Design, Implementation, and Analysis of an LLM-Powered Sports Chatbot

Explore the innovative design, implementation, and analysis of a powerful sports chatbot enhanced with dynamic web augmentation and semantic caching techniques.

Team BitMask

Navigating the Sports Information Landscape

Addressing User Needs in Sports Information

1

Dynamic Sports Information

Sports data is vast and rapidly changing, requiring instant updates on live scores and news.

2

User Demand for Speed

Users seek quick and reliable conversational answers to their sports queries.

3

Limitations of LLMs

Standard LLMs have static knowledge, making them less effective for real-time information.

4

Real-time Data Integration

Combining LLMs with real-time web data presents a significant opportunity.

5

Chatbot Development Goal

Our goal is to create a chatbot that intelligently fetches and utilizes live sports information.

Challenges Integrating LLMs with Live Sports Data

Addressing LLM Limitations in Sports Engagement

Static LLM Limitations

LLMs lack access to real-time game scores, breaking news, or recent results, creating a gap in user expectations.

Technical Integration

The challenge lies in efficiently integrating live web data into the LLM's workflow without redundancy in scraping.

User Expectations

Users expect accurate, up-to-the-minute information when engaging with sports chatbots conversationally.

Data Efficiency Needs

There is a critical need to avoid repeatedly scraping similar information to enhance performance and efficiency.

Goals and Objectives for Sports Chatbot

Key goals and objectives of the project

Main Goal: Functional Prototype

Develop a fully functional prototype Sports Chatbot to enhance user engagement.

Integrate Google Gemini LLM

Utilize Google Gemini LLM for advanced understanding and response generation in sports conversations.

Enable External Tool Usage

Empower the LLM to utilize external tools through function calling for enhanced capabilities.

Dynamic Web Scraping Tool

Build a web scraping tool using Selenium to gather real-time sports data efficiently.

Implement Semantic Caching

Use FAISS for semantic caching to reduce web scraping frequency and speed up responses.

User-Friendly Chat Interface

Create an intuitive chat interface using Streamlit for seamless user interaction.

System Effectiveness Evaluation

Conduct thorough evaluations to measure the system's effectiveness and user satisfaction.

System Architecture for Sports Chatbot

An Overview of the Chatbot's Structural Design

✓ **User Interaction via Streamlit UI**

The user engages with the system through a Streamlit interface for a seamless experience.

✓ **Agent Logic as the Core Processor**

Agent logic serves as the backbone, orchestrating requests between the UI and LLM.

✓ **Gemini LLM for Advanced Reasoning**

Utilizes Google Gemini LLM for reasoning, language processing, and tool integration.

Decision Point for Knowledge Retrieval

A critical juncture where the system decides to fetch internal knowledge or call a tool.

Use of Semantic Caching

FAISS and embeddings are leveraged for efficient semantic similarity searches.

Tool Calls for External Data

If external data is required, the system invokes tool calls to gather live information.

Semantic Cache Check for Efficiency

The system checks the cache to minimize redundant web scraping and speed up responses.

Handling Cache Hits and Misses

Cache hits provide instant responses, while misses trigger a web scraping process.

Web Scraping with Selenium and BeautifulSoup

Utilizes web scraping tools to fetch real-time data from the internet when needed.

Synthesized Response to User

The LLM synthesizes the final response and delivers it back to the user through the UI.

Modular Design for Scalability

The architecture's modular design allows for easy updates and scalability of components.

