

Student Team

Uchit N M & Arya Ajay Gupta

Department of Computer Science and Engineering, PES University

Mentors from Ittiam

Mr. Shailesh Ramamurthy, Mr. Murali Babu Muthukrishnan and
Mr. Arunkumar Mohananchettiar

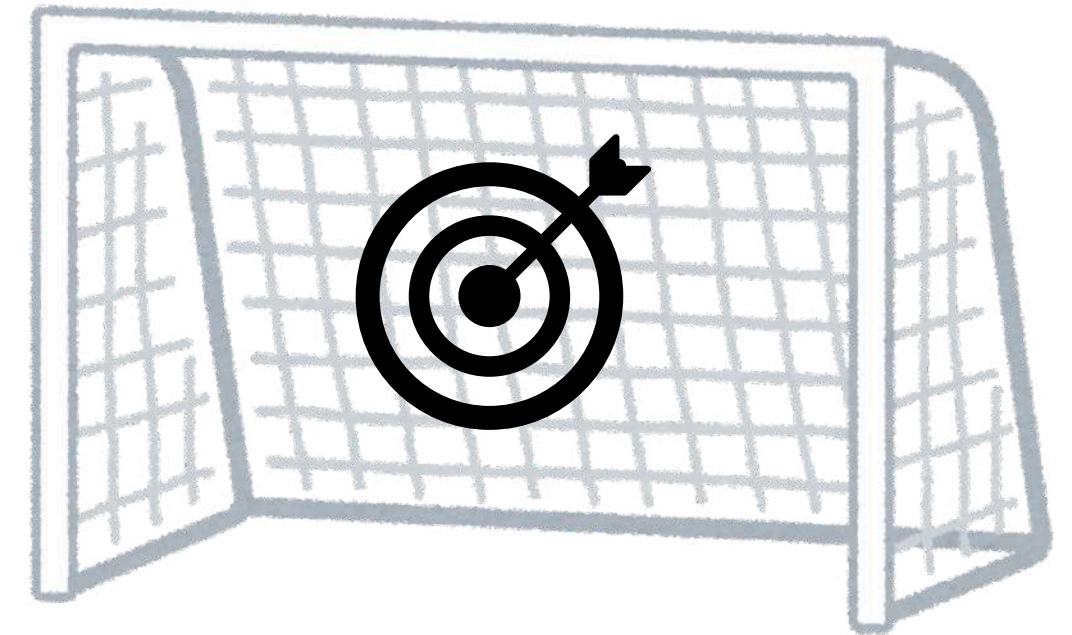
Mentors from PES University

Prof. Prasad H B, Dr. Gowri Srinivasa, Prof. Preet Kanwal

PESU Research Center for Information Security, Forensics and Cyber Resilience
Department of Computer Science and Engineering, PES University

Objective

To develop a system to summarize the soccer match videos.



Domain Knowledge

1. Game Structure:

- Played between two teams of 11 players on a standard 105m x 68m pitch with designated zones (penalty area, goal area, center circle).

2. Match Duration:

- Regulation Time: Two halves of 45 minutes each, with a 15-minute halftime break.
- Extra Time: In knockout stages, if tied, two 15-minute halves are played.
- Penalty Shootout: If still tied, a penalty shootout determines the winner.

3. Ball in Play:

- The ball remains in play unless it crosses the touchline (throw-in) or goal line (goal, corner kick, or goal kick).

4. Governing Body:

- Governed globally by FIFA, with regional federations like UEFA, CONMEBOL, and AFC overseeing competitions.



Domain Knowledge

5. Goals:

- A goal is scored when the ball fully crosses the goal line within the goal frame.

6. Misconducts:

- Yellow Card: Issued as a caution for unsporting behavior or rule violations.
- Red Card: Given for severe offenses, leading to player expulsion.

7. Restarts and Set Pieces:

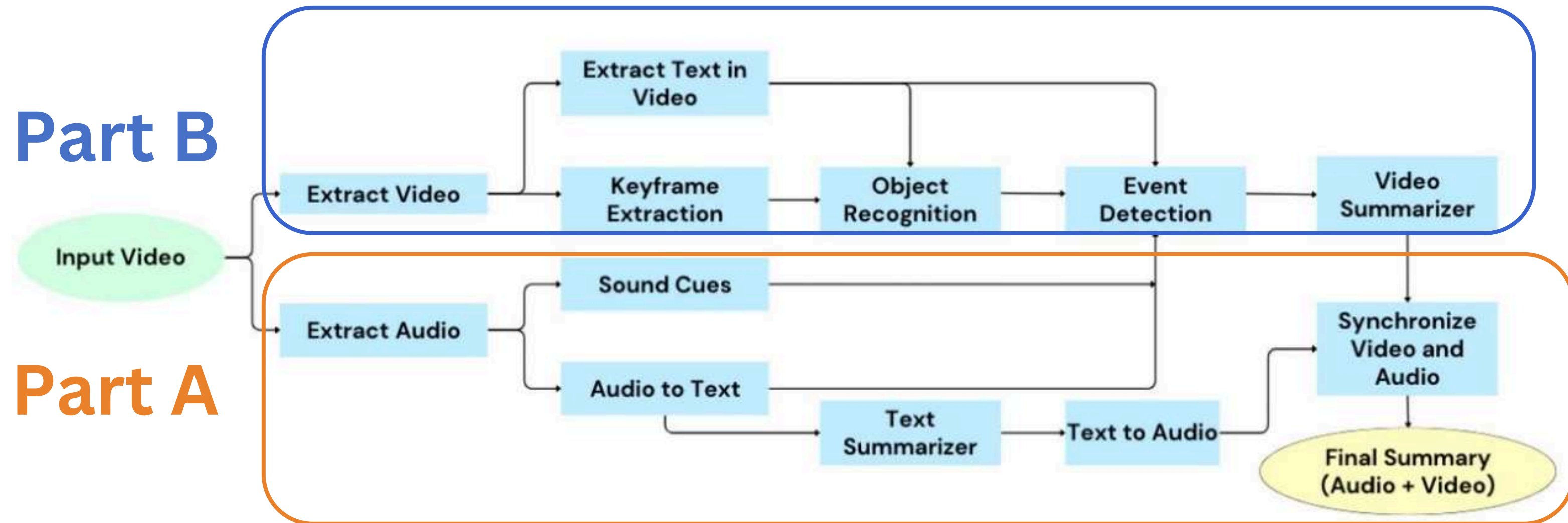
- Kick-Off: Used to start each half or restart play after a goal.
- Throw-In: Awarded when the ball crosses the touchline, last touched by an opponent.
- Corner Kick: Given when the defending team last touches the ball before it crosses their goal line.

8. Free Kicks and Penalties:

- Direct Free Kick: Given for penal fouls, allowing a direct shot on goal.
- Penalty Kick: A one-on-one opportunity from the penalty spot after a foul inside the penalty area.



