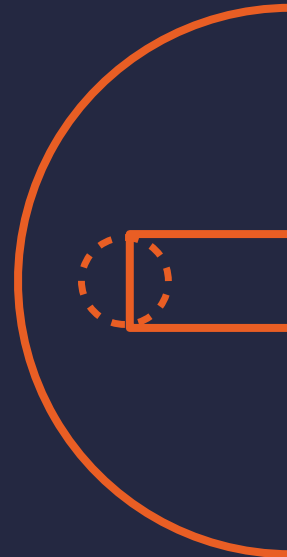


Hoops IQ

**A WNBA Intelligence
Tool**



Meet the Team

Anoop
Nair



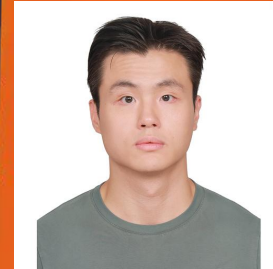
lishaan
Shekhar



Akaash
Venkat



I-Hsiu
Kao



Agenda



1

2



3



4



5

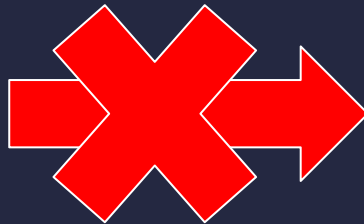
What's the Problem?

WNBA Growth

2.2 New Billion Dollar Media Deal



User-Friendly, Accurate



Extraction

WNBA Data is extremely difficult to extract for non-basketball, data savvy users



Accuracy

Publicly available LLM's consistently report false information

Off the Shelf LLM's - Hallucinations

e-2b28051c8a64

Which player has the highest free-throw percentage in WNBA 2018 season?

Searched 6 sites

In the 2018 WNBA season, Elena Delle Donne had the highest free-throw percentage. She achieved a remarkable free-throw percentage of 95.0%, setting a WNBA record for the highest free-throw percentage in a season with at least 100 attempts. Delle Donne made 207 out of 218 free throws, showcasing her exceptional skill and consistency from the line (WNBA Official) (WNBA Official) (WNBA Official).

proballers.com/basketball/player/181977/elena-delle-donne

ELENA DELLE DONNE STATS

Profile Games Totals Splits Records Teammates

OPPONENT	DATE	LEAGUE	SCORE	PTS	REB	AST	MIN
@ NEW	Sep 20, 2023	WNBA	L 90-85	11	4	1	38
@ NEW	Sep 16, 2023	WNBA	L 90-75	11	5	6	35
vs ATL	Sep 9, 2023	WNBA	L 75-80	13	6	1	26
@ PHO	Sep 6, 2023	WNBA	W 77-100	24	4	0	24
@ LOS	Sep 4, 2023	WNBA	L 72-64	8	6	1	29

stats.wnba.com/leaders/?Season=2018&SeasonType=Regular%20Season&CF=FT_PCT*G*90

ADVANCED STATS Stats Home Players Teams Scores Schedule Standings Quick Links

2018 FT% > 90

#	PLAYER	GP	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	O
3	DIANA TAURASI	33	30.0	20.7	6.1	13.7	44.6	3.2	8.4	38.3	5.2	5.6	92.5	
8	KAYLA MCBRIDE	31	32.3	18.2	6.4	14.2	44.9	1.5	3.9	39.3	3.9	4.3	91.7	
23	KRISTI TOLIVER	33	32.9	13.9	4.5	10.4	43.3	2.2	6.1	36.0	2.7	2.9	91.8	
39	STEFANIE DOLSON	27	27.6	9.7	4.0	8.5	46.7	0.7	1.9	34.6	1.1	1.2	93.8	
48	VICTORIA VIVIAN	34	27.1	8.9	3.1	7.6	40.4	1.9	4.8	39.9	0.8	0.9	93.1	
57	ESSENCE CARSON	34	23.1	7.5	3.0	6.8	43.5	1.1	3.2	36.1	0.4	0.5	93.8	

ChatGPT 4o

"What were the total points scored by Elena Delle Donne in the last five games?"