Key Technologies in LLM Chatbot

Semantic Cache: Google Embedding Model

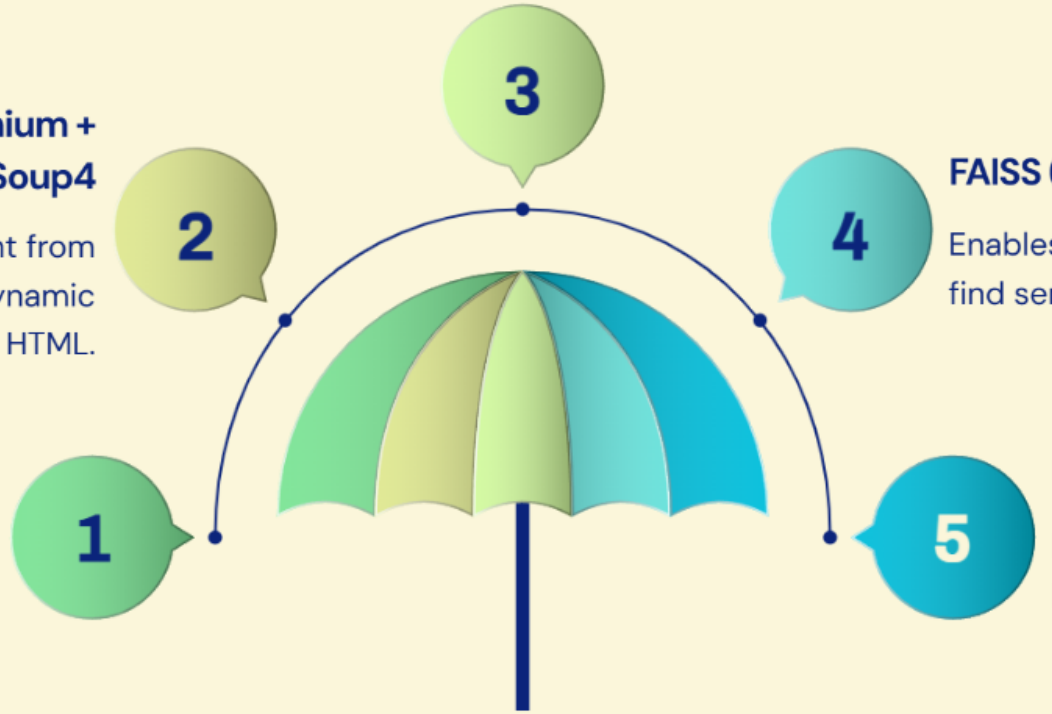
Generates vector representations to enhance understanding of user queries.

Web Scraping: Selenium + BeautifulSoup4

Efficiently extracts content from JavaScript-heavy sites, handling dynamic HTML.

LLM Core: Google Gemini (gemini-pro)

Utilizes strong language skills and supports native Function Calling for advanced interactions.



FAISS (Facebook AI Similarity Search)

Enables fast vector similarity searches to find semantically similar past queries.

User Interface: Streamlit

Facilitates rapid development of interactive Python web applications for users.

Implementation Highlights of LLM Chatbot

User Query Input

Users submit queries through the Streamlit chat interface for processing.

LLM Decision Making

Gemini assesses whether to use internal knowledge or to call external tools.

Cache Check Process

A check is performed to see if the information is cached for efficiency.

Cache Hit Outcomes

If a cache hit occurs, the stored result is returned instantly to the user.

Web Scraping on Cache Miss

If not cached, Selenium scrapes data from Bing to gather relevant information.

Text Extraction Technique

BeautifulSoup is used to extract text content from the fetched web pages.