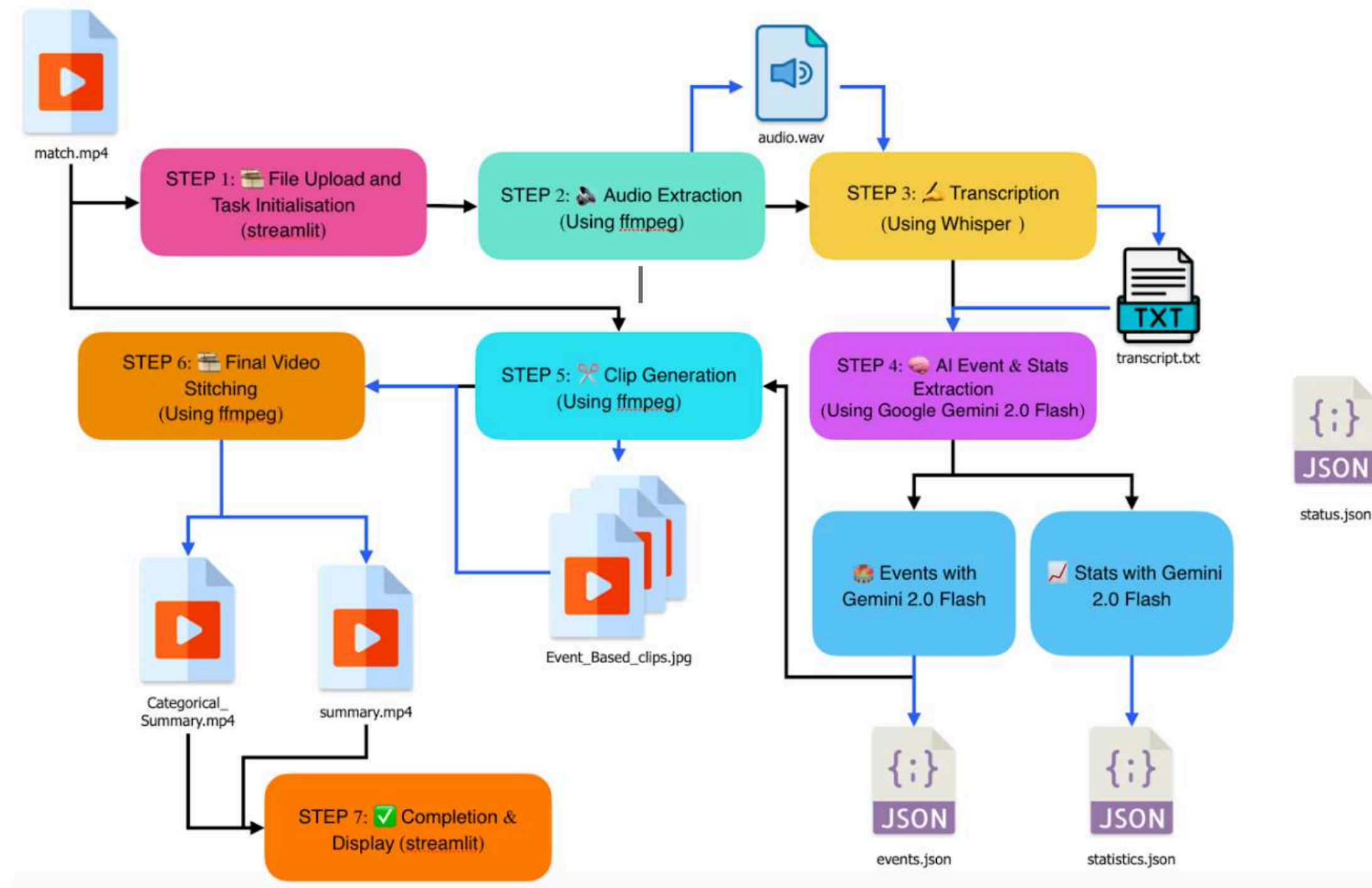
Proposed Framework



PART A : Audio Exploration



System Architecture



Building an End-to-End System and Assess the challenges

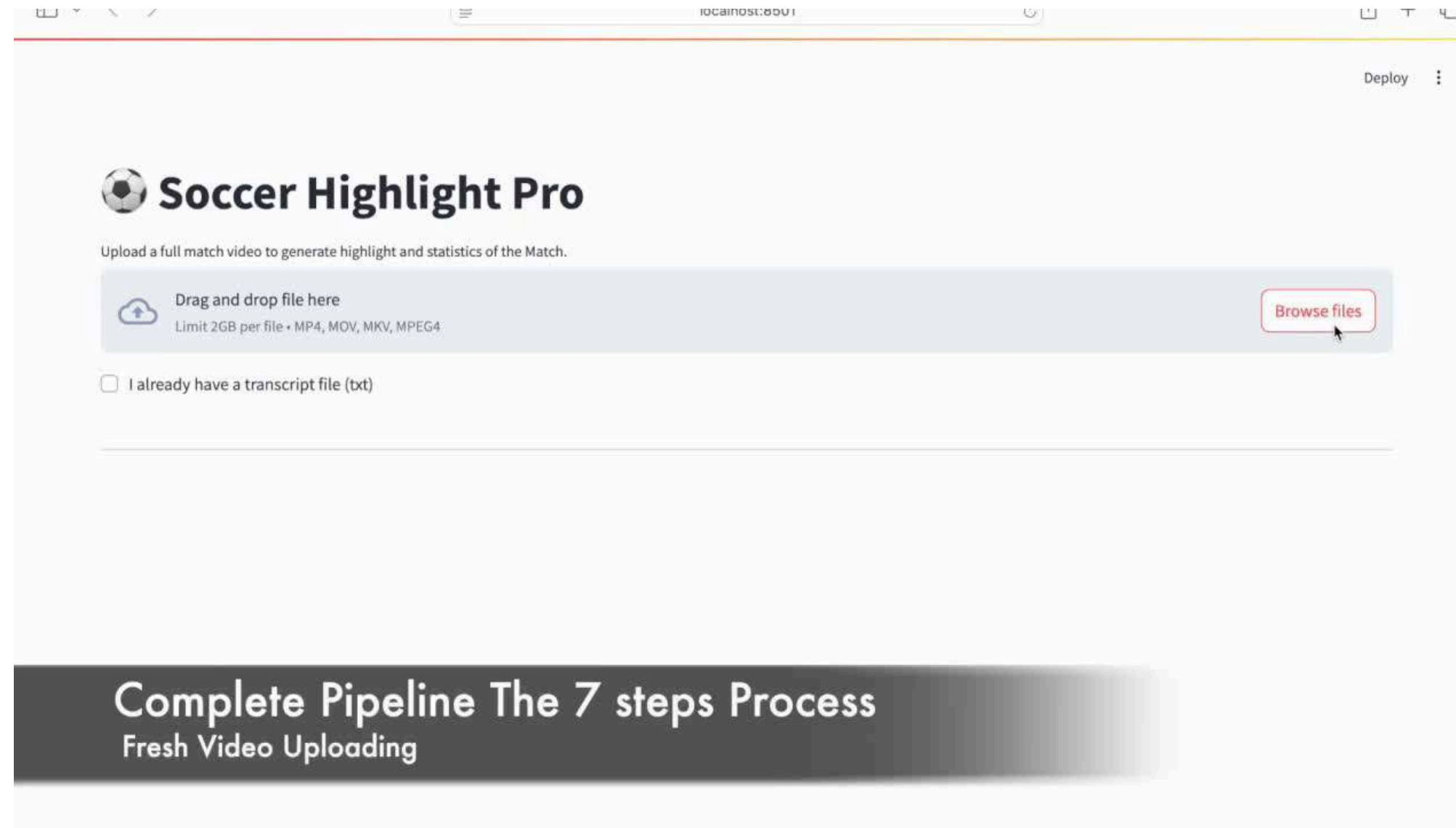
3 key functional blocks:

- The audio transcript generator
- The Gemini-based event timestamp extractor, and
- The video clipping and stitching component.

Key Challenges:

- Maintaining structural integrity across modules
- Parsing and structuring Gemini's output
- Performance bottleneck in long video processing

Demo

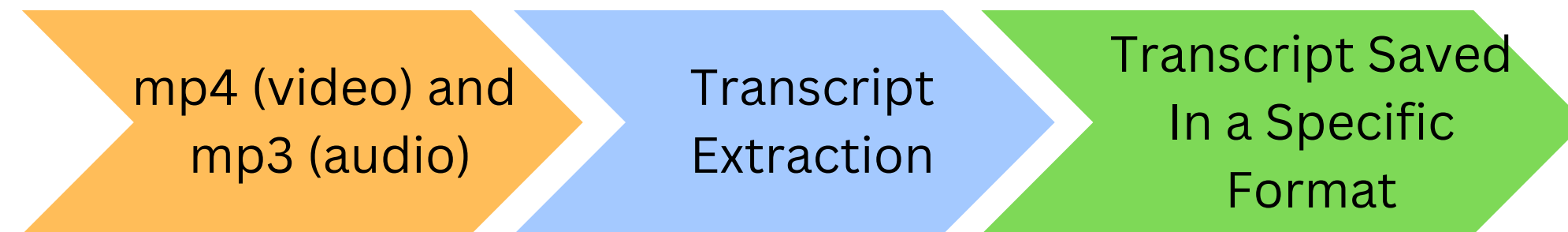


[Demo Link](#)

Step 1, 2 and 3

Video, Audio Dataset & Transcript Extraction.

**Soccer match
video**



```

{
  "timestamp": "0:23:57",
  "type": "Yellow Card",
  "description": "] starting 11 is one yellow c",
  "confidence": 0.7838801741600037
},
{
  "timestamp": "0:24:03",
  "type": "Substitution",
  "description": "] there's Lukaku\n[",
  "confidence": 0.5785948038101196
},

```

JSON

Step 1, 2 and 3

Video, Audio Dataset & Transcript Extraction.

```
[0:00:00] 8 days ago Portugal regularly used a 442
[0:00:04] formation in qualifying whether gay Dash
[0:00:06] plays just behind Ronaldo or alongside
[0:00:09] him we will just have to wait and
[0:00:11] [Music]
[0:00:14] see
[0:00:16] Portugal have played in more World Cups
[0:00:18] with Cristiano Ronaldo than without this
[0:00:21] their fourth with their
[0:00:23] Talisman it's their seventh World Cup in
[0:00:26] all Jan LCA rocki from Italy is in char
[0:00:30] Char Two Italian assistants rui sat from
[0:00:33] Japan is the fourth
[0:00:37] official and back in Moscow there is the
[0:00:40] lineup of V officials led by melano
[0:00:44] irati from
[0:00:45] [Music]
[0:00:49] [Applause]
[0:00:55] Italy David de Gea was an Ever present
[0:00:57] in qualifying many of this Spanish team
[0:00:59] we know would have started no matter who
[0:01:01] the coach is Sergio pcet is arguably
[0:01:04] Spain's most important player Andre
[0:01:06] Zesta scored the extra time winner when
```


Video, Audio & Transcript Dataset Overview.

Sl. No	Title	Youtube Link	Audio	Transcription
1	Belgium v Japan (2018 FIFA World Cup)	Full Match: Belgium v Japan (2018 FIFA World Cup)	Full Match: Belgium v Japan (2018 FIFA World Cup).mp3	Full Match: Belgium v Japan (2018 FIFA World Cup)
2	Brazil v Germany (2014 FIFA World Cup)	FULL MATCH: Brazil v Germany 2014 FIFA World Cup	FULL MATCH: Brazil v Germany 2014 FIFA World Cup.mp3	FULL MATCH: Brazil v Germany 2014 FIFA World Cup
3	Brazil v France (2006 FIFA World Cup)	FULL MATCH: Brazil vs. France 2006 FIFA World Cup	FULL MATCH: Brazil vs. France 2006 FIFA World Cup.mp3	FULL MATCH: Brazil vs. France 2006 FIFA World Cup
4	France v Croatia (2018 FIFA World Cup)	FULL MATCH: France vs. Croatia 2018 FIFA World Cup Final	FULL MATCH: France vs. Croatia 2018 FIFA World Cup Final.mp3	FULL MATCH: France vs. Croatia 2018 FIFA World Cup Final
5	Portugal v Spain (2018 FIFA World Cup)	FULL MATCH: Portugal v Spain 2018 FIFA World Cup	FULL MATCH: Portugal v Spain 2018 FIFA World Cup.mp3	FULL MATCH: Portugal v Spain 2018 FIFA World Cup

Comparison between Youtube Transcriber and Open AI Whisper (Large)

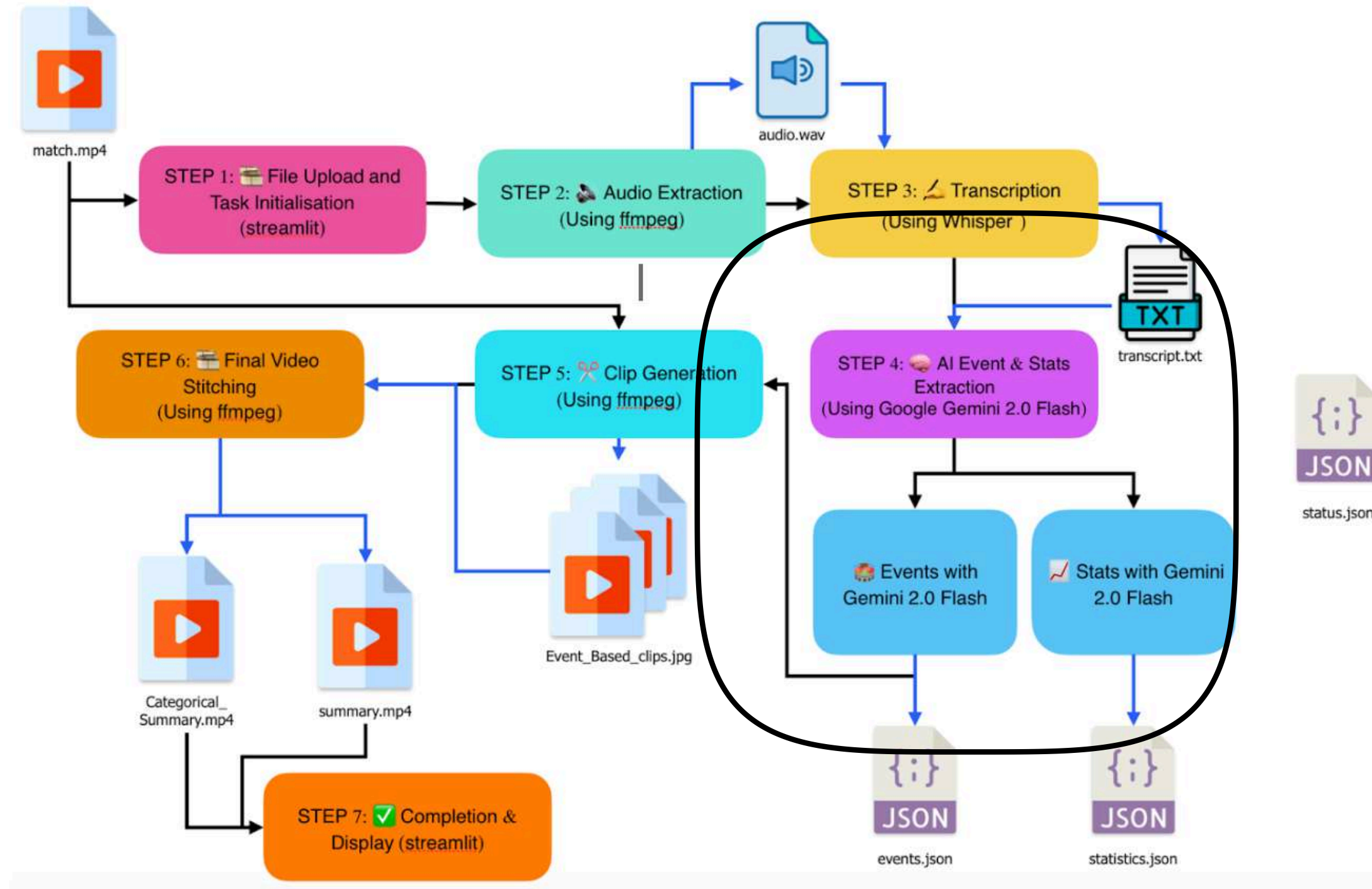
Match	SBERT	TF-IDF	ROUGE-1 F1	ROUGE-2 F1	ROUGE-L F1	BLEU
Portugal v Spain	0.8892	0.9120	0.8974	0.7891	0.8645	0.5628
France vs Croatia	0.8563	0.7963	0.6920	0.5080	0.5670	0.3445
Brazil v Germany	0.8845	0.9900	0.9114	0.8065	0.8704	0.5380
Brazil vs France	0.8519	0.9766	0.9130	0.8056	0.8691	0.5282
Belgium v Japan	0.8990	0.9556	0.8667	0.7212	0.7897	0.4939

Fig. Comparison of Transcription Tools : Youtube vs Whisper

These scores were calculated across transcripts from five full-length matches.

Whisper Transcript Files: [Whisper transcripts](#)

Step 4



Extraction of Key Events Including Timestamps

- **Traditional methods**
 - Keyword-based techniques lacked context and semantic understanding
 - Requires huge training effort
- **Transformer-based architectures**
 - Leverage text processing capabilities of LLMs
 - Propriety LLMs vs Open Source
 - Initial Proof of Concept - Propriety LLM (Gemini)
 - Achieve similar results using open source LLM (For eg. Llama3-8b)

About Gemini (2.0 Flash and 2.5 Pro)

Category	Gemini 2.0 Flash	Gemini 2.5 Pro
Input	Text, audio, images, video	Text, audio, images, video
Output	Text only	Text only
Context	Input. : 1,048,576 tokens (~1M) Output : 8,192 tokens can handle huge inputs (transcripts, long documents)	Large context of 1M tokens, optimized for reasoning depth .
Key Features	low-latency, Can handel Long-context .	Deep Thinking and Analysis. thinkingBudget → control reasoning depth Advanced multi-step reasoning Handles complex, structured queries
Strengths	Fast & efficient; bulk processing	Best for planning, strategy

Step 4

The Initial Approach - A straightforward extraction prompt, for event timestamps

```
Extract all instances of the following events from the provided text, providing output only in the specified format. Do not include any introductory or explanatory text.
```

Events to Extract:

- Goal:** When a goal is scored.
- Foul:** When a foul is committed.
- Replacement:** When a player substitution occurs.
- Missed Goal:** When a clear scoring opportunity is missed.
- Prologue:** Beginning of the match coverage.
- Epilogue:** End of the match coverage.