Searched 3 sites

Elena Delle Donne has been performing exceptionally well in her recent games for the Washington Mystics. Here are her total points scored in the last five games:

- 21 points against the Las Vegas Aces.
- 15 points against the Atlanta Dream.
- 19 points in another match against the Aces.
- 25 points against the Seattle Storm.
- 18 points against the Phoenix Mercury.

This gives her a total of 98 points over the last five games (Fox Sports) (WNBA Official).

Target User - Daily Fantasy Players

Daily Fantasy User Selection

1. Allows us to refine scope to a user group interested in Statistical Q/A
2. We address needs by creating a chatbot that not only provides reliable responses, but also offers advanced visualizations



What's the Solution?

Hoops IQ - A Language Model that answers WNBA specific statistics questions accurately for users to leverage to optimize for daily fantasy gameplay.

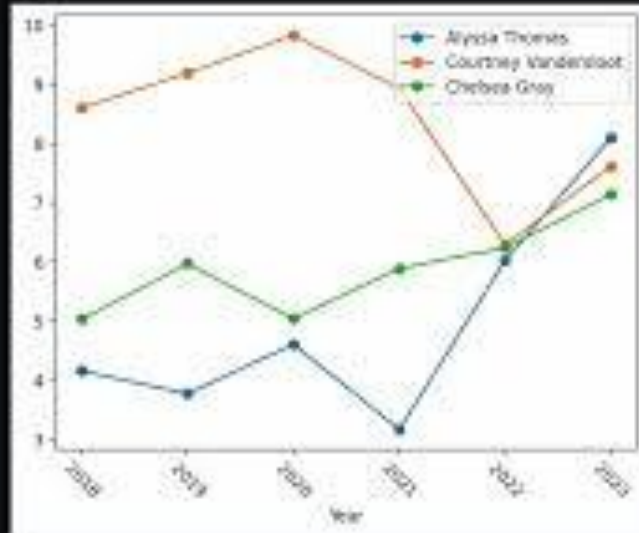
wnba.hoops-iq.com



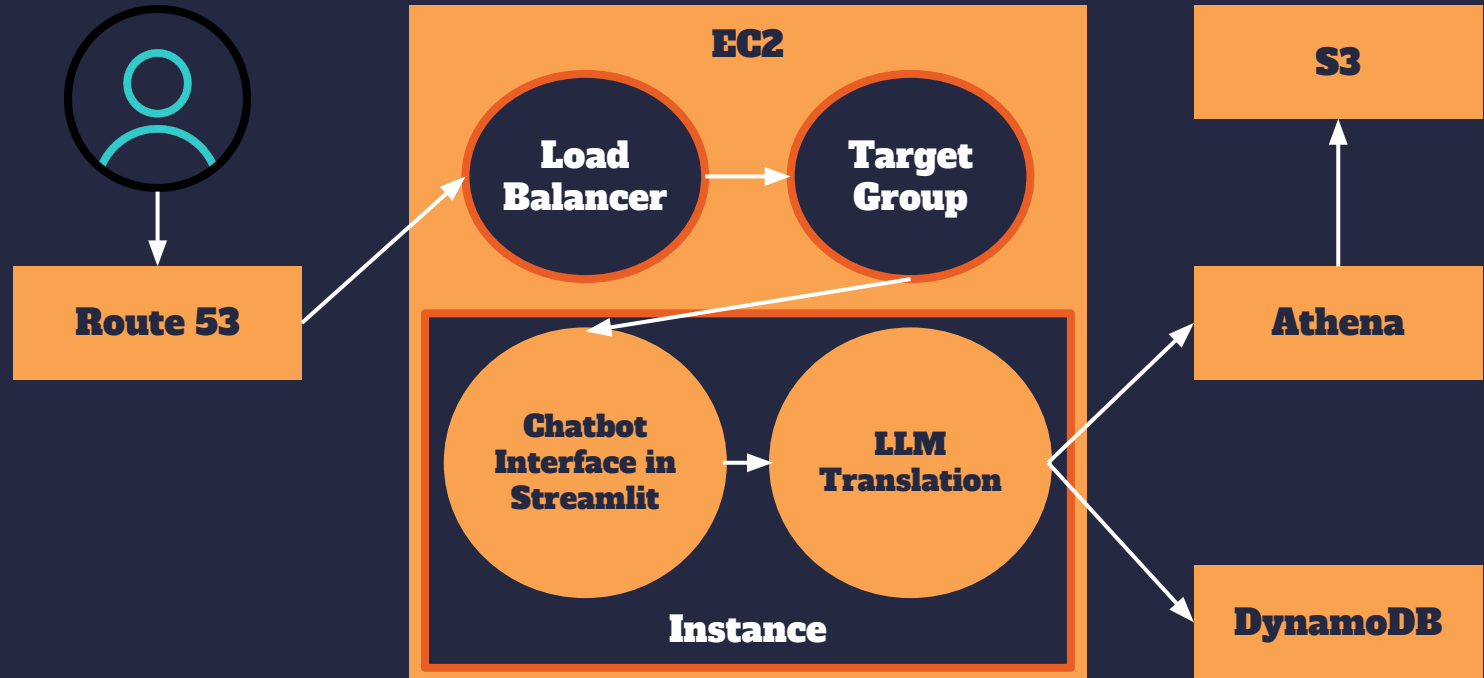
Demo

Courtney Vandenoort and Chelsea Gray

Visualization:



Tech Overview



Data

API's



Data

Metrics

1. **Box: Points, Rebounds, Assists, Steals, Blocks**
2. **Player Info: Height, Years Active, Career Stats**
3. **Schedule: Date, Attendance, Result (Win/Lose)**

Grains

The Dataset was built to be robust to both multiple time dimensions and player/team level assessments

