



# BASNya

Basketball Auto-  
generated  
Statistical News

# Суть проекта

Автоматическая генерация твитов на основе аномальных\интересных выступлений игроков НБА

Демо-стенд проекта: <https://basnya.streamlit.app>  
Пароль: basnya\_weak\_password

Код проекта: <https://github.com/no-one2k/basnya>



# Команда



Алексей  
LLM Center



Богдан  
Signal Forward



Борис  
DB Guard



# Для кого\чего мы делали проект



Coach Nick

- Nick Hauselman
- твиттер: 144.5K
- ютуб: 866K



# Цели проекта

- Автоматизировать создание контента на стыке баскетбола и статистики
  - } Меньше трудозатрат
  - } Чаше публикации
  - } Меньше пауза между окончанием игры и выходом контента
  - } Больше разнообразия
  - } Шире охват команд и игроков
- Освоение техник и инструментов построения ML систем
  - } ML Design Doc как методологический подход
  - } LLM
  - } Инструменты для ETL, ML Observability и прототипирования UI

# Основной сценарий использования

- 1) Пользователь авторизуется и выбирает дату, за которую он хочет проанализировать игры
- 2) Игры для этой даты загружаются в хранилище
- 3) Пользователь выбирает какие игры ему хочется проанализировать
- 4) Для выбранных игр происходит поиск потенциально интересных фактов
- 5) По этим фактам генерируются твиты
- 6) Пользователь публикует твит(ы) в исходном виде или проводит дополнительное ручное исследование на их основе

# Основной сценарий использования

Password



# Основной сценарий использования

## BASNya Demo

enter date

23.12.2023

Get games



# Основной сценарий использования

Select	unseen_game	game_id	game_date	teams	score
<input checked="" type="checkbox"/>	<input type="checkbox"/>	0022300388	12/23/2023	Milwaukee Bucks - New York Knicks	130 - 111
<input type="checkbox"/>	<input type="checkbox"/>	0022300389	12/23/2023	Boston Celtics - LA Clippers	145 - 108
<input checked="" type="checkbox"/>	<input type="checkbox"/>	0022300390	12/23/2023	Charlotte Hornets - Denver Nuggets	95 - 102
<input type="checkbox"/>	<input type="checkbox"/>	0022300391	12/23/2023	Orlando Magic - Indiana Pacers	117 - 110
<input type="checkbox"/>	<input type="checkbox"/>	0022300392	12/23/2023	New Orleans Pelicans - Houston Rockets	104 - 106
<input checked="" type="checkbox"/>	<input type="checkbox"/>	0022300393	12/23/2023	Memphis Grizzlies - Atlanta Hawks	125 - 119
<input type="checkbox"/>	<input type="checkbox"/>	0022300394	12/23/2023	Detroit Pistons - Brooklyn Nets	115 - 126
<input type="checkbox"/>	<input type="checkbox"/>	0022300395	12/23/2023	Toronto Raptors - Utah Jazz	119 - 126
<input type="checkbox"/>	<input type="checkbox"/>	0022300396	12/23/2023	Cleveland Cavaliers - Chicago Bulls	109 - 95
<input type="checkbox"/>	<input type="checkbox"/>	0022300397	12/23/2023	Los Angeles Lakers - Oklahoma City Thunder	129 - 120
<input type="checkbox"/>	<input type="checkbox"/>				

Generate tweets

# Основной сценарий использования

🇷🇺🏀 BIG changes in the Season Top 10 players by MIN! 🇷🇺 Nikola Jokic from Serbia has entered the top with a new rating of 7, while Anthony Davis from USA has been knocked out with a new rating of 11. 🤖🔥 #NBA #MIN #FGM #FGA #FG3M #FG3A #FTM #FTA #REB #AST #STL #BLK #TO #PF #PTS #PLUS\_MINUS

"🇷🇺 Huge changes in the Season Top 10 players by FGM rating! Jalen Brunson from 🇺🇸 has entered the top with a new value of 255 and a new rating of 9, while Kyle Kuzma from 🇺🇸 has sadly dropped out with the same old value but a lower new rating of 12. #NBA #FGM #Top10 🏀"

Copy to clipboard

"🇷🇺 Huge shakeup in the Season Top 10 players by FGA! 🏀 Desmond Bane and Jalen Brunson enter the top with impressive performances, while Jayson Tatum and Tyrese Maxey drop out. 📈 Keep an eye on these rising stars! #NBA #FGA #MIN #FGM #FGA #FG3M #FG3A #FTM #FTA #REB #AST #STL #BLK #TO #PF #PTS #PLUS\_MINUS"