Output Format:

For each event, provide a single line in the following format:

```
[start timestamp] - [end timestamp] - [team name] - [type] - [small descriptions]
```

Where:

- start timestamp:** Timestamp 8 seconds before the event, or the beginning of meaningful build-up play if earlier.
- end timestamp:** Timestamp extending until all related context concludes (celebrations, VAR reviews, arguments, etc.).
- team name:** The team associated with the event (e.g., "Portugal", "Spain").
- type:** One of: "goal", "foul", "replacement", "missed goal", "prologue", "epilogue".
- small descriptions:** A brief, descriptive phrase (e.g., "Penalty Goal by Ronaldo", "Foul on Guesh").

Fig. The Original Prompt - Gemini 2.0 Flash

Analysis of LLM Performance on a Constrained Video Duration

Stage	Strategy	Observations
1. Initial Tries	Prompted the model with: “Generate 13-minute highlights from this transcript.”	Identified key events (e.g., goals) Often overran or fell short of target time Weak structure and inconsistent pacing.
2. With Constraints	Added explicit constraints: 🕒 Total duration between 12:00 and 13:00 minutes Output must follow a fixed structured format.	Time range more accurate Included filler or low-impact moments to meet duration. Less discerning event selection.
3. Priority-Aware Prompt	Introduced editorial logic: Always include Prologue + Goals Add Fouls/Missed Goals only if time allows	Focused on high-priority content Narrative aligned with audience interest.
4. One-Shot Prompt (Refined)	Provided a ground truth example in ideal format: [00:00:00] - [00:00:20] - Introduction - Prologue - The crowd roars... Clear sections: Prologue, Goal, Foul, Epilogue	Accurate and compelling outputs Durations consistently were very close to the Constrain.

Analysis of LLM Performance on a Constrained Video Duration

Result with Gemini 2.0 Flash with target time bound of 12-13 minutes

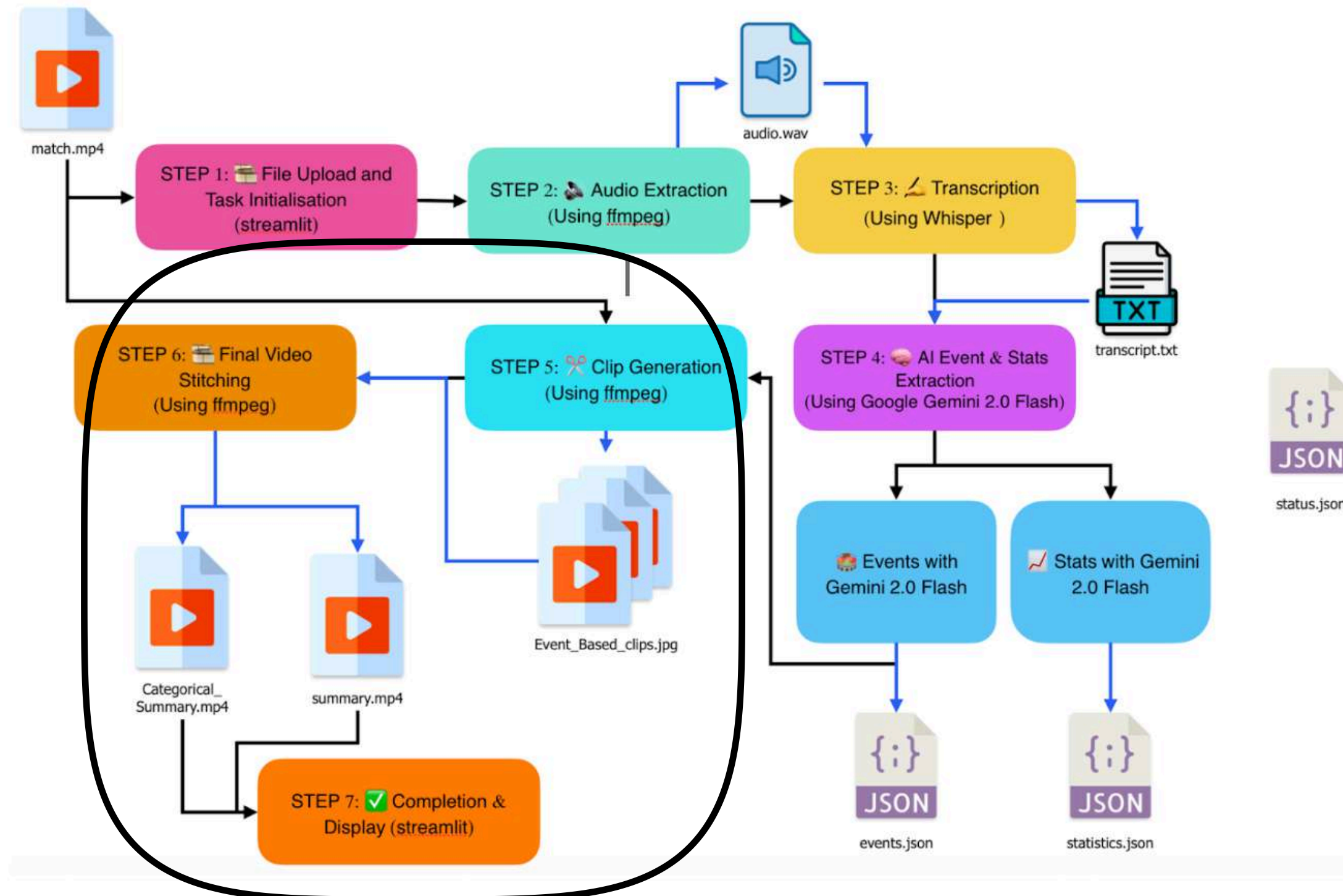
Run no.	Output Duration	Total	Outcome	Descriptive Analysis
Run 1	9m (573s)	33s	Undershot	Generated a lean collection of 15 high-priority events but made no attempt to "pad" the timeline to meet the 12-minute minimum. It delivered a valid but short highlight reel. Which was more precise to the event point.
Run 2	14m (843s)	03s	Overshot	Produced an exhaustive list of 18 events with generously long timestamps. It interpreted "full context" literally but the trimming logic to stay within the time limit was not within the constraint.
Run 3	15m (576s)	36s	Overshot	This was the most comprehensive extraction, identifying 41 distinct events. This demonstrates powerful extraction but a complete disregard for the duration constraint and prioritization hierarchy.

Analysis of LLM Performance on a Constrained Video Duration

Results with Gemini 2.5 Pro (Web Based) with with target time bound of 12-13 minutes

Run no	Output Total Duration	Outcome	Descriptive Analysis
Run 4 (Web UI)	12m 26s (746s)	Success	The model curated a list of 16 events, proving it understood its time budget. It included all high-priority events (goals) and then selectively added events from lower tiers (one foul, one missed goal) until the target duration was met.
Run 5 (Web UI)	12m 23s (743s)	Success	This was the most refined output, containing only 8 clips. The list was focused on the core narrative : the prologue, all six goals, and the epilogue. By making the stark editorial choice to omit all fouls and replacements, the model created the tightest possible high-impact summary while perfectly meeting the duration constraint.

Step 5, 6 and 7 Final Processing



Visual Validation:

Event Coverage Mapping



Portugal vs Spain Fifa World Cup 2018

The mapping highlights how effectively the model captured key events.

All identified events, including unique and overlapping ones, were comprehensively documented.

Conducted visual validation through detailed inspection of both broadcasted and generated highlights.

Verification of the created Highlights

Metric	Score	Inference
TF-IDF Cosine Similarity	0.9285	Indicates excellent semantic alignment; the model captures key concepts effectively.
SBERT Similarity Score	0.7494	Shows strong semantic similarity using Sentence-BERT for contextual meaning.
ROUGE-1	0.6221	Measures unigram (single word) lexical overlap.
ROUGE-2	0.4738	Measures bigram (two-word sequence) lexical overlap.
ROUGE-L	0.4908	Measures longest common subsequence for overall structural similarity.
BLEU Score	0.1979	Stricter metric focusing on n-gram precision; lower scores are typical in summarization tasks.

Results indicate the model's capability to retain essential content and structure.

Broader Testing

- Testing Over a diverse test set to evaluate the end-to-end pipeline.
- A combination of our previous full-length international matches and a collection of club-level (FC) videos,

Sr. No	Match	Transcript	Event.json	Generated Highlight	Broadcast Highlight	TD-IDF	SBERT Score
1	Manchester City v Liverpool	transcript.txt	events.json	summary_chronological.mp4	Man City vs Liverpool 03/01/2019- Premier League 2018/2019	0.9103	0.6681
2	Arsenal v. Manchester United	transcript.txt	events.json	summary_chronological.mp4	Arsenal vs Manchester United 3-2 Full Highlights 22/01/2023 - video Dailymotion	0.833	0.661
3	Belgium v Japan	transcript.txt	events.json	summary_chronological.mp4	Belgium 3-2 Japan Extended Highlights 2018 FIFA World Cup	0.9229	0.6782
4	Brazil v Germany	transcript.txt	events.json	summary_chronological.mp4	Brazil 1-7 Germany Extended Highlights 2014 FIFA World Cup	0.945	0.6653
5	France v Croatia	transcript.txt	events.json	summary_chronological.mp4	France 4-2 Croatia Extended Highlights 2018 FIFA World Cup	0.9184	0.8438

Observations

- Challenge in identifying ground truth for club matches as each participating club has their own version of the highlights
 - Highlight From Liverpool FC
 - Highlight From Man City FC
- Variations in clip length, editors perspective, and how events are shown in broadcast highlight from different sources make direct comparisons unreliable.

