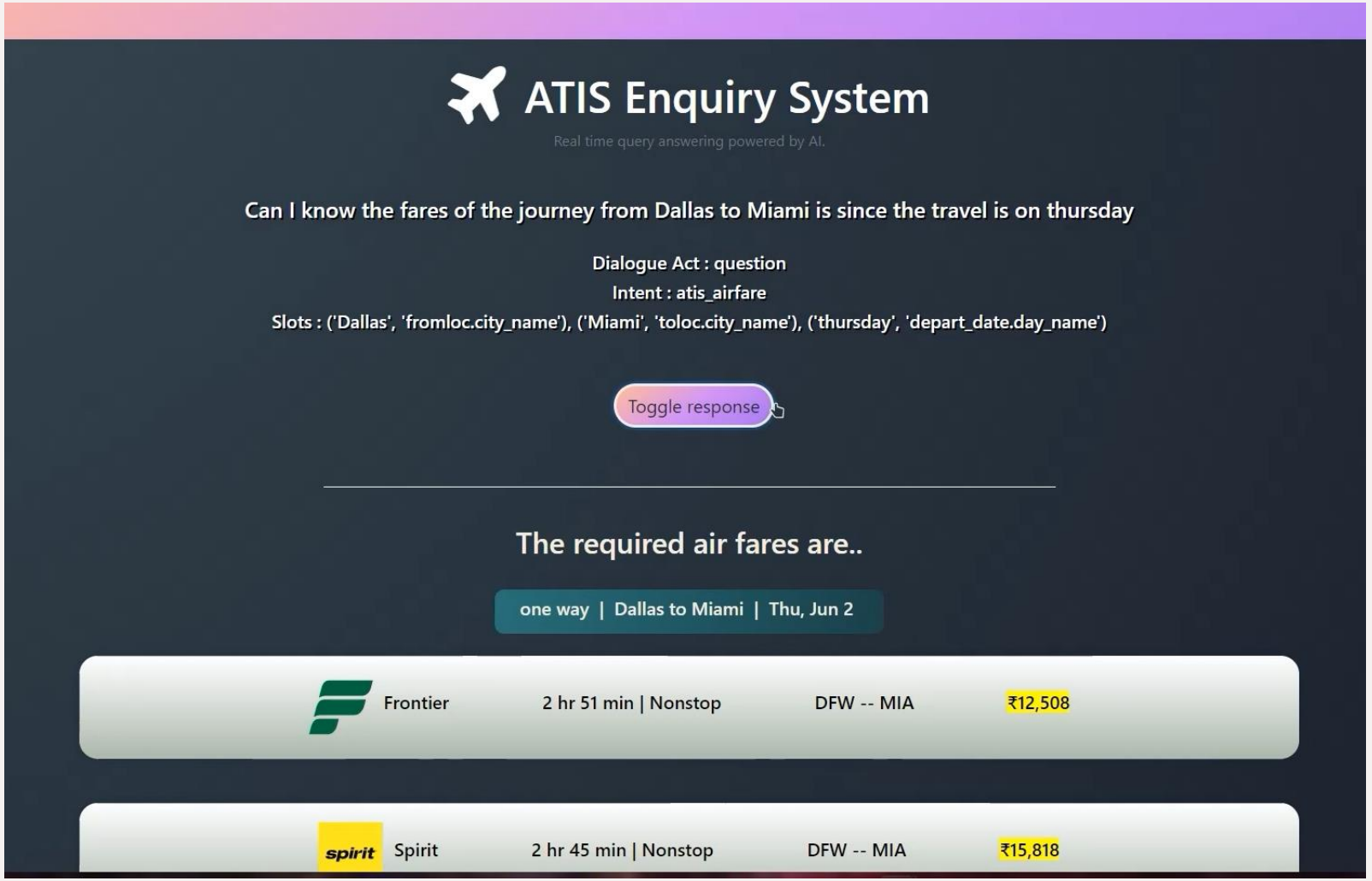
Module 6: Query Response Retrieval



Module 6: Query Response Retrieval



Module 6: Query Response Retrieval



DATASET

Dataset	Dialogue Acts	Intents	Slots	Total Utterances	Percentage Composition
ATIS	16	3	127	Train - 4978 Validation - 498 Test - 893	Train – 75% Validation – 10% Test – 15%
TRAINS	12	5	32	Train - 4819 Validation - 536 Test - 1336	Train – 73% Validation – 8% Test – 19%
FRAMES	24	10	136	Train - 19155 Validation - 2129 Test - 5321	Train – 72% Validation – 8% Test – 20%