Storing Results Efficiently

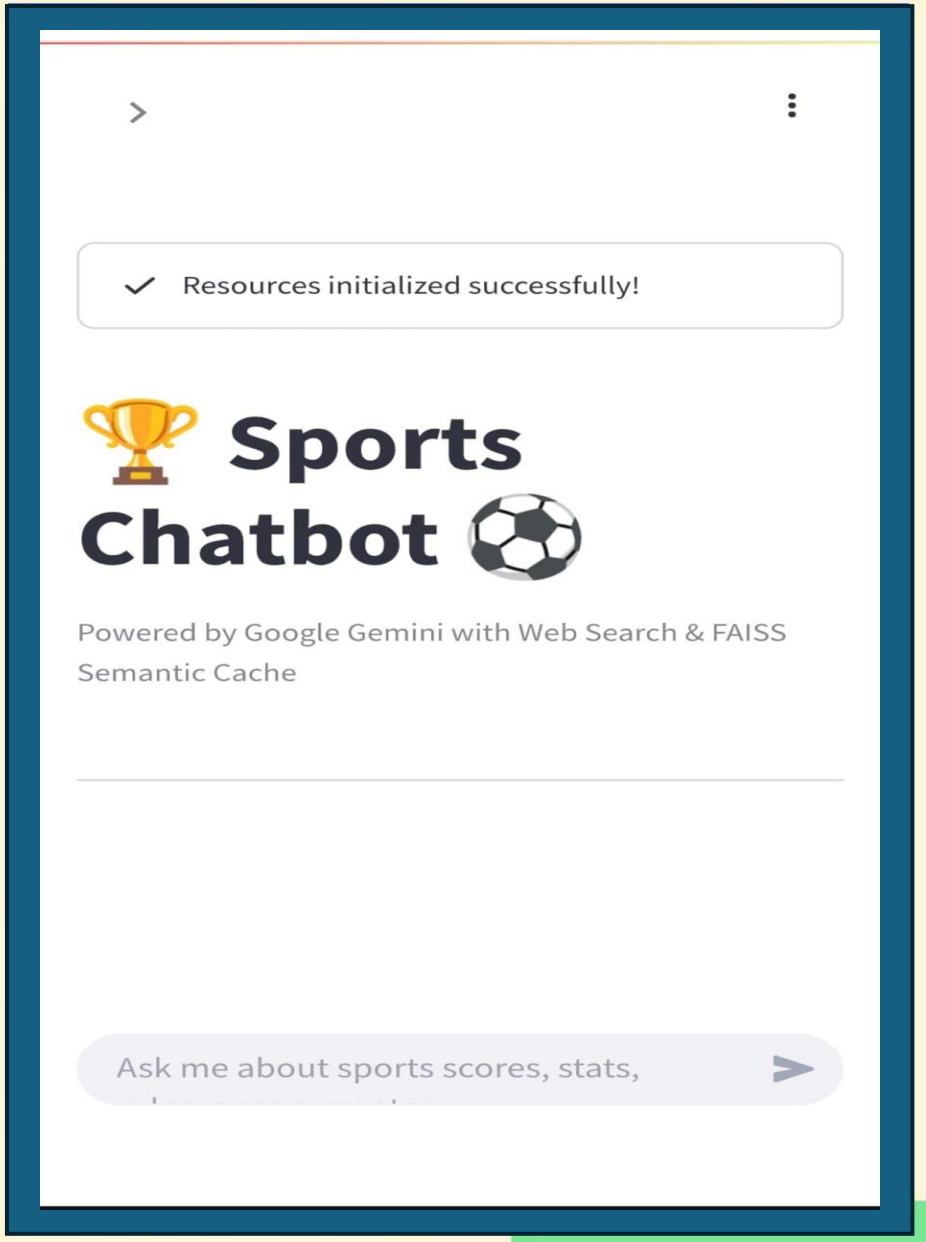
The scraped results are stored in a cache as a JSON file and indexed in FAISS.