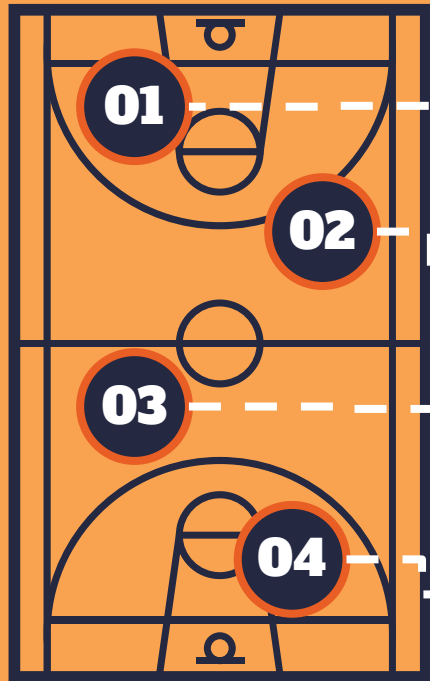
Time

1. **Event**
2. **Game**
3. **Season**
4. **Career**

Type

1. **Player**
2. **Team**

Dataset Creation



**Data
Download**

**Data
Cleaning**

**Staging
Layer**

- Data Tables



PBP

Team Box

Player Box

Schedule

Player Information

RAG Architecture

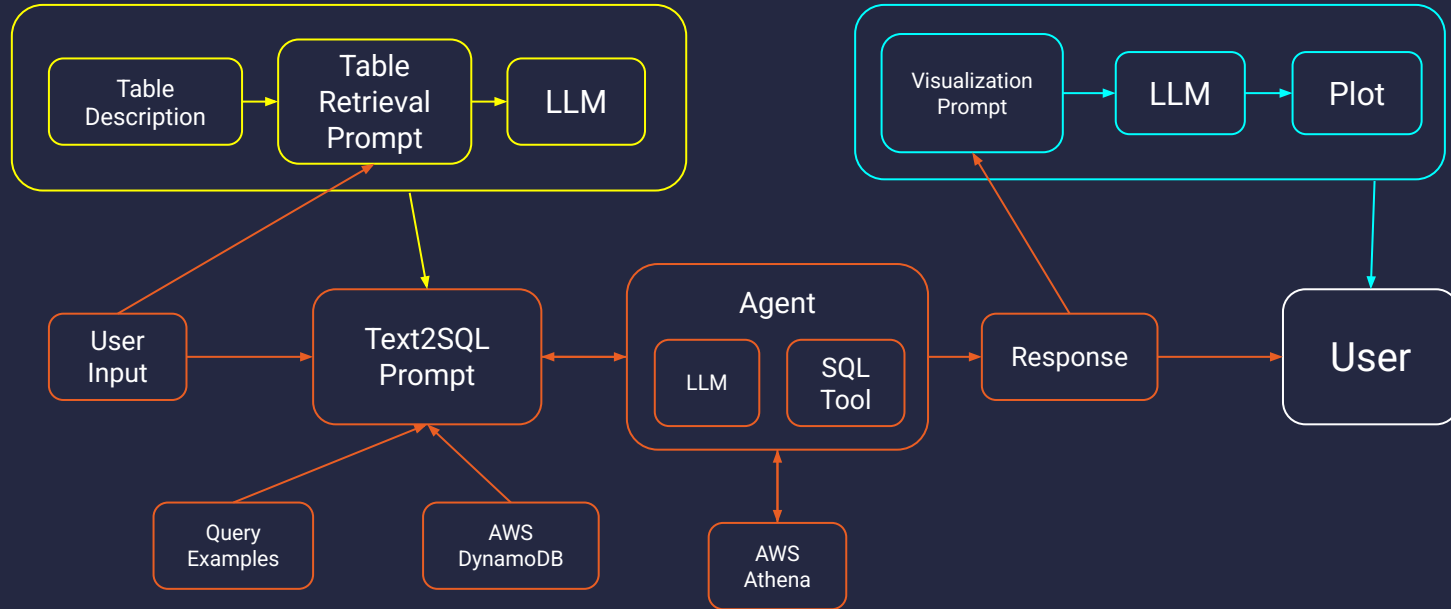
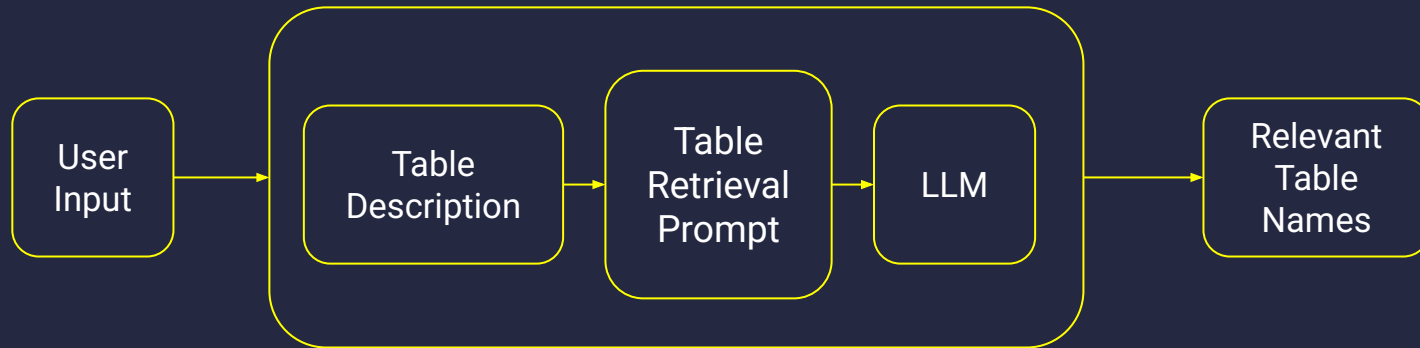