PERFORMANCE METRICS

- **Precision**

Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

- **Recall**

Recall (also known as sensitivity) is the fraction of the total amount of relevant instances that were actually retrieved.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

- **Accuracy**

Accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

- **F1- Score**

The F1 score is defined as the weighted harmonic mean of the test's precision and recall.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Results – Non Residual Networks

Self-Attention	Embedding	#Units	ATIS			FRAMES			TRAINS		
		(LSTM)	DA	Intent	Slot	DA	Intent	Slot	DA	Intent	Slot
Multiplicative	BERT + Char CNN	100	***0.991***	***0.9854***	***0.9889***	0.4877	0.6309	0.9352	***0.7889***	***0.8121***	***0.9407***
Multiplicative	BERT + Char CNN	128	0.9843	0.9698	0.9806	0.4713	0.6288	0.9470	0.8054	0.6624	0.9410
Multiplicative	BERT	100	0.9765	0.9821	0.9841	0.4767	0.6221	0.9164	0.9368	0.6722	0.9407
Multiplicative	BERT	128	0.9597	0.9854	0.9889	0.4815	0.6196	0.9491	0.7912	0.8061	0.9419
Regularizer	BERT + Char CNN	100	0.9745	0.9776	0.9832	0.4670	0.6213	0.9476	0.7747	0.7702	0.9417
Regularizer	BERT + Char CNN	128	0.9698	0.9787	0.9864	0.4620	0.6226	0.9538	0.8174	0.7934	0.9422
Regularizer	BERT	100	0.9821	0.9821	0.9839	0.4653	0.6180	0.9424	0.7949	0.6886	0.9402
Regularizer	BERT	128	0.9664	0.9698	0.9821	0.4745	0.6176	0.9192	0.8061	0.7949	0.9382

*The metrics for Dialogue Act Classification and Intent Detection are indicated in accuracy while Slot Filling is measured using F1-Score.

Results – Residual Networks

Self-Attention	Embedding	#Units	ATIS			FRAMES			TRAINS		
		(LSTM)	DA	Intent	Slot	DA	Intent	Slot	DA	Intent	Slot
Multiplicative	BERT + Char CNN	100	0.9765	0.9776	0.9814	0.4886	0.6236	0.9647	0.8046	0.6729	0.9410
Multiplicative	BERT + Char CNN	128	0.9854	0.9832	0.9749	0.451	0.6180	0.9676	0.8054	0.6624	0.9406
Multiplicative	BERT	100	0.9798	0.9709	0.9794	0.4801	0.6236	0.9629	0.8136	0.6909	0.9405
Multiplicative	BERT	128	0.9742	0.9832	0.9796	0.4751	0.6184	0.9656	0.8114	0.6677	0.9400
Regularizer	BERT + Char CNN	100	0.9854	0.9787	0.9850	0.458	0.6327	0.9657	0.8009	0.8061	0.9400
Regularizer	BERT + Char CNN	128	0.9765	0.9810	0.9772	***0.4763***	***0.6161***	***0.9677***	0.8121	0.7590	0.9412
Regularizer	BERT	100	0.9742	0.9810	0.9818	0.4811	0.6246	0.9539	0.8054	0.7949	0.9392
Regularizer	BERT	128	0.9798	0.9832	0.9861	0.4659	0.6292	0.9604	0.8099	0.7028	0.9397

*The metrics for Dialogue Act Classification and Intent Detection are indicated in accuracy while Slot Filling is measured using F1-Score.

