

# Chatbot Report

Shreya Valaboju, Soham Mukherjee

CS 4395.001

Github Links:

[https://github.com/Zakenmaru/CS4395\\_Portfolio](https://github.com/Zakenmaru/CS4395_Portfolio)

[https://github.com/shreyavala/nlp\\_portfolio](https://github.com/shreyavala/nlp_portfolio)

System Description	2
Dialog Logic	4
Sample Dialog Interactions	5
Knowledge Base Appendix	7
User Model Appendix	8
Evaluation and Analysis	9

# System Description

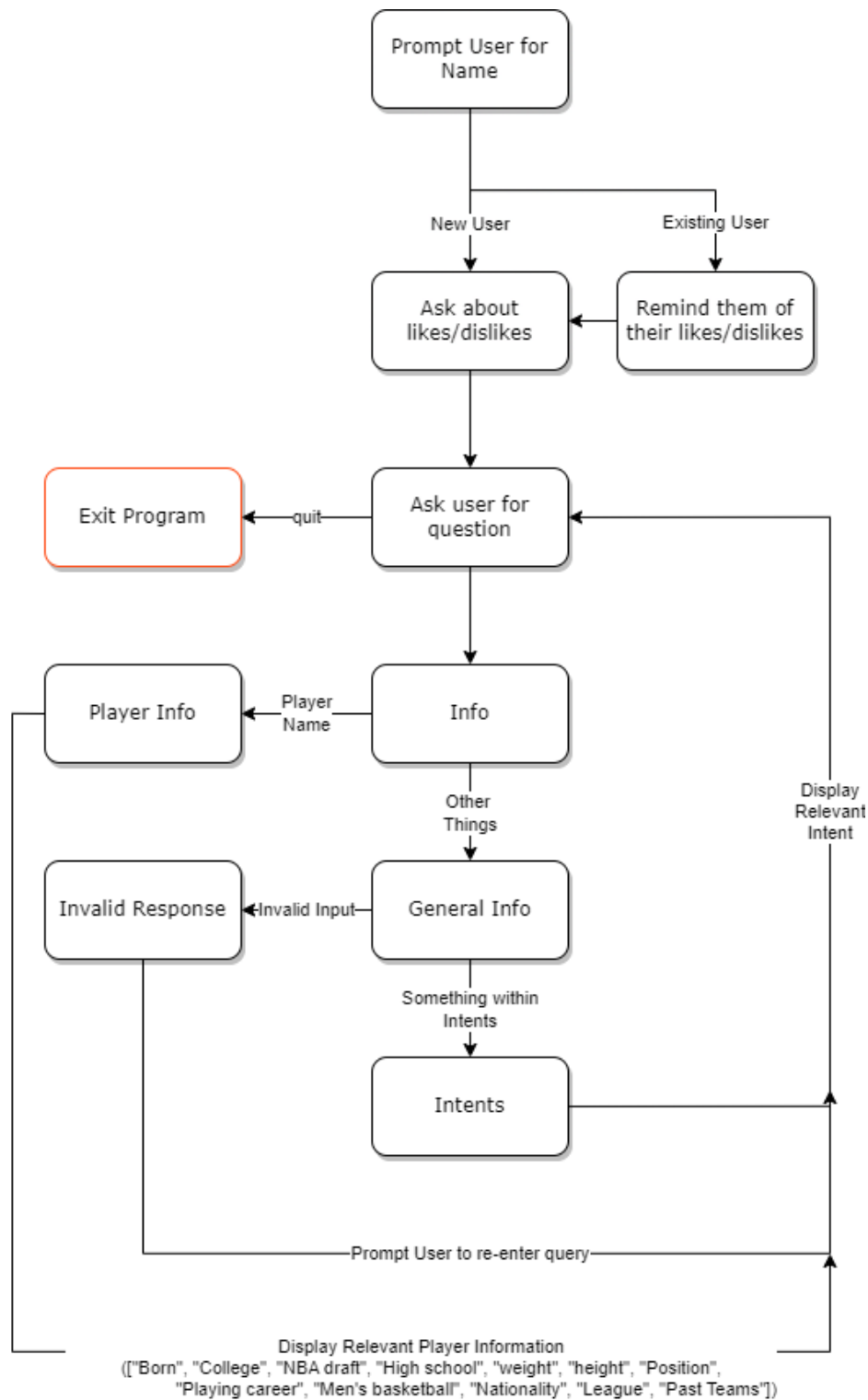
The chatbot we created is named “Champ.” Champ was built to answer specific and general questions about the players on the Dallas Mavericks NBA team. Champ is implemented in python. It is capable of retrieving a variety of basic information about a player. For instance, Champ can answer questions about what a player’s height or weight is, where he was born, where he went to college and high school, what his NBA draft was like, playing career, and more.