Table Retrieval Chain

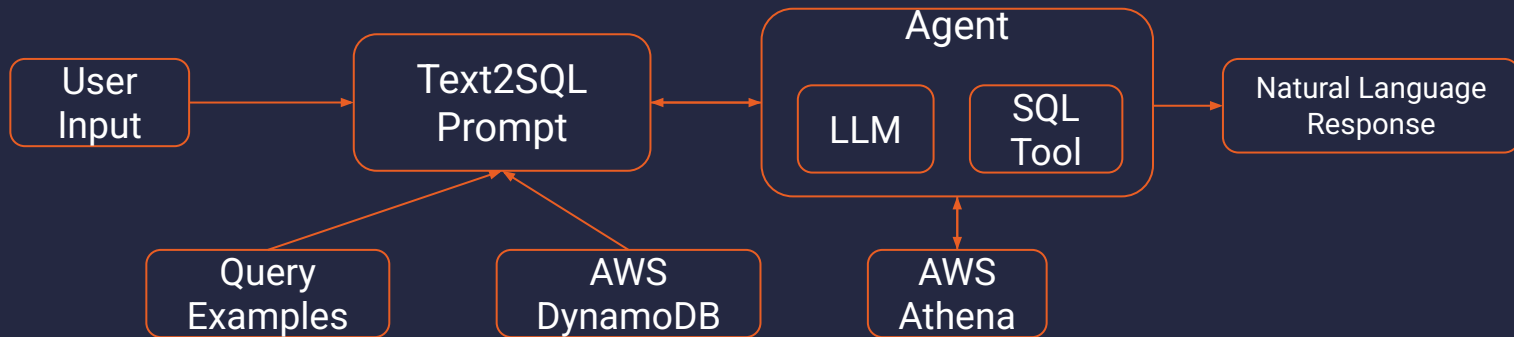


Q: *Which player had the most steals in the 2023 season?*

Without Table Context: I encountered an issue with retrieving the player's name who had the most steals in the 2023 season. However, I can provide you with the player's ID. The player with the most steals in the 2023 season is player with ID 2047.

With Table Context: `select sum(steals), athlete_display_name from wnba_player_box where season = 2023 group by athlete_display_name order by 1 desc limit 1;`

Text2SQL Chain



Q: *Which player had the highest free throw percentage in the 2020 season (min. 25 FT's attempted)?*

Without Ex: `SELECT player_name, free_throw_percentage FROM player_stats WHERE season = 2020 AND free_throws_attempted >= 25 ORDER BY free_throw_percentage DESC LIMIT 1`

With Ex: `SELECT athlete_display_name, SUM(free_throws_made) / SUM(free_throws_attempted) AS free_throw_pct FROM wnba_player_box WHERE season = 2020 GROUP BY athlete_display_name HAVING SUM(free_throws_attempted) >= 25 ORDER BY free_throw_pct DESC LIMIT 1;`

SQL DB vs Vector Store

Q: Which player made an 18-foot step back jumpshot in 2018?