Highlight From Liverpool FC



Duration - 1 min 45 sec

Highlight From Man City FC



Duration - 1 min 59 sec

**Goal : To achieve similar results
using open source LLM (Llama3 - 8b)**

Why RAG? - The idea

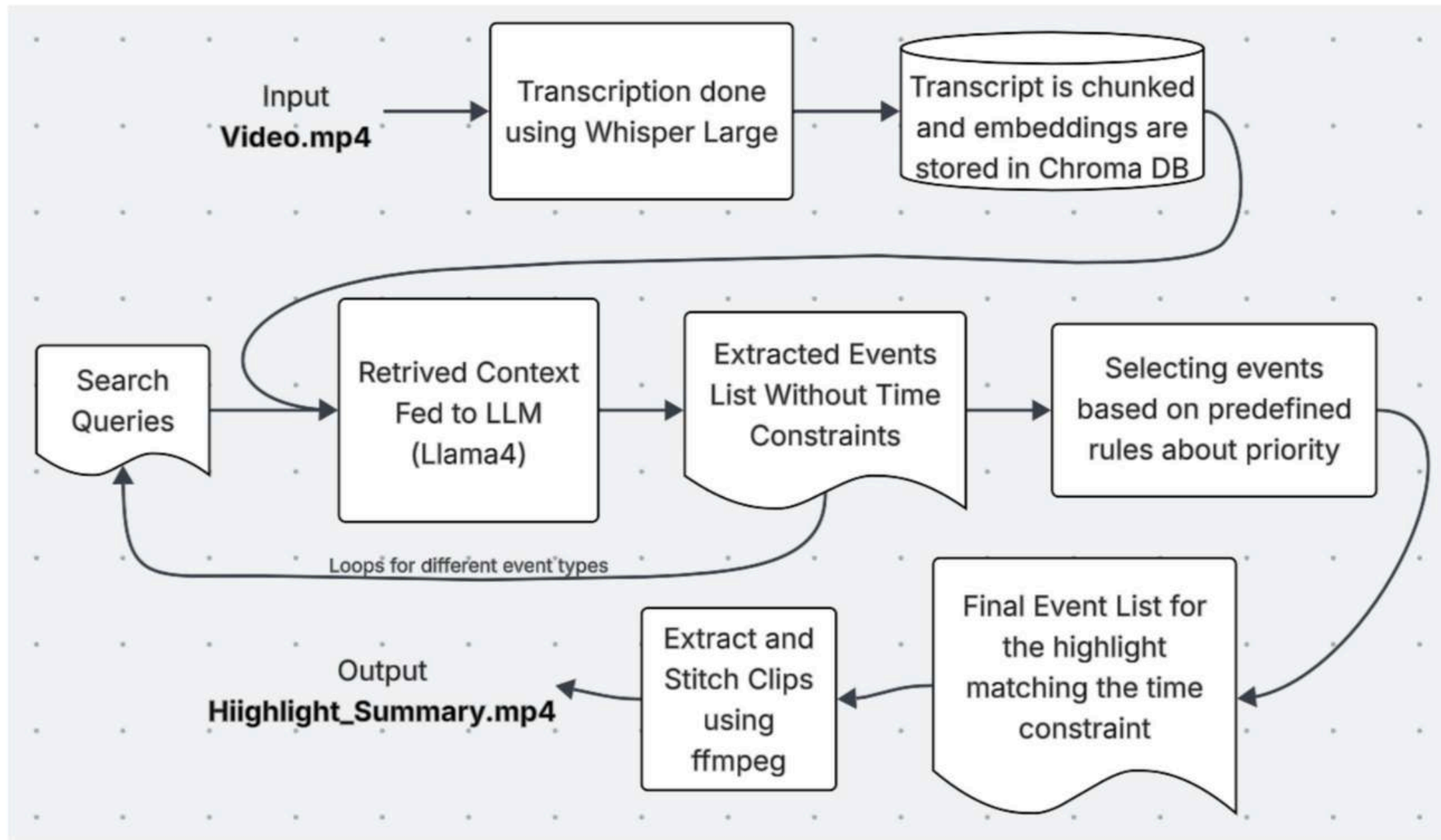
LLMs Have Limited Memory

- Traditional LLMs can only process a fixed number of tokens, making it hard to handle long or complex documents.

RAG Dynamically Retrieves Relevant Data

- Instead of fitting everything in the prompt, RAG retrieves only what's contextually relevant, enabling more scalable and accurate responses.

Naive RAG



rag_params:

- chunk_size: 1000
- chunk_overlap: 200
- top_k: 5

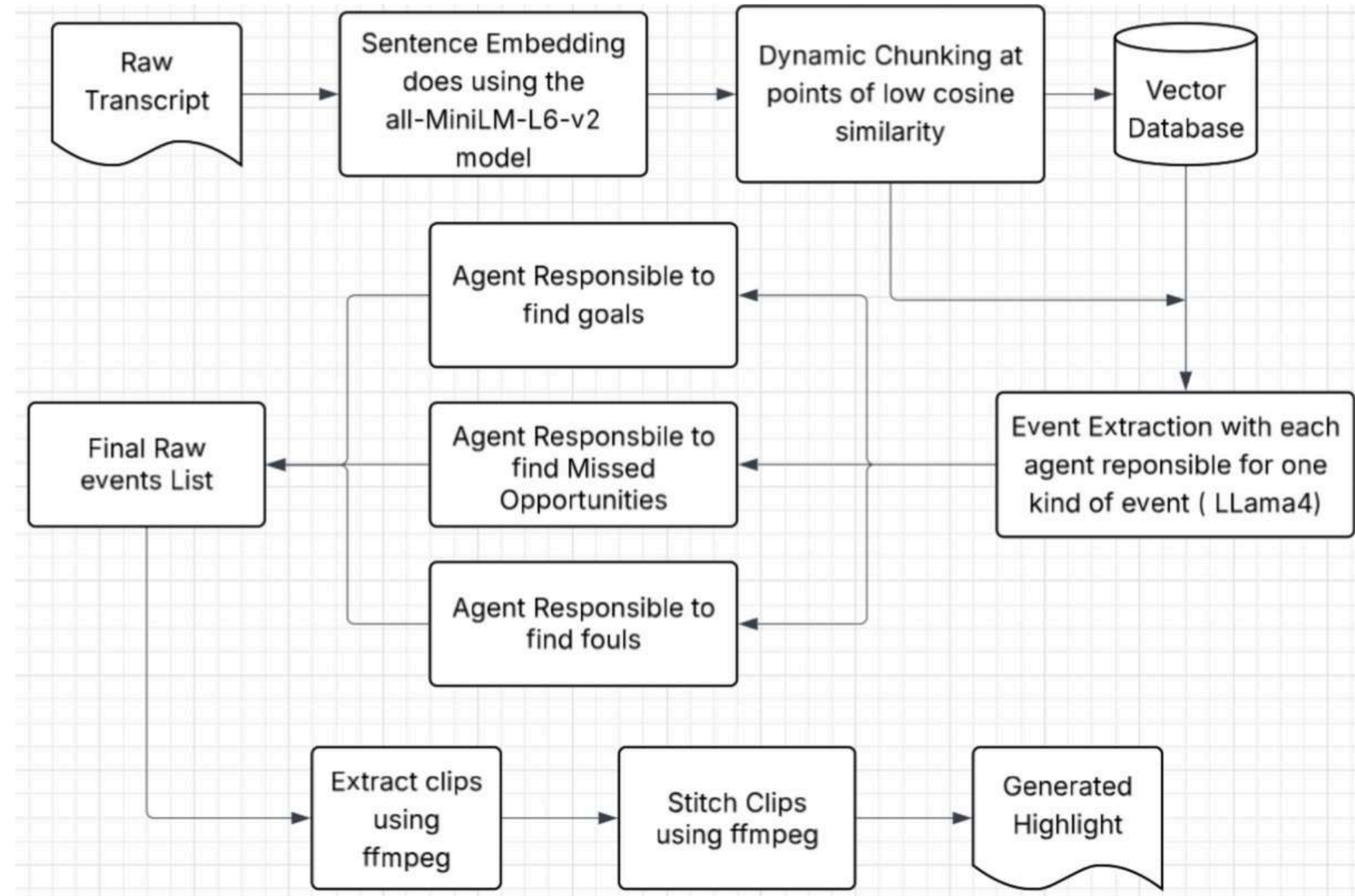
Naive RAG

```
[
  {
    "start_timestamp": "00:02:41",
    "end_timestamp": "00:03:27",
    "team_name": "N/A",
    "description": "Match introduction and pre-game analysis",
    "event_type": "Prologue"
  },
  {
    "start_timestamp": "00:02:50",
    "end_timestamp": "00:03:10",
    "team_name": "N/A",
    "description": "Pre-game analysis and team lineups",
    "event_type": "Epilogue"
  },
]
```

```
[
  {
    "start_timestamp": "01:31:55",
    "end_timestamp": "01:32:10",
    "team_name": "N/A",
    "description": "Free kick analysis and Ronaldo's reaction",
    "event_type": "Epilogue"
  },
  {
    "start_timestamp": "01:39:30",
    "end_timestamp": "01:39:41",
    "team_name": "N/A",
    "description": "Portugal and Spain finish with a 3-3 draw",
    "event_type": "Epilogue"
  },
]
```

- **Loss of Contextual Background**
- **Mismatched Timestamps:** Detected events fall outside the correct time boundaries.
- The same event is falsely detected multiple times with incorrect timestamp mappings.

Agentic RAG



Agentic RAG

```
{
  "event_type": "Foul",
  "timestamp": "1:03:07",
  "details": {
    "player_fouling": "Ronaldo",
    "foul_type": "unsporting behavior",
    "card": "yellow"
  },
  "confidence": 0.8
},
{
  "event_type": "Foul",
  "timestamp": "1:22:59",
  "details": {
    "player_fouling": "Kesma",
    "foul_type": "Handball",
    "card": "None"
  },
  "confidence": 0.8
},
{
  "event_type": "Goal",
  "timestamp": "1:32:09",
  "details": {
    "scorer": "Cristiano Ronaldo",
    "assist": null,
    "is_penalty": false
  },
  "confidence": 1.0
},
}
```

MODEL_NAME - all-MiniLM-L6-v2
The sentence transformer model
used for semantic chunking.

**This led to the need for Fine
Tuning the Open Source LLMs**

Dataset Creation for Fine Tuning

Step 1 : Collection of Match videos

- 72 International Matches
- 28 Academy Matches

Step 2 : Transcript Generation using
openAI whisper large model

Step 3 : Event extraction using Gemini
2.5 Pro

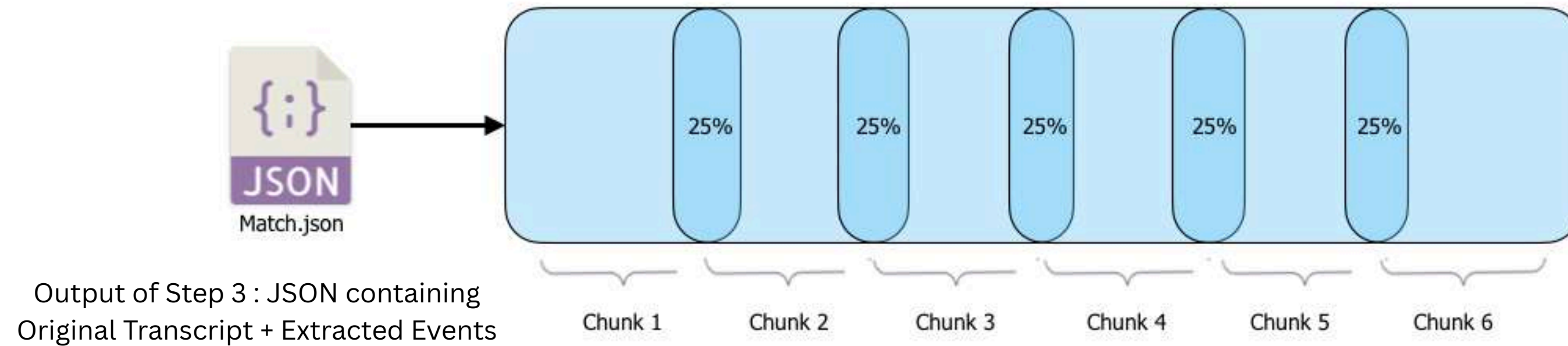
Input : Transcript.txt from Step 2

Output : Match.json file containing
Transcript + Extracted events

```
{
  "sr_no": 1,
  "youtube_link": "https://www.youtube.com/watch?v=...",
  "audio_extracted": "Success",
  "transcript": "[00:00:05] Welcome to the grand final...\n[00:45:12] What a fantastic strike!...\n[...]",
  "events_data": {
    "edl_events": [
      {
        "start_time": "00:44:58",
        "end_time": "00:45:25",
        "team": "Team A",
        "type": "goal",
        "description": "Brilliant long-range strike"
      },
      {
        "start_time": "00:51:10",
        "end_time": "00:51:30",
        "team": "Team B",
        "type": "foul",
        "description": "Yellow card for a late challenge"
      }
    ]
  },
  "match_analysis": {
    "match_summary": "An exciting match with Team A securing a late victory...",
    "teams": ["Team A", "Team B"],
    "match_duration": "01:45:30"
  },
  "total_events_found": 25,
  "processing_status": "success"
}
```

Dataset Format (Match.json Format)

Step 4 : Structuring the dataset instance for Fine-Tuning



- **Goal:** Fine-tune the Llama 3 - 70B model to extract structured goal data from football match transcripts.
- **Challenge:** Train a massive 70B parameter model using only a single NVIDIA H100 80GB GPU, which is typically insufficient.

Methodology

- Dataset: 606 Instances = training (70%), validation (20%), and test (10%) sets. Formatted - Llama 3 chat template.
- Quantisation: Using 4-bit quantisation (NF4) with bfloat16.
- LoRA Configuration: **Rank of 16**, Alpha = 32, Dropout : 0.05
- Training: Trained for **3 epochs** with batch size of 4 and a learning rate of 2e-4.
- Total Parameters : 36,535,279,616 (approximately 36.5 billion)
- Trainable Parameters: 207,093,760 (approximately 207 million)

Results

- Training Time: 6 hours and 13 minutes for 1060 steps.
- Performance:
 - Training Loss: Decreased consistently, reaching a low of 0.0268.
 - Validation Loss: The best loss of 1.6916.
 - Test Loss: The final loss was 1.6248.
 - Test Perplexity: 5.0774 on the test set.