We created Champ from scratch using several NLP techniques. In our implementation, we started off by providing a description of the bot and the kinds of information it has to the user. Next, we stored some details, such as name and likes and dislikes, from the user. After retrieving information about the user for the user model, Champ prompts them to ask a question. The user’s question or statement is preprocessed using nltk as the query is tokenized, stopwords and punctuation are removed, and lemmatized. The next phase is training, having Champ understand and correctly answer the question. Champ retrieves information about the player using a knowledge base we created by scraping the Dallas Mavericks Wikipedia page. The knowledge base is stored as a python dictionary.

Our implementation uses tf-idf vectorization and cosine similarity to understand what and who the user is inquiring about. First, we try to identify whether the user mentions a player we have information on. So, we construct a tf-idf vectorizer between each player and the user query. Next, the cosine similarity is computed between each player and the user query. Finally, we identify the player name by finding which player had the highest similarity score to the user query. Similarly, we try to find a topic, like height, weight, playing career, etc., in the user query using tf-idf vectorization and cosine similarity. However, sometimes, the user may refer to a topic using a different word/phrase. For example, if a user wanted to know where a certain player was born, they may say “birthplace” or “born.” To deal with this, we used WordNet to identify synonyms for a topic to which a user may refer. Finally, if the exact player and topic are found in the knowledge base, Champ prints that information. If a player is identified and a topic is not, Champ prints general information about that player. We used nltk to implement the NLP techniques described.

We also have additional functionality, such as sentiment analysis and dealing with information unrelated to the players on the team. Specifically, if the user expresses negative sentiment, Champ acknowledges the user and apologizes. We used VADER to implement the sentiment analysis. For Champ to reply to basic greetings and random questions about Mark Cuban or Dirk Nowitzki, we implemented a small, basic intents file that we use to match key words and phrases and randomly select a response back to the user.

# Dialog Logic



# Sample Dialog Interactions

Users can converse with Champ through a variety of interactions. Here is some sample dialog:

1. Asking something specific about a player that Champ has info on:

```
Champ: Howdy Shreya! Go ahead, ask me question
Shreya: Did Reggie Bullock go to college? I can't remember.
Champ: Here's some information about Reggie Bullock relating to College
'North Carolina (2010-2013)'
```

2. Getting general player info:

```
Shreya: tell me about kyrie
Champ: Could not find that specific information about Kyrie Irving
Champ: Here are some facts about Kyrie Irving:
    Kyrie Irving is a player on the 2022-2023 Dallas Mavericks team.
    He has been playing basketball since 2011-present
    He plays Point guard / Shooting guard
```

3. User displaying negative emotion not about any player:

```
Shreya: THIS TEAM IS SO BAD I HATE WATCHING THE MAVS SO TRASH
Champ: Sorry you feel this way. Try asking me something about a player (general/specific). I will try my best to help you.
Shreya: |
```

4. Asking something specific about a player that Champ doesn't have info on:

```
Shreya: Do you have information about josh green's nationality?
Champ: Could not find information about Josh Green and their Nationality; it may not exist. Try again?
Shreya: What about irving's nationality?
Champ: Here's some information about Kyrie Irving relating to Nationality
'American'
Shreya: Did luka go to high school, If so, where?
Champ: Could not find information about Luka Doncic and their High school; it may not exist. Try again?
```

## 5. Asking about past teams a player has been on

```
Shreya: What are javale's past team?
Champ: Here's some information about JaVale McGee relating to Past Teams
1. 2008-2012: Washington Wizards
2. 2012-2015: Denver Nuggets
3. 2015: Philadelphia 76ers
4. 2015-2016: Dallas Mavericks
5. 2016-2018: Golden State Warriors
6. 2018-2020: Los Angeles Lakers
7. 2020-2021: Cleveland Cavaliers
8. 2021: Denver Nuggets
9. 2021-2022: Phoenix Suns
10. 2022-present: Dallas Mavericks
Shreya: Luka past teams
Champ: Here's some information about Luka Doncic relating to Past Teams
1. 2015-2018: Real Madrid
2. 2018-present: Dallas Mavericks
```