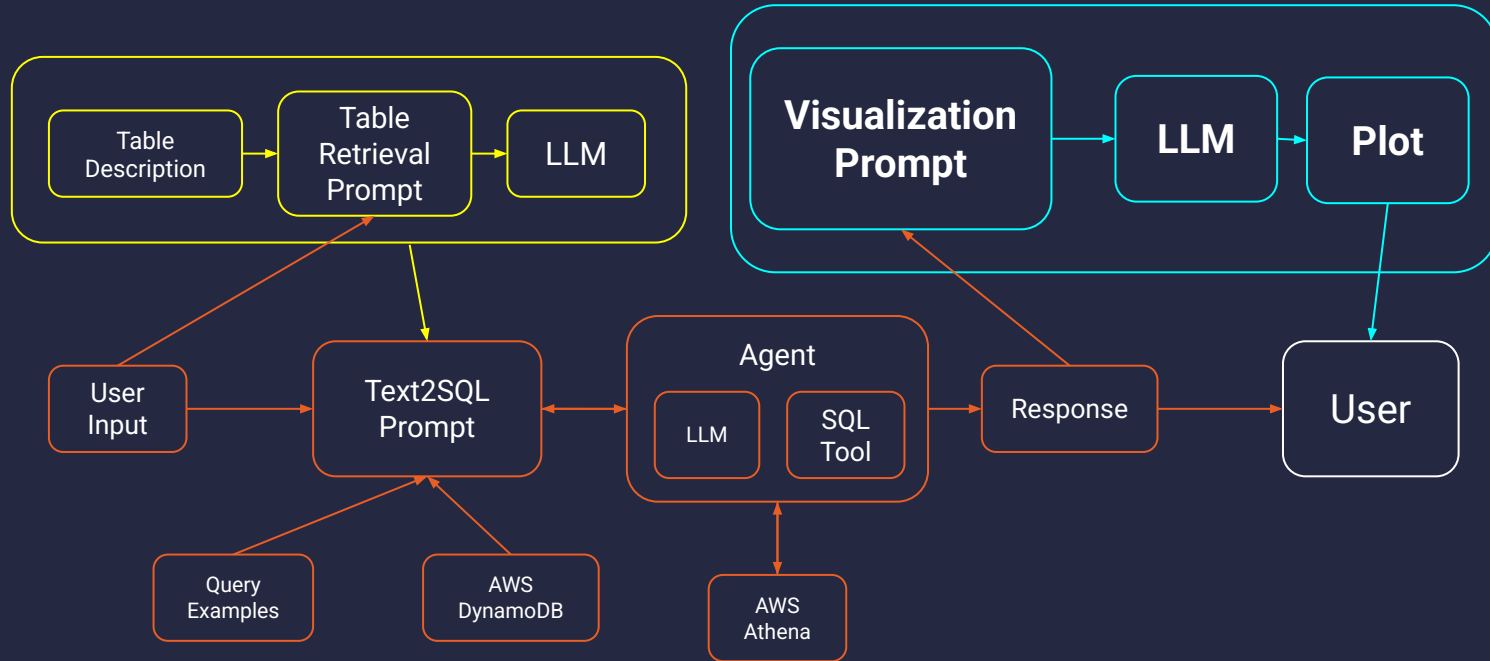
Vector Store

- Maya Moore makes **16-foot** step back jumpshot (Rebekkah Brunson assists)
- Jonquel Jones makes **26-foot** step back jumpshot (Courtney Williams assists)
- Jewell Loyd makes 18-foot step back jumpshot (Sue Bird assists)
- Kristi Toliver makes **21-foot** step back jumpshot
- Kristi Toliver makes 18-foot step back jumpshot (Elena Delle Donne assists)