# Что мы строили

Система сбора баскетбольной статистики

+

Анализ и выявление потенциально интересных фактов

+

Генерация контента



# Предположения и ограничения

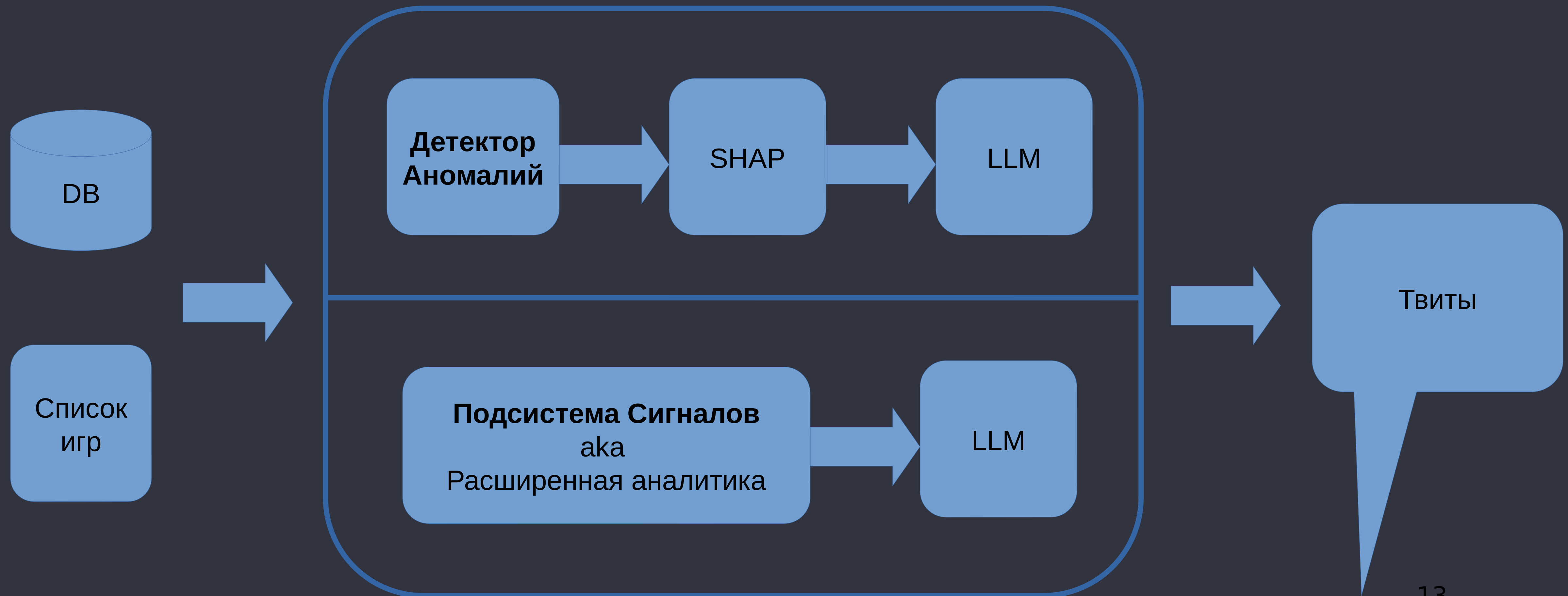
## 1. Знания о пользователе

- англоязычный
- владеет python и средствами анализа данных

## 2. Знания о задаче:

- разработка и эксплуатация должны быть дешёвыми
- 1 игровой день обрабатывается менее, чем за 1 час
- интересность контента сложно оценить => нужен отбор человеком

# Карточка модели. Схема



# Карточка модели. Детали

1. Данные получены с портала `stats.nba.com` с помощью пакета ``nba_api``
  - для обучения детектора аномалий — протоколы игр регулярных сезонов НБА 2021-22 и 2022-23
  - для подсистемы сигналов — протоколы игр текущего сезона НБА (2023-24)
2. стек: `python + prefect + sklearn + shap + langchain + openAI + streamlit`
3. Fairness
  - не используем пол\расу\национальность\возраст
  - сторонняя LLM как чёрный ящик



# Какие были сложности

1. Слишком широкая постановка задачи в самом начале
2. Отсутствие прямых метрик для более узкой задачи: как измерить и сравнивать интересность?
3. PYOD и SHAP не дружат между собой
4. Отсутствие документации на данные
- 5\*. Галлюцинации и некорректное поведение LLM
- 6\*. Малое разнообразие и «искусственность» контента

Спасибо за  
внимание