Results – Best Performing Hyperparameter

Dataset	Self-Attention	Embedding	LSTM units	Residual Connections
ATIS	Multiplicative - 15	BERT + Char CNN	100	No
FRAMES	Regularizer - 15	BERT + Char CNN	128	Yes
TRAINS	Multiplicative - 15	BERT + Char CNN	100	No

Future Scope

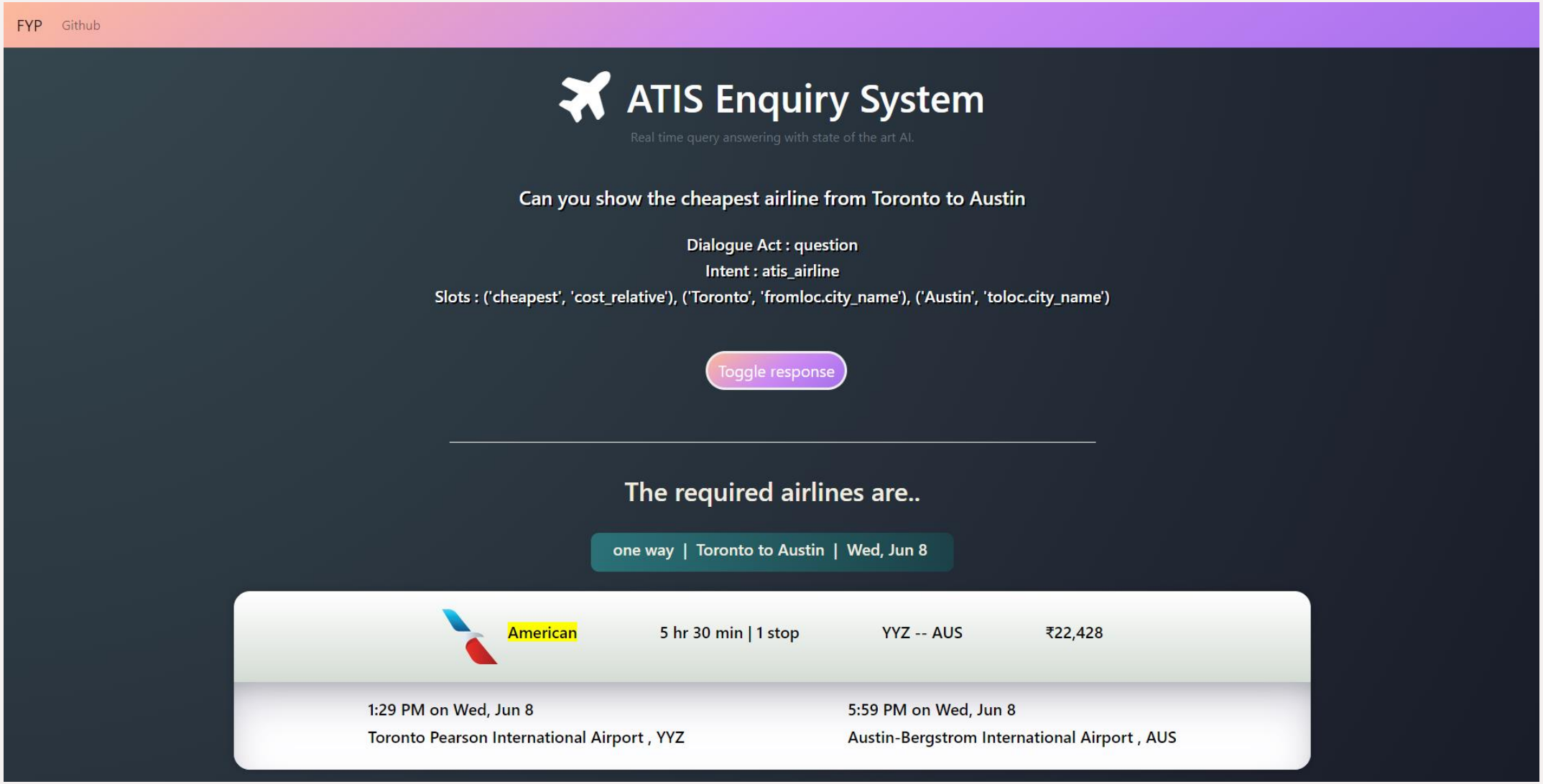
- Making the system **more stateful** rather than being one-shot.
- Employing Graph Convolutional Neural Networks (GCN) and Semantic Web structures that **inject more structural and semantic dependencies** into the system architecture.
- Adopting an extension of automatic inquiry system to life saving situations like an **army combat field**.