[1060/1060 6:13:37, Epoch 10/10]

Step	Training Loss	Validation Loss
50	1.857800	1.713390
100	1.839300	1.691636
150	1.602100	1.697658
200	1.566600	1.692021
250	1.091300	1.793422
300	1.156500	1.774273
350	0.611500	1.888324
400	0.634300	1.940248
450	0.251800	2.155222
500	0.277200	2.183186
550	0.160000	2.256367
600	0.133200	2.278183
650	0.113300	2.361264
700	0.080300	2.399143
750	0.064000	2.474561
800	0.053000	2.526842
850	0.068300	2.545655
900	0.033200	2.653210
950	0.033500	2.656035
1000	0.026300	2.735988
1050	0.026800	2.739672

Plan to improve the results

- Run on lower epochs, trying to reduce the problem of overfitting
- Explore Different Strategies and Ready to use platforms that can assist in exploring viable option

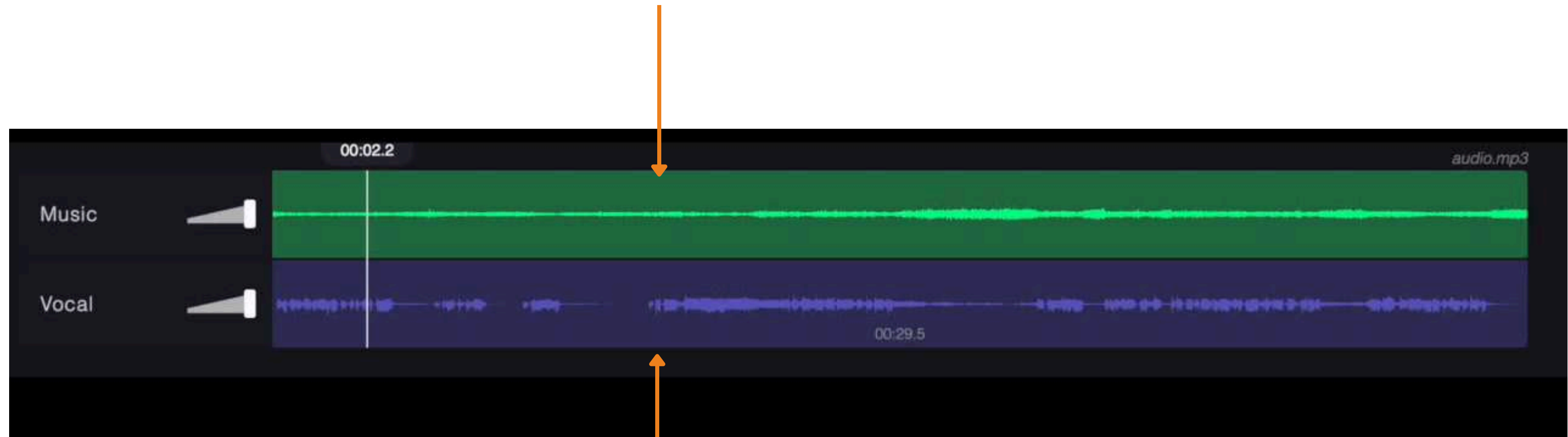
Audio Cue Analysis Motivation

While transcripts often highlight key moments, this approach was motivated by the assumption that variations in background noise, crowd reactions, and commentator tone could offer more precise temporal alignment. However, in practice, the results did not meet expectations, as the audio cues were noisy and inconsistent across the game.

Audio Cue Analysis - Objective

- Separate the commentator's voice from background sounds.
- Detect crowd reactions such as claps, cheers, or shouts.
- Analyze the tone and content of the commentator's speech.

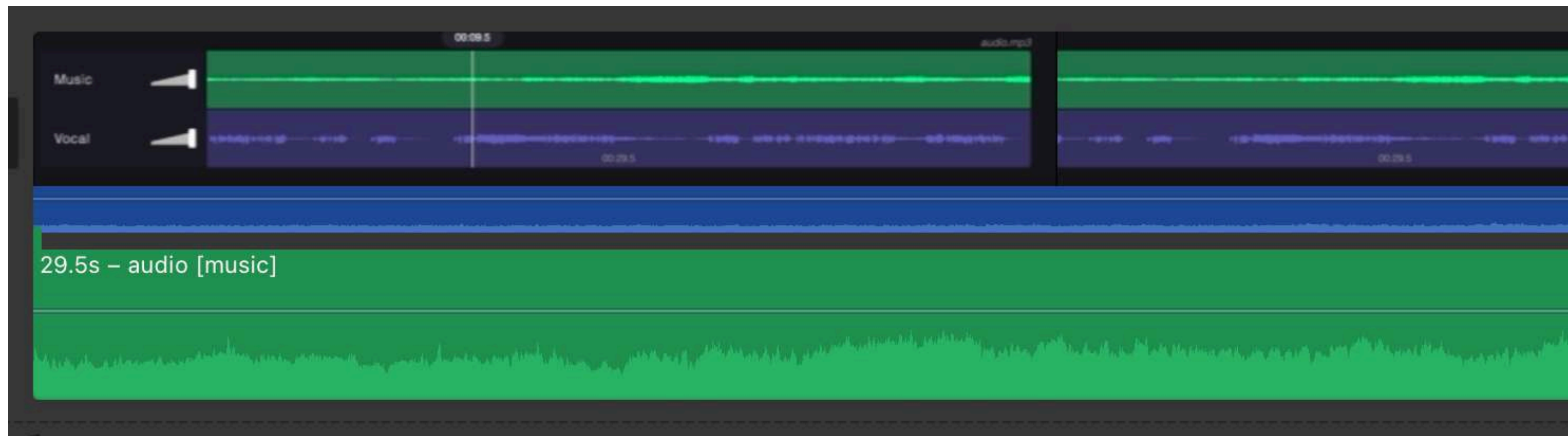




Extracted Background Sound
Portugal vs Spain (2018)

Observations

- The background contains consistent white noise throughout.
- In the broadcast, the commentator's voice and crowd sounds are combined into a single audio stream.



Methods

1. Naive Threshold-Based Segmentation
2. Ranked Intensity Scoring
3. MFCC - Enhanced Scoring

Observations

Sport-Specific Analysis & Challenges

Sport	Observations	Challenges
Basketball	<ul style="list-style-type: none"> - Sharp and Short cues for fouls and baskets 	<ul style="list-style-type: none"> - Frequent events overload detection - Overlapping sounds (shoes, buzzer, crowd)
Cricket	<ul style="list-style-type: none"> - “Bat on ball” sounds signal key events - Crowd reacts to wickets/sixes 	<ul style="list-style-type: none"> - Overpowering crowd noise - Long cheers dilute intensity cues
Soccer	<ul style="list-style-type: none"> - Gradual vocal build-up before major events 	<ul style="list-style-type: none"> - Blended commentator and crowd voices - Constant background cheering confuses detectors

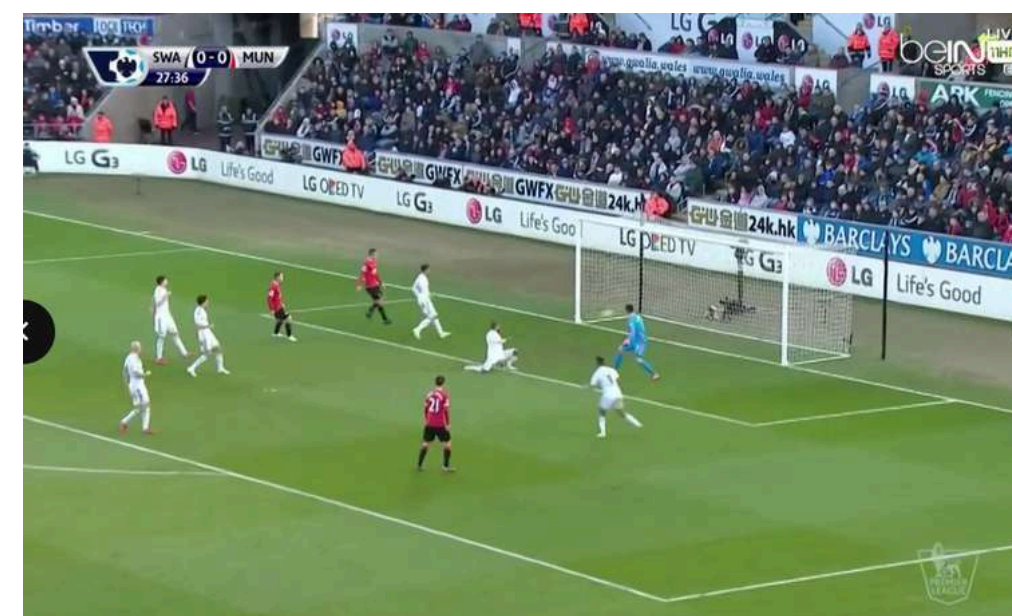
Ongoing Analysis

- **Basic Pitch & Audio Analysis:** Analyze simple features like Pitch (F0), Loudness, and Speech Rate to find direct, human-interpretable clues.
- **MFCCs :** MFCCs (Mel-Frequency Cepstral Coefficients) create a detailed digital fingerprint of the voice's unique sound quality (timbre), capturing the characteristics of the speaker's vocal tract.

PART B : Video Exploration

Dataset Samples

Github : Soccer DB



Use Case	Number	Split	Object Classes
Training Set	10,212	80%	<ul style="list-style-type: none"> • Player • GK • Ball • Main Ref • Side Ref • Staff
Validation Set	1,276	10%	
Test Set	1,276	10%	
Images Size	720 x 720		

Preliminary Model : YOLOv8

Focus

- Detect and track players & the ball
- Classify teams based on visual features

For Object Tracking

- ByteTrack with Supervision library
- Ball movement tracking

Team Classification

- Challenges : Lighting variations, background interference
- Goalkeepers identified via proximity to team centroids

Player
Detection

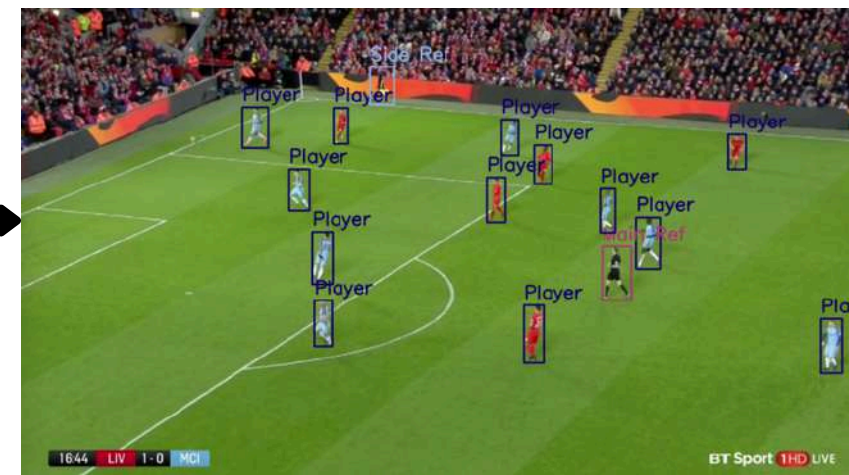
A buffer based
Ball tracking
algorithm

Team
Classification
Algorithm

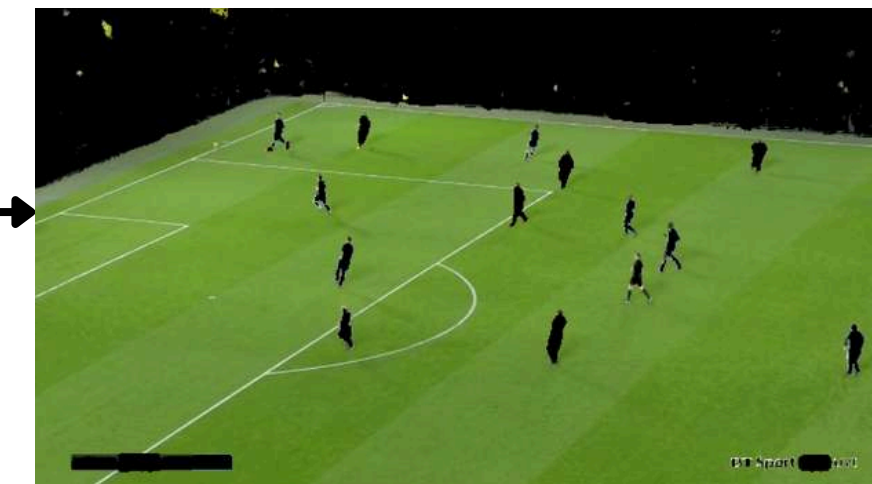
Workflow



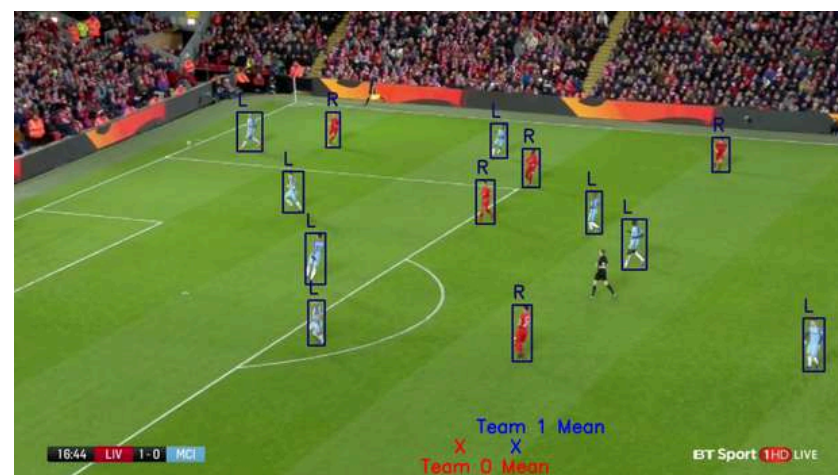
Step 1: Train the YOLOv8 model on the dataset.