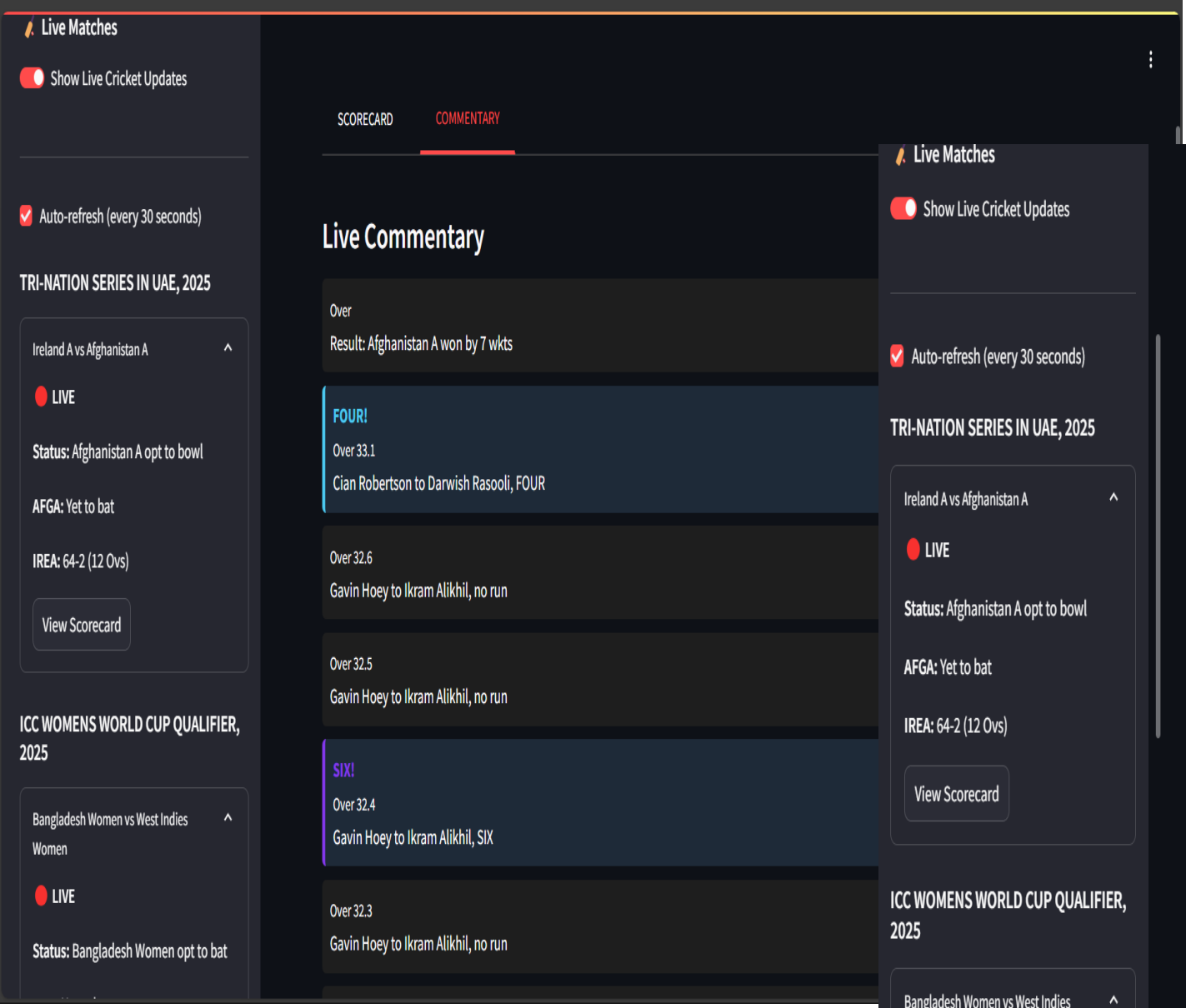
Final Response Generation

The LLM synthesizes an answer using either internal knowledge or retrieved data.