REFERENCES

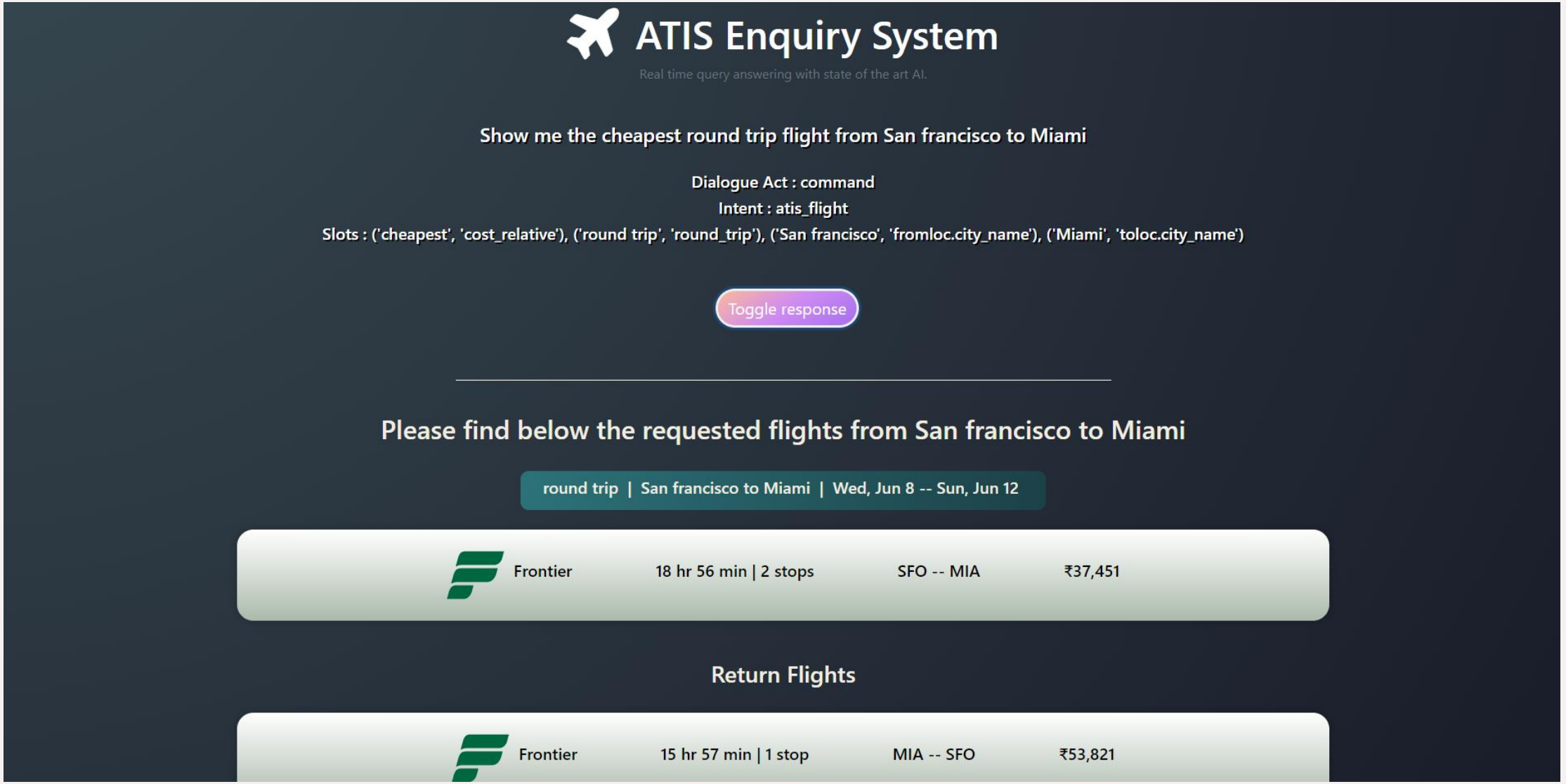
1. "A deep multi-task model for dialogue act classification, intent detection and slot filling." Firdaus, Mauajama, Hitesh Golchha, Asif Ekbal, and Pushpak Bhattacharyya. *Cognitive Computation* 13, no. 3 (2020)
2. "AISE: Attending to Intent and Slots Explicitly for better spoken language understanding." Yang, Peng, Dong Ji, Chengming Ai, and Bing Li. *Knowledge-Based Systems* 211 (2021)
3. "Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction." Ni, Pin, Yuming Li, Gangmin Li, and Victor Chang. *Neural Computing & Applications* 32, no. 20 (2020)
4. "Multi-turn intent determination and slot filling with neural networks and regular expressions." Abro, Waheed Ahmed, Guilin Qi, Zafar Ali, Yansong Feng, and Muhammad Aamir. *Knowledge-Based Systems* 208 (2020)
5. "Dual learning for semi-supervised natural language understanding." Zhu, Su, Ruisheng Cao, and Kai Yu. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020)
6. "End-to-end masked graph-based CRF for joint slot filling and intent detection." Tang, Hao, Donghong Ji, and Qiji Zhou. *Neurocomputing* 413 (2020)
7. "A multi-task hierarchical approach for intent detection and slot filling." Firdaus, Mauajama, Ankit Kumar, Asif Ekbal, and Pushpak Bhattacharyya. *Knowledge-Based Systems* 183 (2019)