## 6. Random/Funny statements or questions

```
Shreya: Who owns this team?
Champ: Mark Cuban, owner of the Dallas Mavs
Shreya: Who is the GOAT
Champ: All time GOAT, #41 - Dirk Nowitzki
Shreya: who is the best player of all time
Champ: Our very own Hall of Famer - Dirk Nowitzki!!
Shreya: Who is devin booker?
Champ: Devin Booker is Luka's son, born on May 15, 2022.
Shreya: I love you, you are the best bot ever
Champ: I'm happy to help you, but try asking me something about a player.
```

## 7. General Mavs/NBA question

```
Shreya: who are the dallas mavs?
Champ: The Dallas Mavericks is a basketball team part of the NBA (National Basketball Association)
Shreya: what is the nba?
Champ: Dallas Mavericks is an NBA (National Basketball Association) team located in Dallas, Texas.
```

# Knowledge Base Appendix

This is an example of how the knowledge base is formatted. It is a python dictionary that is extracted from a pickle file. Below are 2 of 15 players used as examples to showcase the format.

```
{'Christian_Wood': {'Born': '(1995-09-27) September 27, 1995 (age 27)Long '
                    'Beach, California U.S.',
                    'College': 'UNLV (2013-2015)',
                    'High school': 'Los Alamitos(Los Alamitos, '
                                   'California)~Knight(Palmdale, '
                                   'California)~Findlay(Henderson, Nevada)',
                    'League': 'NBA',
                    'NBA draft': '2015 / Undrafted',
                    'Past Teams': [{'2015-2016': 'Delaware 87ers',
                                   '2016-2017': 'Greensboro Swarm',
                                   '2017-2018': 'Delaware 87ers',
                                   '2018-2019': 'Wisconsin Herd',
                                   '2019': 'New Orleans Pelicans',
                                   '2019-2020': 'Detroit Pistons',
                                   '2020-2022': 'Houston Rockets',
                                   '2022-present': 'Dallas Mavericks'}],
                    'Playing career': '2015-present',
                    'Position': 'Power forward / Center',
                    'height': '6 ft 9 in (2.06 m)',
                    'name': 'Christian Wood',
                    'weight': '214 lb (97 kg)'}}
```

```
'Tim Jr._Hardaway': {'Born': '(1992-03-16) March 16, 1992 (age 31)Alameda, '
                       'California, U.S.',
                       'College': 'Michigan (2010-2013)',
                       'High school': 'Miami Palmetto (Pinecrest, Florida)',
                       'League': 'NBA',
                       'NBA draft': '2013 / Round: 1 / Pick: 24th overall',
                       'Past Teams': [{'2013-2015': 'New York Knicks',
                                       '2015': 'Canton Charge',
                                       '2015-2016': 'Austin Spurs',
                                       '2015-2017': 'Atlanta Hawks',
                                       '2017-2019': 'New York Knicks',
                                       '2019-present': 'Dallas Mavericks'}],
                       'Playing career': '2013-present',
                       'Position': 'Shooting guard / Small forward',
                       'height': '6 ft 5 in (1.96 m)',
                       'name': 'Tim Hardaway Jr.',
                       'weight': '205 lb (93 kg)'}}
```