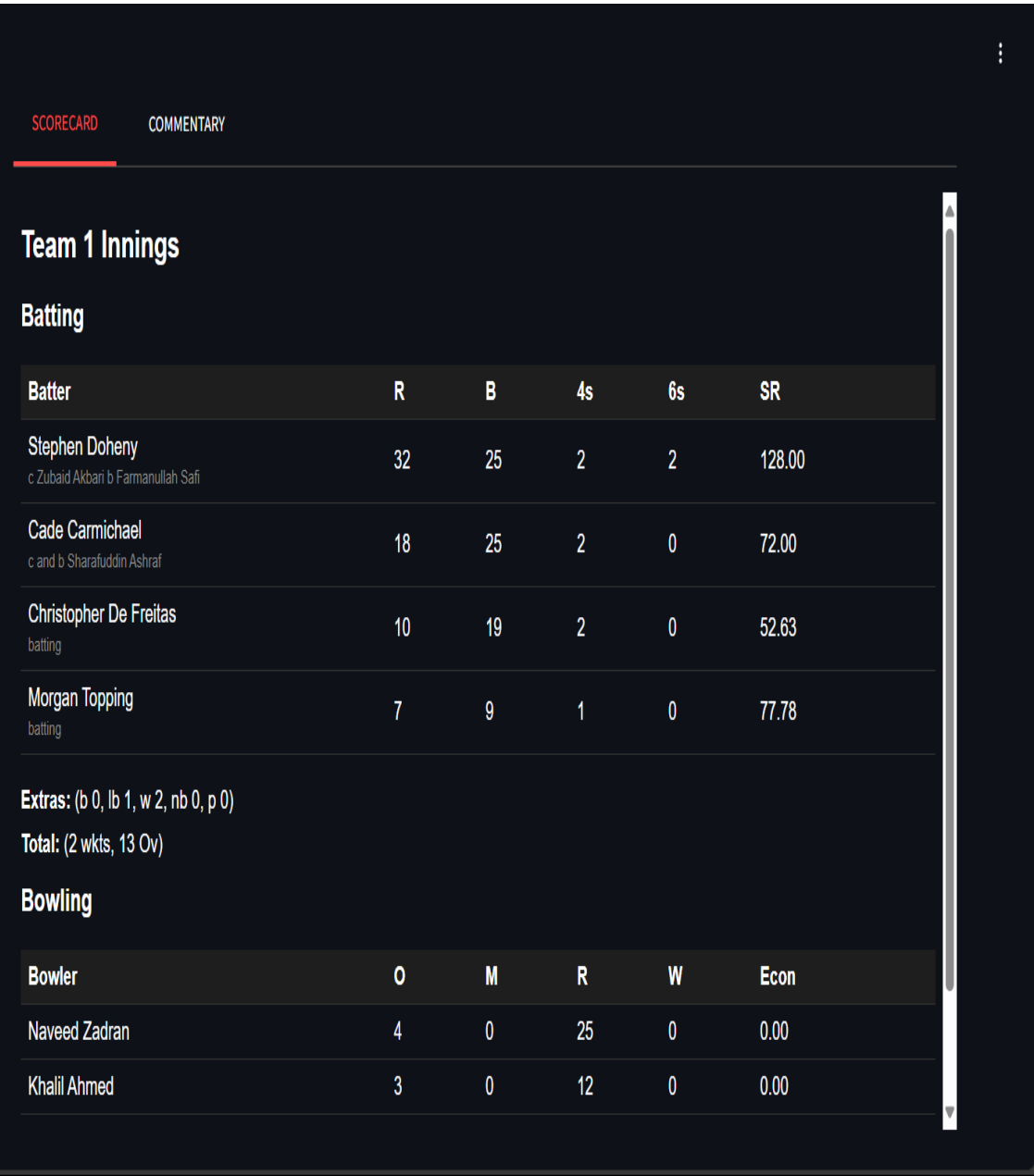
Results & Demonstration of Sports Chatbot