THANK YOU

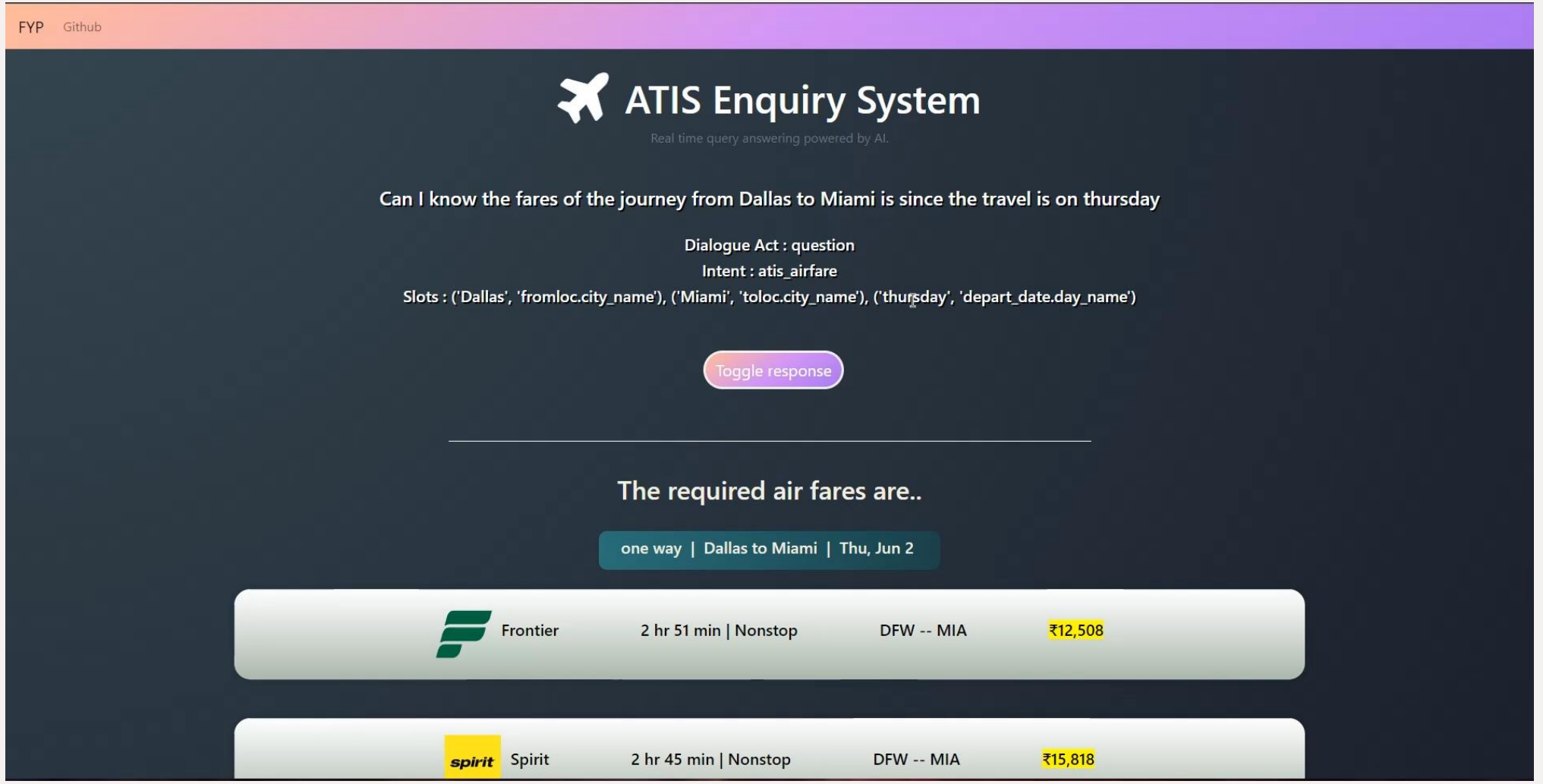
Cheapest One-way



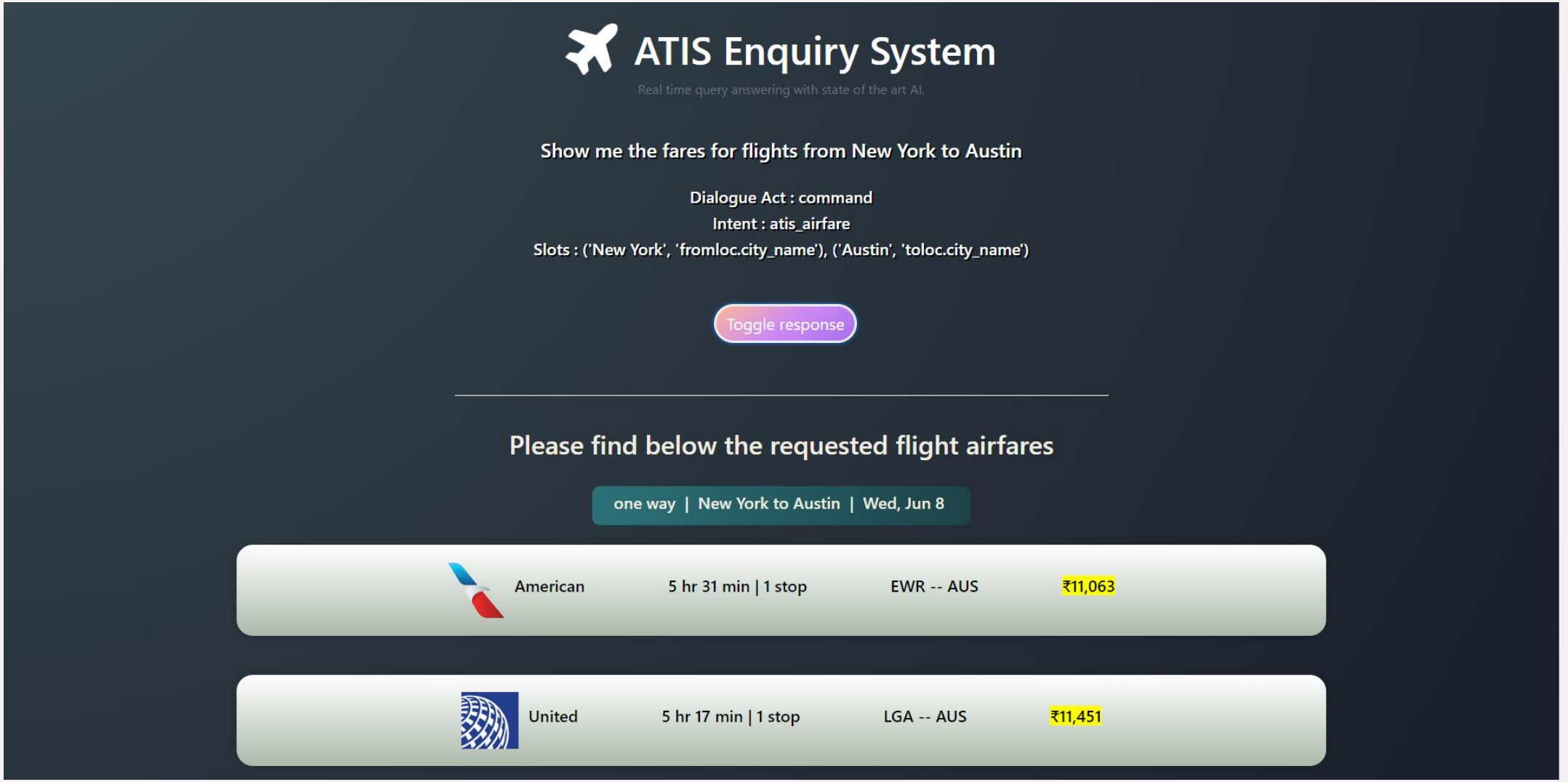