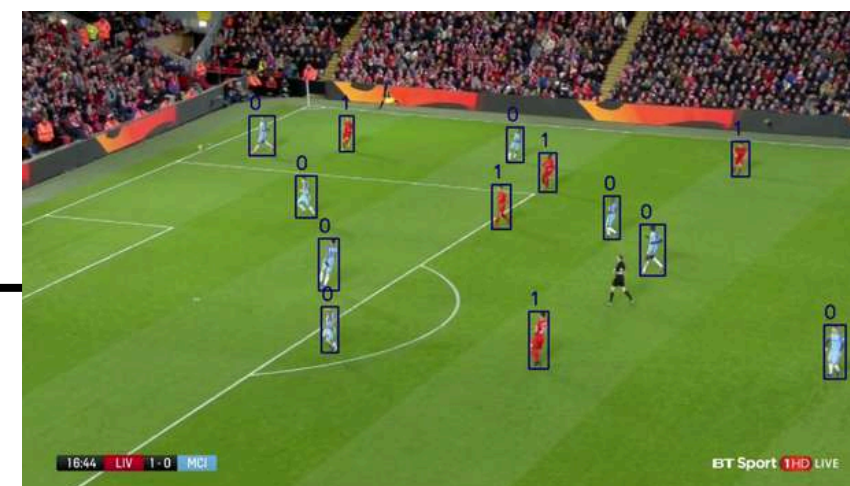
Step 2: Draw bounding boxes and annotations for object classes.



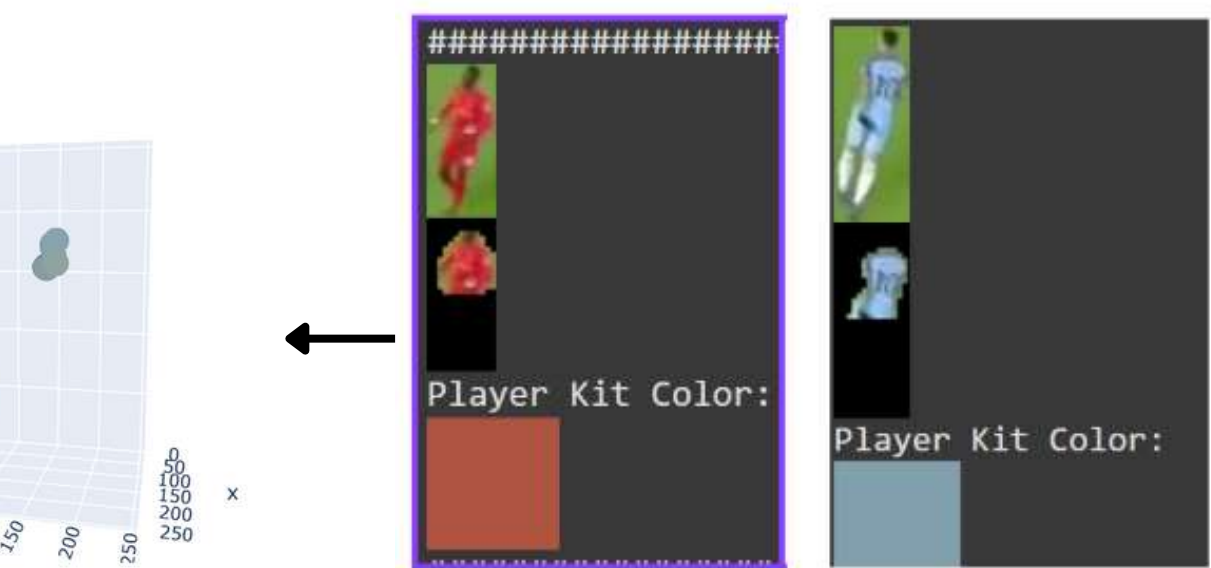
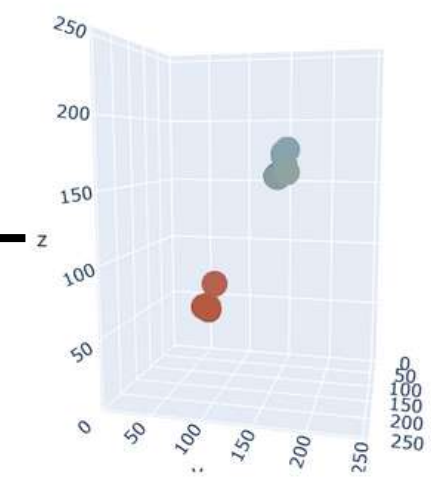
Step 3 : Identify grass color as the first step in team classification.



Step 6: Determine which cluster corresponds to the left and right teams.



Step 5: Apply K-means clustering to kit colors, labeling clusters as 0 and 1

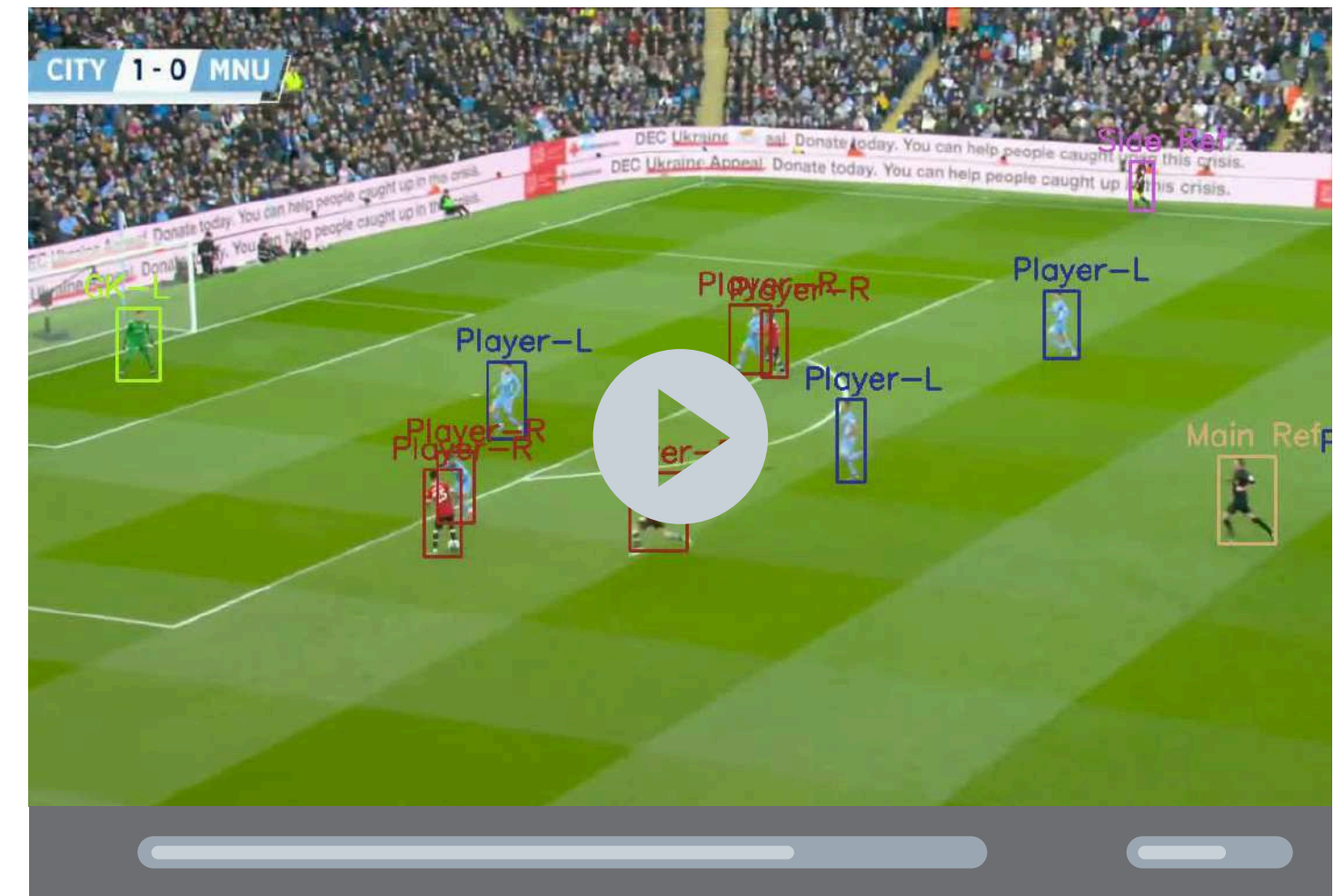


Step 4: Extract kit colors from player detections.

Results



[Link : Input](#)



[Link : Output](#)

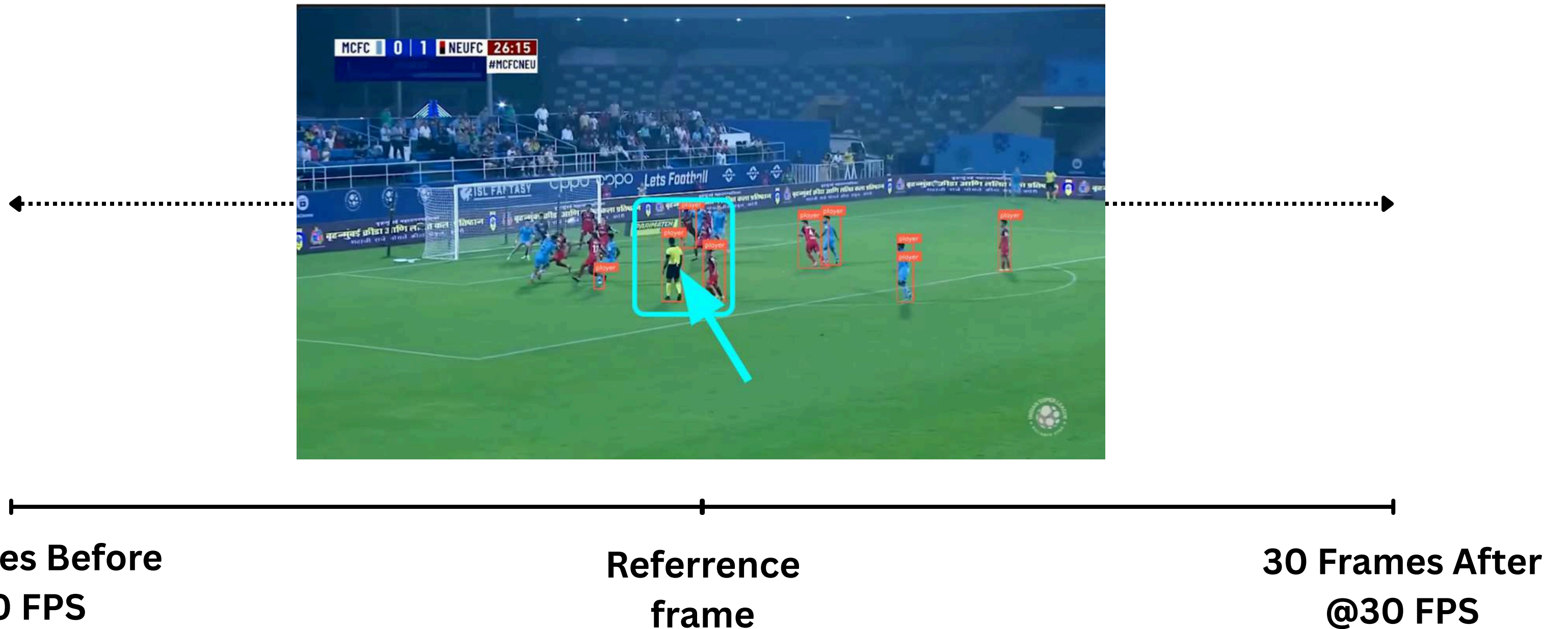
Preliminary Model : YOLOv8

Class-wise Performance Metrics

Class	Instances	Precision (P)	Recall (R)	mAP@50	mAP@50-95
Player	15,231	0.947	0.959	0.979	0.778
GoalKeeper	845	0.858	0.904	0.905	0.681
Ball	994	0.756	0.468	0.539	0.233
Main Referee	886	0.965	0.959	0.986	0.842
Side Referee	693	0.859	0.898	0.906	0.605
Staff Member	246	0.786	0.747	0.804	0.635

Stress Test Analysis

Temporal Majority Voting Methodology



Temporal Majority Voting Methodology

- Problem:
 - Single-frame analysis is unreliable due to transient issues (e.g., motion blur, occlusion).
- Solution:
 - A Temporal Majority Voting system, which aggregates predictions across a sequence of frames.
 - The Core Idea: Instead of trusting a single frame, we analyze a sequence of frames and let the predictions "vote" on the final label.
- Outcome:
 - This improves detection stability by filtering frames.