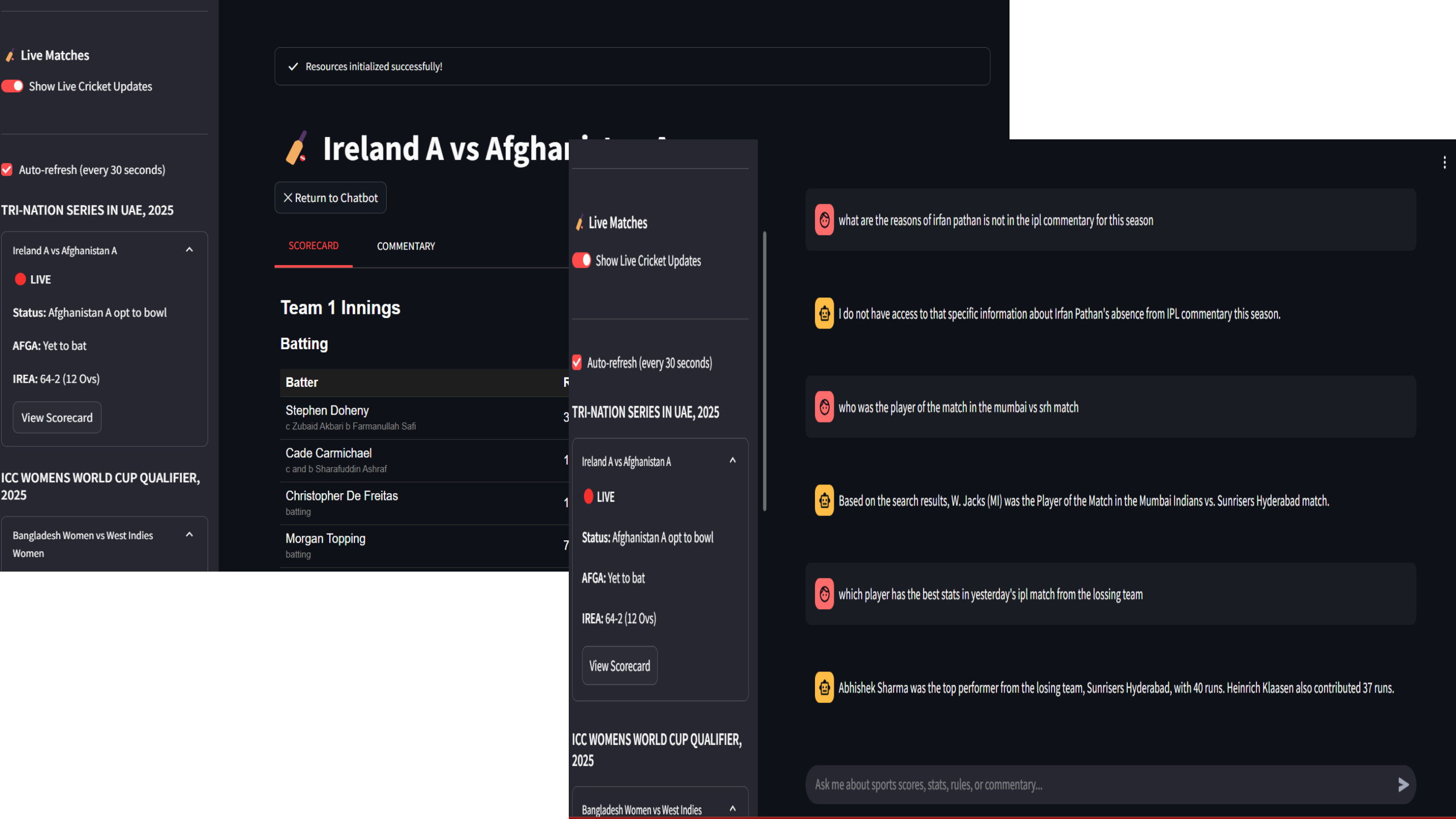
Explore the effectiveness of our sports chatbot in answering diverse queries, from rules to live scores, with impressive speed and cache efficiency.




Live Matches
Show Live Cricket Updates
SCORECARD
COMMENTARY
Auto-refresh (every 30 seconds)
TRI-NATION SERIES IN UAE, 2025
Ireland A vs Afghanistan A
LIVE
Status: Afghanistan A opt to bowl
AFGA: Yet to bat
IREA: 64-2 (12 Ovs)
View Scorecard
ICC WOMENS WORLD CUP QUALIFIER, 2025
Bangladesh Women vs West Indies Women
LIVE
Status: Bangladesh Women opt to bat
Live Commentary
Over
Result: Afghanistan A won by 7 wkts
FOUR!
Over 33.1
Cian Robertson to Darwish Rasooli, FOUR
Over 32.6
Gavin Hoey to Ikram Alikhil, no run
Over 32.5
Gavin Hoey to Ikram Alikhil, no run
SIX!
Over 32.4
Gavin Hoey to Ikram Alikhil, SIX
Over 32.3
Gavin Hoey to Ikram Alikhil, no run

Live Matches
Show Live Cricket Updates
Auto-refresh (every 30 seconds)
TRI-NATION SERIES IN UAE, 2025
Ireland A vs Afghanistan A
LIVE
Status: Afghanistan A opt to bowl
AFGA: Yet to bat
IREA: 64-2 (12 Ovs)
View Scorecard
ICC WOMENS WORLD CUP QUALIFIER, 2025
Bangladesh Women vs West Indies Women
LIVE
Status: Bangladesh Women opt to bat