# User Model Appendix

Below is a screenshot of the user model with various fields storing information about users:

```
{
  "Howdy": {
    "name": "Howdy",
    "personal_info": {},
    "likes": [],
    "dislikes": []
  },
  "Siadls": {
    "name": "Siadls",
    "personal_info": {},
    "likes": ["pie", "cheese"],
    "dislikes": []
  },
  "Siadl": {
    "name": "Siadl",
    "personal_info": {},
    "likes": ["pie"],
    "dislikes": ["cheese"]
  },
  "Choco": {
    "name": "Choco",
    "personal_info": {},
    "likes": ["chocolate pie and herbs", "carrots"],
    "dislikes": ["carrots"]
  },
  "SM": {
    "name": "SM",
    "personal_info": {},
    "likes": ["Japan and Claymation"],
    "dislikes": ["the Bulls"]
  },
  "ADSJ": {
    "name": "ADSJ",
    "personal_info": {},
    "likes": ["fried rice and potatoes"],
    "dislikes": ["oranges."]
  },
  "Beans": {
    "name": "Beans",
    "personal_info": {},
    "likes": ["radishes but I"],
    "dislikes": ["oranges"]
  },
  "SDIUA": {
    "name": "SDIUA",
    "personal_info": {},
    "likes": ["peaches", "japan"],
    "dislikes": ["mangoes"]
  },
  "John Likebody": {
    "name": "John Likebody",
    "personal_info": {},
    "likes": [],
    "dislikes": ["crows", "kettles"]
  },
  "John Like": {
    "name": "John Like",
    "personal_info": {},
    "likes": [],
    "dislikes": ["carrots."]
  },
  "Cheese Like": {
    "name": "Cheese Like",
    "personal_info": {},
    "likes": ["cheese", "china"],
    "dislikes": ["chocolate", "pineapples", "chicharron"]
  },
  "Sana": {
    "name": "Sana",
    "personal_info": {},
    "likes": ["Cetaphil"],
    "dislikes": ["the"]
  },
  "John": {
    "name": "John",
    "personal_info": {},
    "likes": ["candy", "remotes"],
    "dislikes": []
  },
  "Soda": {
    "name": "Soda",
    "personal_info": {},
    "likes": [],
    "dislikes": []
  },
  "Shreya": {
    "name": "Shreya",
    "personal_info": {},
    "likes": [],
    "dislikes": ["jason"]
  },
  "Rekha": {
    "name": "Rekha",
    "personal_info": {},
    "likes": ["skincare"],
    "dislikes": []
  }
}
```

The users can choose whether or not they'd like to name likes/dislikes. Champ remembers a user using the information stored in this user model.



## Evaluation and Analysis

Champ is a simple chatbot that leverages a variety of NLP techniques. We hope to continue to improve and build upon Champ's features. Currently, Champ is able to answer specific questions about a player very well. Champ's range of topics is limited to a player's birth-place/hometown, playing career, past teams, college, high school, height, weight, position, and draft. Therefore, we aim to add more details to the knowledge base, such as advanced analytics stats, awards, and personal life information. This way, Champ can answer a larger range of topics and questions. In addition to adding more topics, we want to create a GUI for Champ to make the user experience more enjoyable and interactive than a command line console.

Champ is easy to talk to and provides responses to any question a user may have. Some strengths of Champ is that it can identify sentiment, particularly frustration or negative emotions. Another impressive attribute is being able to answer any question on the players. Meaning, Champ can support dynamic querying of any Mavs player and respond promptly and correctly. Additionally, it can remember details about previous users and uses personalized remarks to address them.

Although Champ is able to carry a conversation with a user, it is limited. This is partly because the implementation is completely from scratch. An area of improvement we look to address is broadening the types of questions Champ can answer. Meaning, Champ cannot answer queries like "Who is the tallest player on the team?" or "Which player had a higher draft: Luka or Javale?" It cannot compare different attributes between players. Therefore, we could allow our bot to process more than 1 player, act upon terms such as "compare," and read from our knowledge base accordingly. Another weakness of Champ is that its domain is specific and the user is expected to have some prior knowledge. For example, the user should know that the Dallas Mavericks is an NBA team and players have certain biographical information specific to basketball players. Nevertheless, we hope that we can cater Champ to a more newbie/general audience, have it better answer general questions about the Mavericks outside of who Mark Cuban is. Next, currently Champ's responses are not generated through some machine learning technique. Champ's responses are fixed and coded based on a user's query. Our chatbot can be enhanced by incorporating NLP generation techniques, through rules-based and NLP

approaches, thus having Champ create customized, logical, and human-like replies. Another improvement we want to make is adding more details about the user to the user model. As of now, the user model stores likes/dislikes and names at the beginning of a session after Champ's prompt. However, that is only at the beginning. Perhaps we could identify when the user says they like or dislike something even in the middle of a conversation with Champ. Overall, there are many improvements and enhancements that can be made. By adding more sophisticated NLP and machine learning techniques, we can make Champ more versatile and intelligent.