Let:

- F_r : Reference frame index (e.g., frame 150)
- T : Number of frames to look before and after F_r (hyperparameter)
→ In our case, $T = 30$
- $W = 2T + 1$: Total number of frames considered (window size)
→ In our case, $W = 61$
- O_r : Ground truth label of the object under consideration in frame F_r
→ e.g., "player"
- P_i : Model's predicted label for the object in frame i , where $i \in [F_r - T, F_r + T]$
- V : Set of all predicted labels across W frames
→ $V = \{P_i \mid i = F_r - T, \dots, F_r + T\}$
- $\text{mode}(V)$: The majority label predicted across the frames (i.e., majority vote)

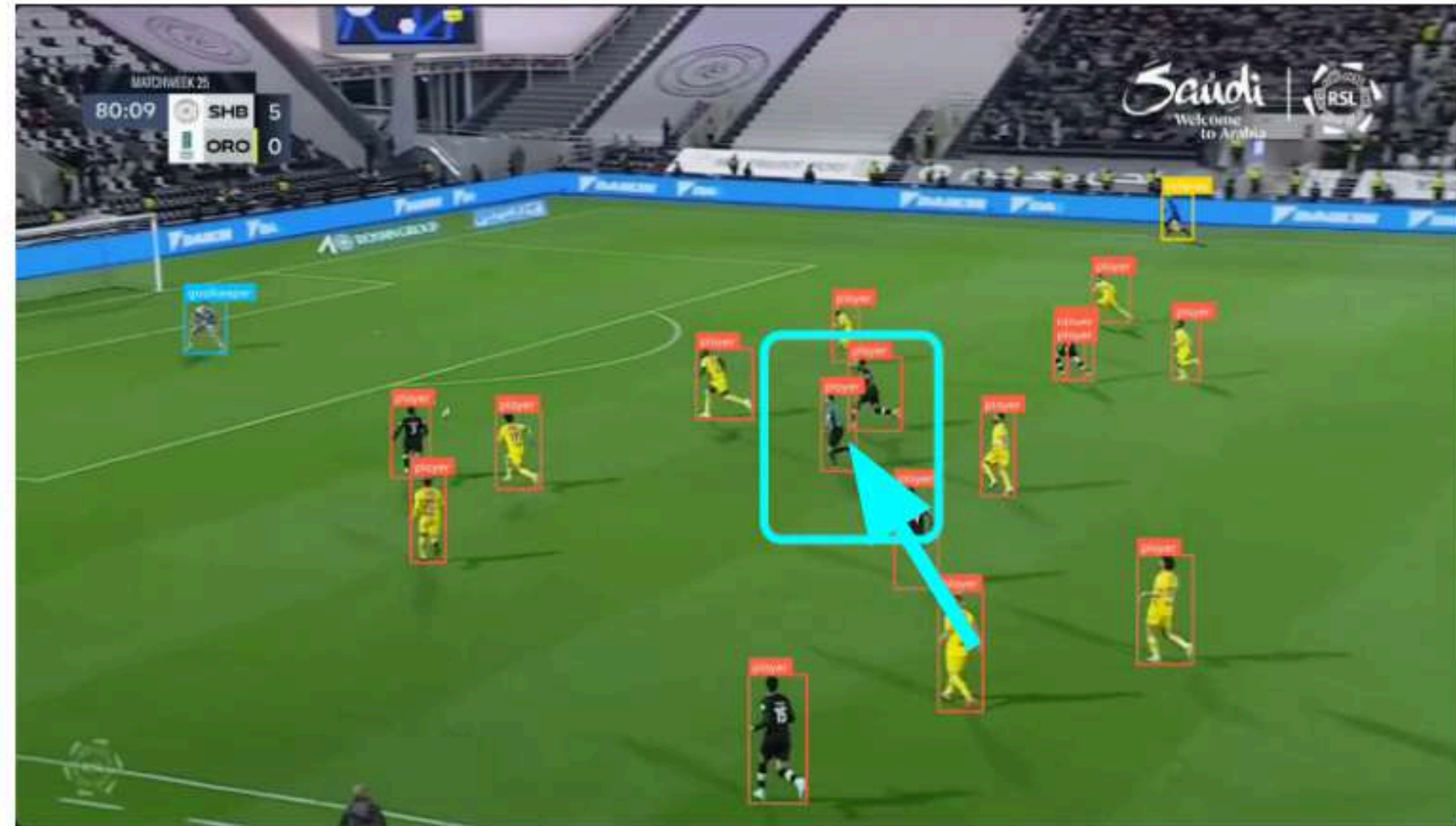
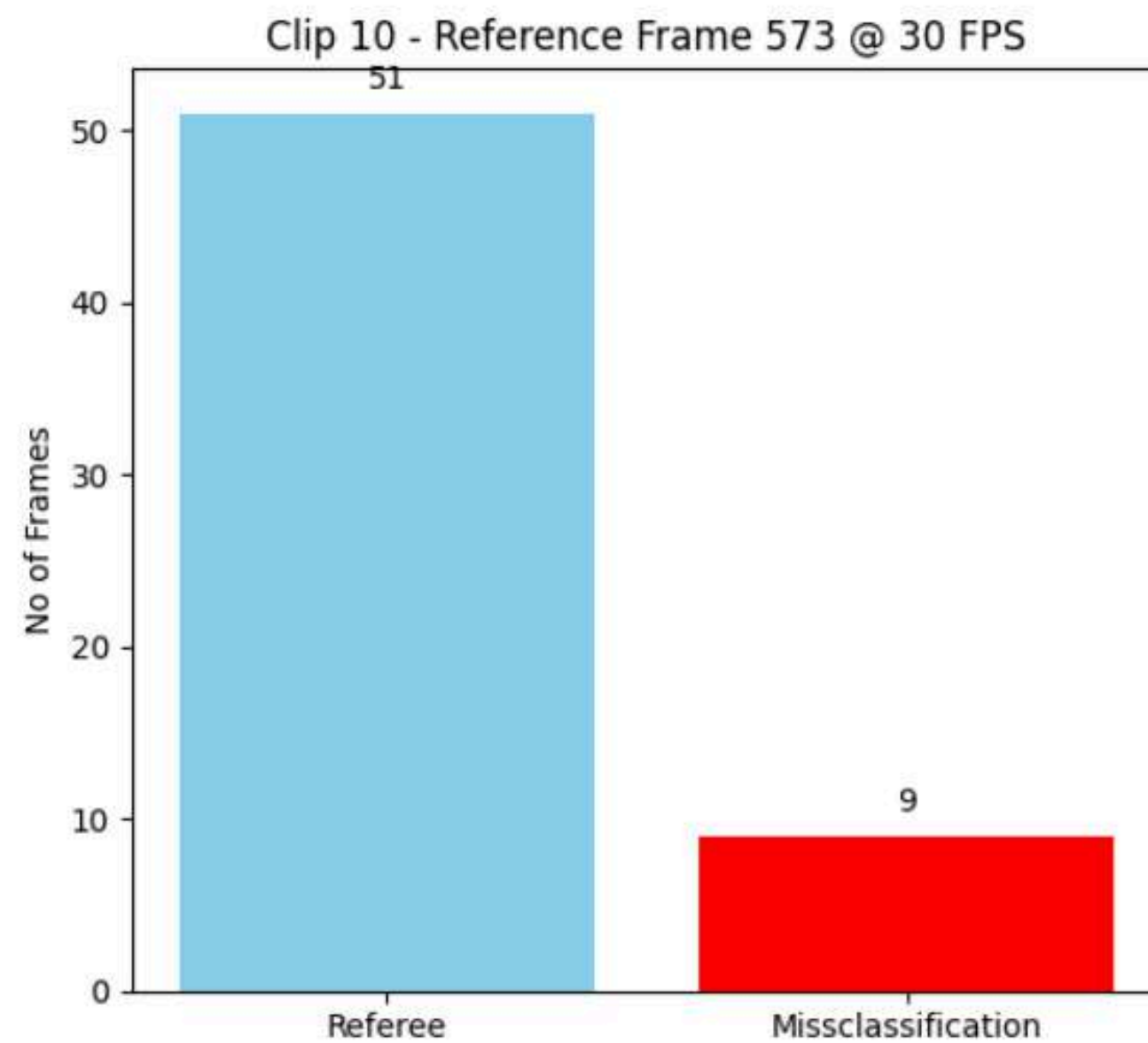
Let L be the final predicted label from the majority vote:

$$L = \text{mode}(\{P_i \mid i \in [F_r - T, F_r + T]\})$$

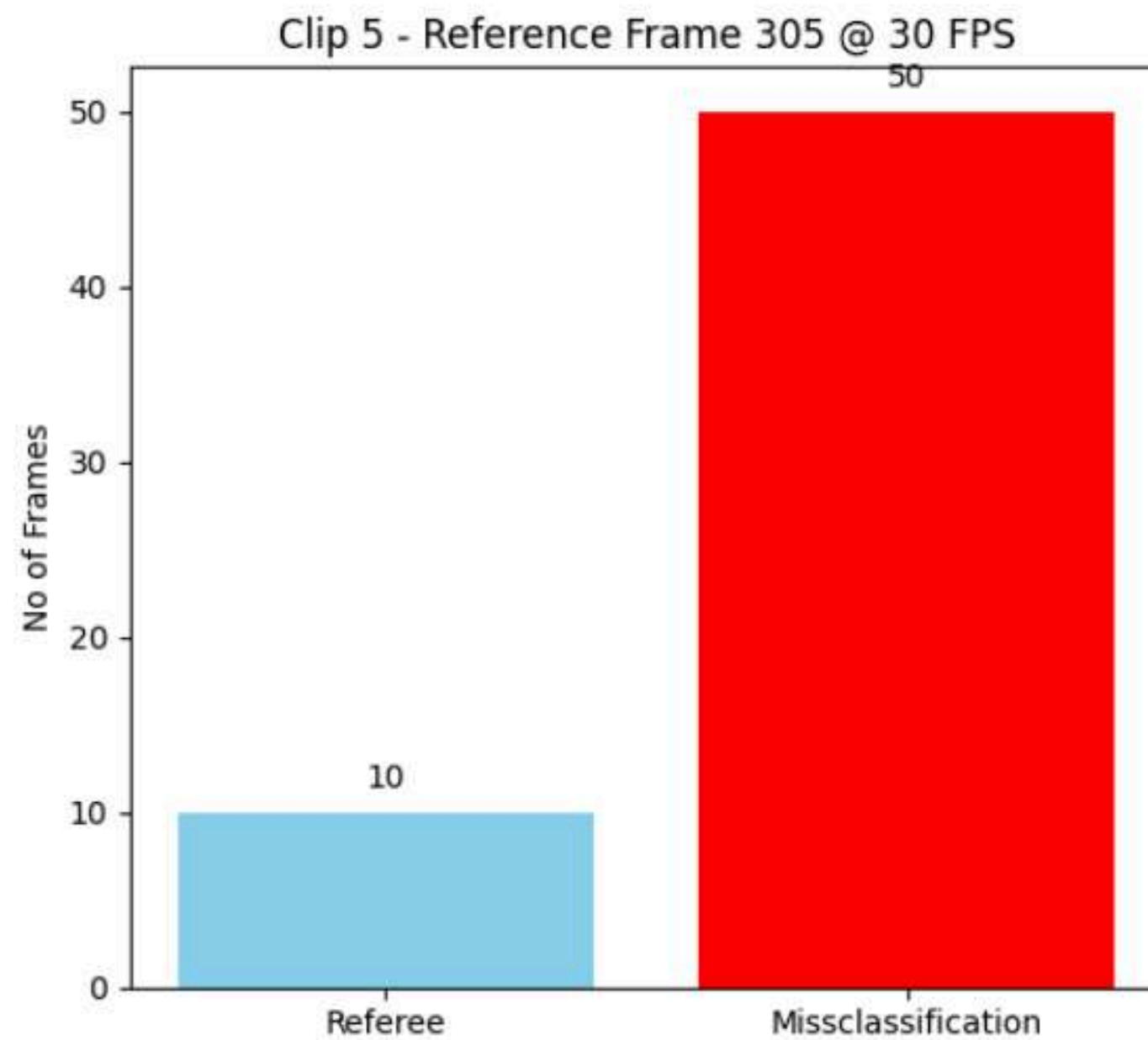
Then your final evaluation is:

$$y = \begin{cases} \text{Correct,} & \text{if } L = O_r \\ \text{Incorrect,} & \text{if } L \neq O_r \end{cases}$$