SCORECARD
COMMENTARY
Team 1 Innings
Batting
Batter R B 4s 6s SR
Stephen Doheny
c Zubaid Akbari b Farmanullah Safi
32 25 2 2 128.00
Cade Carmichael
c and b Sharafuddin Ashraf
18 25 2 0 72.00
Christopher De Freitas
batting
10 19 2 0 52.63
Morgan Topping
batting
7 9 1 0 77.78
Extras: (b 0, lb 1, w 2, nb 0, p 0)
Total: (2 wkts, 13 Ov)
Bowling
Bowler O M R W Econ
Naveed Zadran 4 0 25 0 0.00
Khalil Ahmed 3 0 12 0 0.00


Live Matches
Show Live Cricket Updates
Auto-refresh (every 30 seconds)
TRI-NATION SERIES IN UAE, 2025
Ireland A vs Afghanistan A
LIVE
Status: Afghanistan A opt to bowl
AFGA: Yet to bat
IREA: 64-2 (12 Ovs)
View Scorecard
ICC WOMENS WORLD CUP QUALIFIER, 2025
Bangladesh Women vs West Indies Women
Resources initialized successfully!
Ireland A vs Afgha
Return to Chatbot
SCORECARD
COMMENTARY
Team 1 Innings
Batting
Batter
Stephen Doheny
c Zubaid Akbari b Farmanullah Safi
Cade Carmichael
c and b Sharafuddin Ashraf
Christopher De Freitas
batting
Morgan Topping
batting
Live Matches
Show Live Cricket Updates
Auto-refresh (every 30 seconds)
TRI-NATION SERIES IN UAE, 2025
Ireland A vs Afghanistan A
LIVE
Status: Afghanistan A opt to bowl
AFGA: Yet to bat
IREA: 64-2 (12 Ovs)
View Scorecard
ICC WOMENS WORLD CUP QUALIFIER, 2025
Bangladesh Women vs West Indies
what are the reasons of irfan pathan is not in the ipl commentary for this season
I do not have access to that specific information about Irfan Pathan's absence from IPL commentary this season.
who was the player of the match in the mumbai vs srh match
Based on the search results, W. Jacks (MI) was the Player of the Match in the Mumbai Indians vs. Sunrisers Hyderabad match.
which player has the best stats in yesterday's ipl match from the lossing team
Abhishek Sharma was the top performer from the losing team, Sunrisers Hyderabad, with 40 runs. Heinrich Klaasen also contributed 37 runs.
Ask me about sports scores, stats, rules, or commentary...

Performance Insights of Sports Chatbot

Understanding the capabilities and efficiency

Chatbot Functionality Overview

The chatbot effectively addresses a wide range of sports queries, enhancing user engagement.

Query Types Handled

Covers General Knowledge, Recent Scores, and Commentary Requests using LLM knowledge and web searches.

Response Speed Performance

Fast responses (1-5s) for LLM-only queries, while live scrapes take longer (15-40s).

Cache Effectiveness

Semantic caching reduces unnecessary web scrapes, improving overall performance.

Examples of Query Handling

Showcases various query types with screenshots for clarity and demonstration purposes.

Enhancing Future Work Strategies

Strategic Recommendations for LLM Sports Chatbot

Replace Selenium with APIs

Switch to official Search APIs for improved reliability and maintenance.

Use requests/httpx + BeautifulSoup

Leverage these libraries for content scraping and processing efficiently.

Consider Scraping API services

Evaluate dedicated services for scalable and efficient data extraction.

Enhance Caching with Cloud Vector DB

Employ Pinecone or Vertex AI for persistence and scalability in caching.

Implement Time-To-Live (TTL)

Set TTL for cache entries to optimize storage and performance.

Add Persistence for Chat History

Store chat history in a database to improve user experience and data tracking.

Improve User Experience

Focus on source attribution, streaming responses, and interactive voice features.

Thank You!!!

We are open to questions now.