Cheapest round trip flight



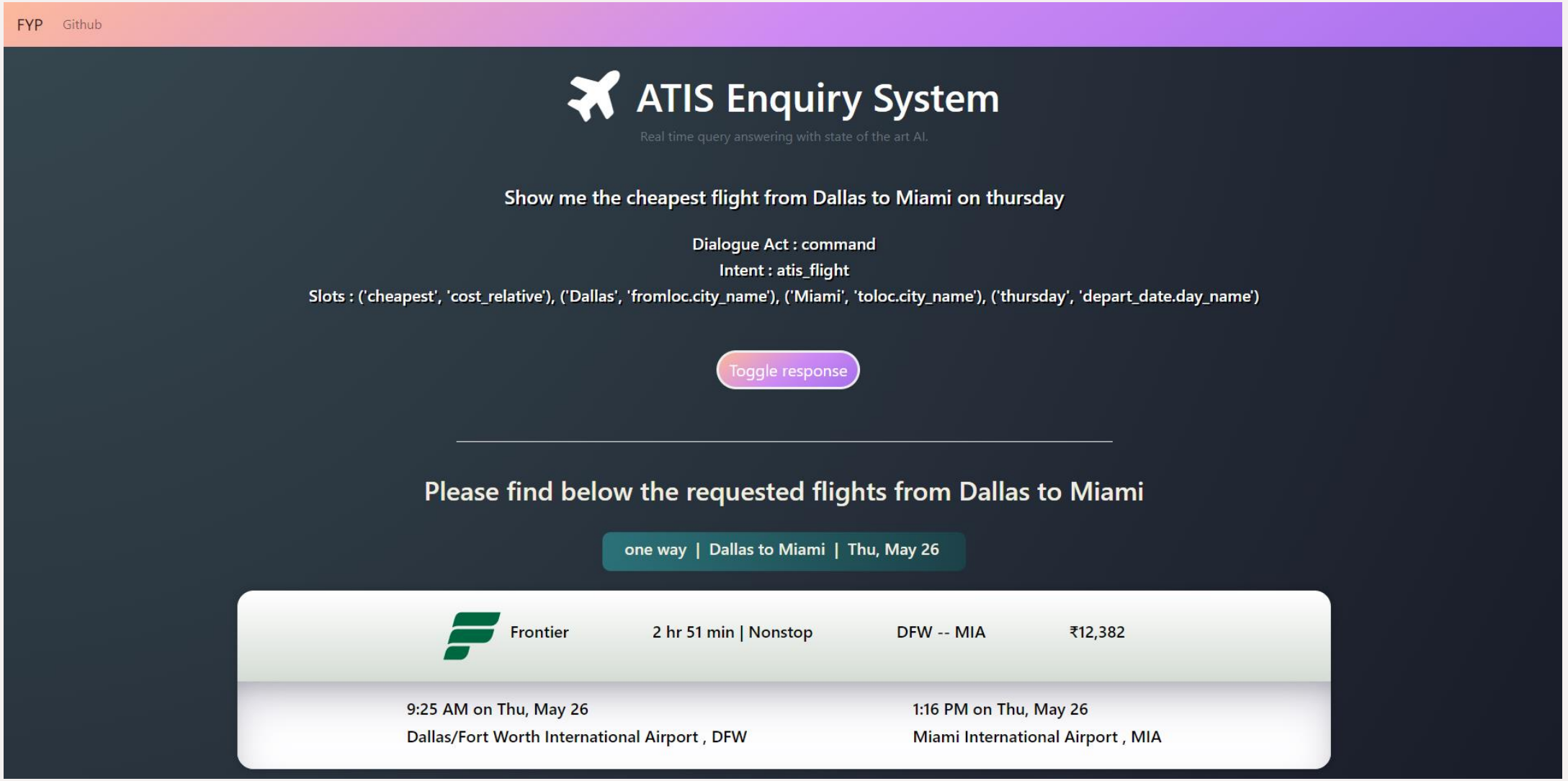