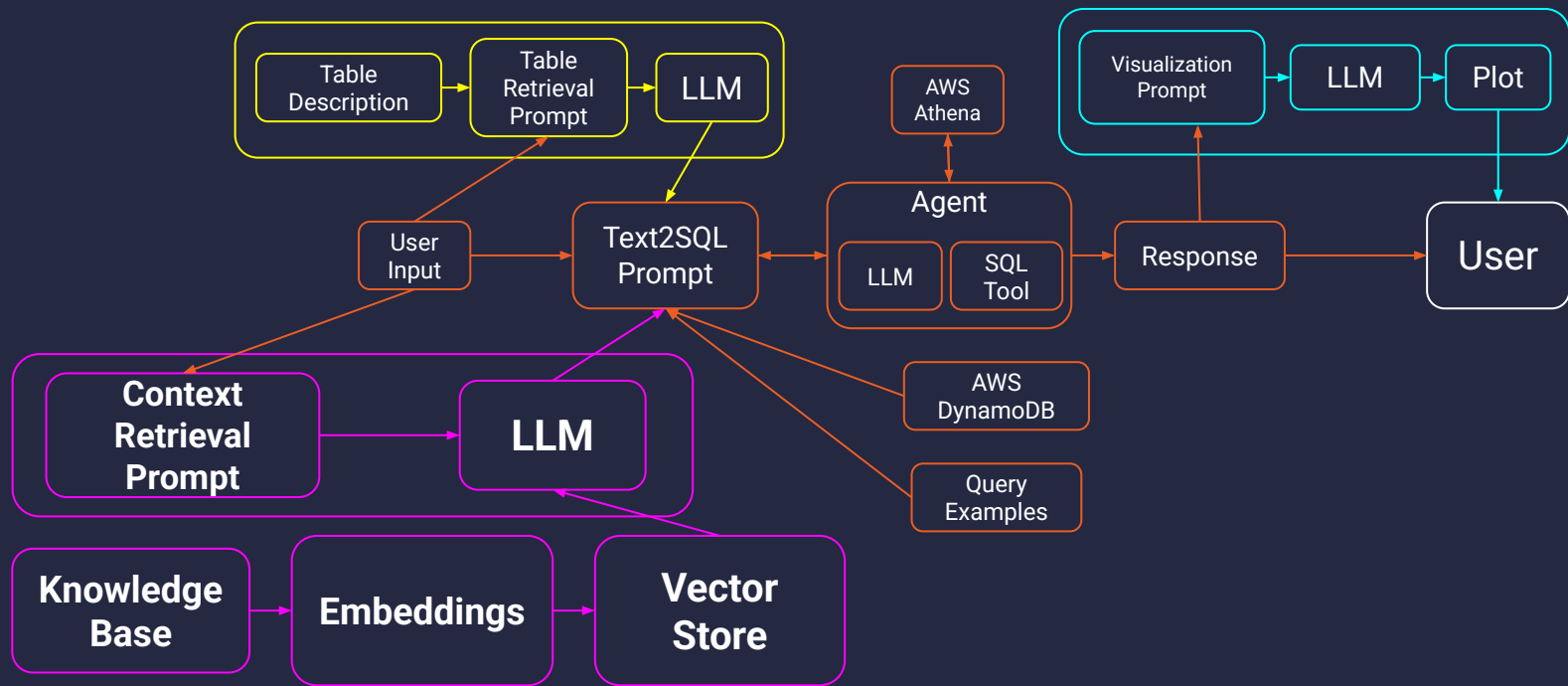
Text2SQL

- In 2018, the players who made an 18-foot step back jumpshot were: Kristi Toliver (assisted by Elena Delle Donne) and Jewell Loyd (assisted by Sue Bird)

RAG Architecture



Future RAG Architecture



How to Measure

Method

Assess SOTA LLM's ability to respond *accurately* to a gold-standard WNBA question set, and create a cost function for *hallucinations* from the LLM (to penalize incorrect data retrieval).

Core KPI: Accuracy

$$\text{Accuracy} = \frac{\text{Right Answers} - \text{Hallucinations}}{\text{Total Questions} = 38}$$

Measurement

Core KPI: Accuracy* (-Hallucinations)

<i>Model</i>	<i>Baseline</i>	<i>Hoops IQ</i>
GPT 3.5 Turbo	-58% (-25)	16% (-13)
GPT 4 Turbo	-11% (-12)	45% (-9)
GPT 4o**	-24% (-18)	47% (-9)
GPT 4o mini	-32% (-21)	42% (-9)
Claude 3.5 Sonnet	0% (-8)	45% (-10)

Results - Strengths

Text2SQL Strengths - General Statistics

Question: *"Which player played the most minutes in the 2021 season?"*

Accuracy: Text2SQL: **100%**

Regular LLM: **-100%**

The player who played the most minutes in the 2021 WNBA season was **Skylar Diggins-Smith**, with a total of 1440 minutes.

In the 2021 WNBA season, the player who played the most minutes was **Courtney Vandersloot** of the Chicago Sky. Vandersloot logged a total of 1,082 minutes over the course of the season.

In the 2021 WNBA season, the player who logged the most minutes was **Courtney Vandersloot** of the Chicago Sky. She played a total of 1,074 minutes over the course of the regular season.

In the 2021 WNBA season, the player who logged the most minutes was **DeWanna Bonner** of the Connecticut Sun. She played a total of 1,034 minutes over the course of the season, showcasing her durability and importance to her team.

Future Path - Model Tuning

Our analysis uncovered opportunities for improving the Text2SQL architecture; the following represent the top priorities for future enhancements to our model.

**Nuanced Terminology (e.g.
Defensive Rating, Usage Rate)**

**Predictive Capabilities (e.g.
forecasted performance)**

**Large Request Synthesis (e.g.
multi-year comparisons across
several WNBA athletes)**

Future Path - New User Groups



Fantasy

01

Comprehensive Dataset

Can we build this robust to all needs (betting lines, seasons, new kpi's)?



Front Office

02

WNBA CBA Details

Can we capture player value & salary details?



Coaching

03

Hawk Eye Data

Can we capture spatial information?

Mission Statement

*Empowering fantasy women's basketball players with **accessible, accurate WNBA data** and **advanced analytics** to make informed decisions, enhance their understanding of the game, and gain a competitive edge in their fantasy leagues.*