Flight Fares on a Specific Day



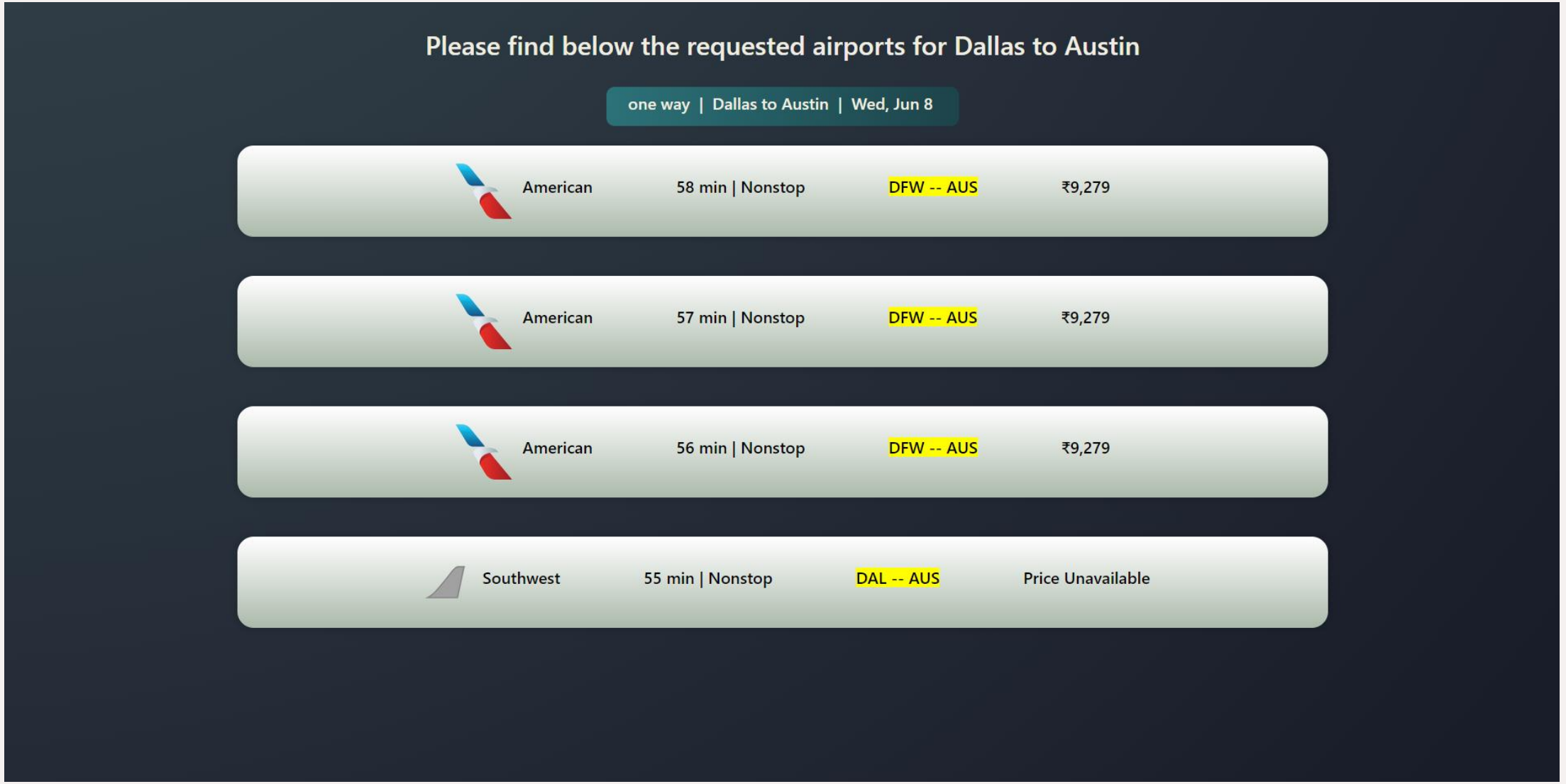
Specific Intent – Airfares



Flights on a specific day



Specific Intent – Airports



Specific Intent – Flight Timings