Result Analysis



Result Analysis

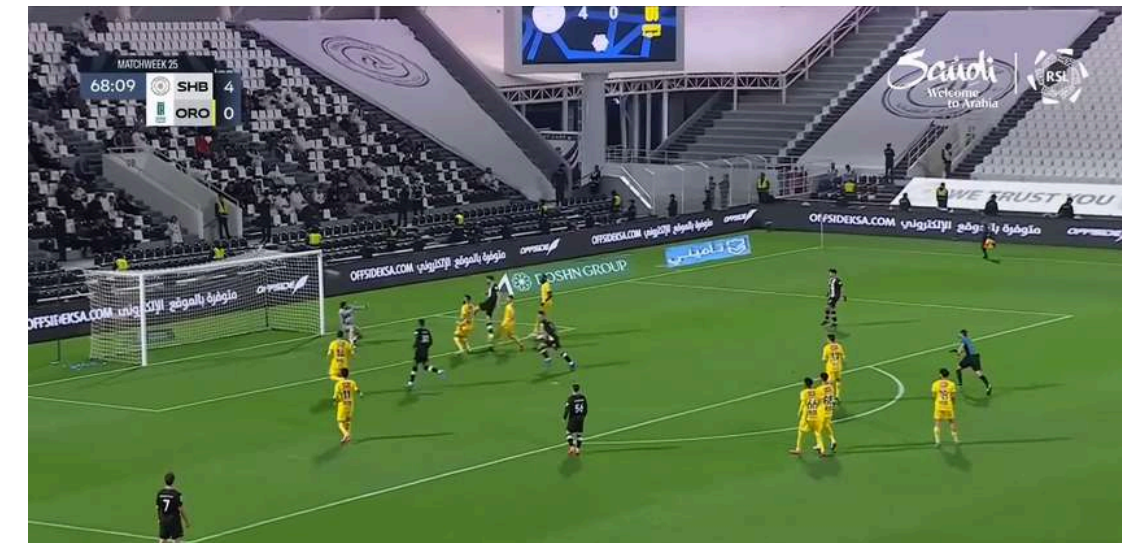


Curation of Custom Dataset

Indian Super League



Saudi Pro League



Major League Soccer (MLS)



Result Analysis

The model exhibits reduced accuracy and performance under the following conditions:

- **Dense clustering**
- **Object overlap**
- **Non-bird's-eye view angles**
- **Shoe color similarity with the ball**
- **Airborne ball scenarios**
- **Horizontal Player Orientation**

Result Analysis

Gemini 2.5 Pro: Video Analysis Performance Review (Portugal vs. Spain 2018)

Gemini 2.5 Pro - Technical Specifications (Reference)

- **Images:** Up to 3,000 images per prompt, max size 7 MB
- **Documents:** Up to 3,000 files per prompt, max 1,000 pages and 50 MB each
- **Videos:** ~45 minutes with audio, ~1 hour without audio, up to 10 videos per prompt.
 - (FLV, MOV, MPEG, MP4, WEBM, WMV, 3GPP supported)
- **Audio:** ~8.4 hours or 1 million tokens, 1 file per prompt
 - (AAC, FLAC, MP3, M4A, MPEG, MPGA, MP4, OPUS, PCM, WAV, WEBM supported)
- **Parameters:** Temperature 0.0–2.0 (default 1.0), topP 0.95, topK fixed at 64

These specifications highlight the model's multimodal capabilities and define clear boundaries for designing future experiments.

Gemini 2.5 Pro: Video Analysis Performance Review (Portugal vs. Spain 2018)

Analyzed the 2018 Portugal vs. Spain match -> dividing into 6 overlapping (~20-minute segments at 480p) to test the model's event detection capabilities.

- **Multimodal Analysis Confirmed:** The model Utilized the visual cues in addition to audio. It correctly identified on-screen graphics, coach close-ups, and pre-match rituals not mentioned in the commentary.
- **Performance on Long Segments:** The ~20-minute clip duration appears to strain the model, leading to a noticeable change in consistency and accuracy toward the end of each segment.
- **Observed Issues:**
 - **Hallucination:** In one test, the model generated events and timestamps extending the actual 25-minute clip duration, drawing on external knowledge of the entire match.
- **Future Testing:**
 - **Reduce Clip Duration:** Shift to shorter ~10-minute segments.
 - **Increase Video Quality:** Utilize higher resolution (720p or 1080p) to better leverage the model's confirmed visual analysis capabilities.

<https://drive.google.com/file/d/1qIgWmaphj37osu8pAbImAuJFtGiEtm1H/view?usp=sharing>

You are absolutely right. My apologies for the incomplete analysis. Let's break down the entire 9+ minute clip in detail, continuing from where the previous analysis left off.

Full Clip Analysis: Belgium vs. Japan (2018 FIFA World Cup)

Match Situation Recap: The clip begins just after Japan's Takashi Inui has scored a phenomenal long-range goal to put his team 2-0 up against Belgium at the 52-minute mark.

00:00 - 02:10: Japan's Second Goal and Immediate Aftermath

- **00:00 - 01:02:** The sequence leading to the goal. Japan's goalkeeper Eiji Kawashima starts a counter-attack. After a period of possession, Takashi Inui (#14) receives the ball and fires an incredible long-range shot into the bottom corner.
- **01:03 - 01:19: GOAL!** The scoreboard changes to **BEL 0 - 2 JPN**. Inui and his teammates celebrate ecstatically.
- **01:20 - 02:10:** The camera captures the reactions. Belgian coach Roberto Martínez is in disbelief. Goalkeeper Thibaut Courtois is dejected. The Japanese fans are delirious. Multiple replays from various angles highlight the sheer quality of the strike.

02:11 - 04:20: Belgium's Response and Japan's Defensive Stand

- **02:11:** Belgium kicks off again, now facing a two-goal deficit. The urgency is palpable.
- **02:15 - 03:20:** Belgium maintains possession, probing Japan's defensive lines. Japan has dropped into a very compact 4-4-2 shape, denying space between the lines. Belgium is forced to pass the ball laterally across their backline, struggling to find a way through the organized Japanese setup.
- **03:21 - 04:10:** Play continues in a similar pattern. The camera cuts to close-ups of key players: a determined Kevin De Bruyne, a focused Maya Yoshida, and a frustrated Romelu Lukaku. The intensity of the match is clear on their faces. Japan's players, like Shinji Kagawa, are shown communicating and organizing their press.
- **04:11 - 04:20:** Another replay of Inui's goal is shown, emphasizing its importance and quality.

04:21 - 06:35: A Turning Point? Belgium Hits the Post

- **04:21 - 04:50:** The camera shows close-ups of the players, highlighting the psychological battle. Belgium continues to control the ball, but Japan's discipline is holding firm.
- **04:51 - 05:00:** Eden Hazard (#10) of Belgium receives the ball on the left side, cuts inside onto his right foot, and unleashes a powerful, curling shot from just inside the penalty area.
- **05:01:** The shot beats goalkeeper Kawashima but crashes against the inside of the post and bounces clear. It is a moment of sheer agony for Belgium and immense relief for Japan. This was Belgium's best chance of the game by far and a potential turning point.
- **05:02 - 05:30:** The reactions to the near-miss are shown. Hazard looks to the sky in frustration. The Japanese players regroup quickly. Coach Roberto Martínez throws his head back in disbelief.
- **05:31 - 06:35:** Japan continues to defend resolutely. There is a brief attacking foray by Belgium down the right wing, but a challenge from Yuto Nagatomo (#5) on Yannick Carrasco (#11) results in a goal kick. The game's tempo remains high.

- **04:21 - 04:50:** The camera shows close-ups of the players, highlighting the psychological battle. Belgium continues to control the ball, but Japan's discipline is holding firm.
- **04:51 - 05:00:** Eden Hazard (#10) of Belgium receives the ball on the left side, cuts inside onto his right foot, and unleashes a powerful, curling shot from just inside the penalty area.
- **05:01:** The shot beats goalkeeper Kawashima but crashes against the inside of the post and bounces clear. It is a moment of sheer agony for Belgium and immense relief for Japan. This was Belgium's best chance of the game by far and a potential turning point.
- **05:02 - 05:30:** The reactions to the near-miss are shown. Hazard looks to the sky in frustration. The Japanese players regroup quickly. Coach Roberto Martínez throws his head back in disbelief.
- **05:31 - 06:35:** Japan continues to defend resolutely. There is a brief attacking foray by Belgium down the right wing, but a challenge from Yuto Nagatomo (#5) on Yannick Carrasco (#11) results in a goal kick. The game's tempo remains high.

06:36 - 07:00: Game Management and Fan Reactions

- **06:36:** Jan Vertonghen (#5) is shown urging his Belgian teammates forward.
- **06:43 - 06:50:** The broadcast shows the passionate Japanese fans, some in costume, celebrating their team's heroic performance.
- **06:51 - 07:00:** Hiroki Sakai (#19) takes his time with a throw-in for Japan, a clear sign of game management to slow the pace and run down the clock.

07:01 - End of Clip: Promotional Overlay and Continued Play

- **07:01:** A large promotional overlay for "FIFA+" appears on the left and bottom of the screen, featuring a QR code. The live match continues in the remaining portion of the frame.
- **07:13 - 07:42:** While the overlay is on screen, the match continues. Belgium is attacking. Graphics appear showing the shot accuracy statistics for both teams, highlighting Belgium's inefficiency (17% accuracy) versus Japan's clinical finishing (50% accuracy).
- **07:43 - 08:26:** Belgium creates another chance. A cross comes in from the right, and Axel Witsel (#6) gets a head on it, but the ball is cleared by the Japanese defense.
- **08:27 - 08:58:** Japan wins a corner. Gaku Shibasaki (#7) prepares to take it, using the opportunity to further manage the clock. He takes the corner, which is headed clear by the Belgian defense.
- **08:59 - 09:40:** The cleared corner leads to a quick Belgian counter-attack, but Japan's goalkeeper Kawashima is alert and rushes out of his box to claim the ball near the corner flag, averting the danger.
- **09:41 - End:** The clip concludes with more footage of the ongoing match under the promotional overlay, with Belgium in possession, desperately trying to find a way back into the game as the clock ticks past the 60-minute mark. Close-ups show the focus and determination on the faces of players from both sides.

Overall Summary of the Extended Clip

This nearly 10-minute segment captures a pivotal period in the match. It showcases Japan's second goal by Takashi Inui, a moment of individual brilliance that put them in a commanding 2-0 lead. The subsequent minutes highlight the stark contrast in reactions: Japan's disciplined and organized defense versus Belgium's increasingly desperate attacks. The key moment of this period is Eden Hazard's shot hitting the post, a fraction of an inch away from changing the entire complexion of the game. The clip ends with Belgium still trailing by two goals but relentlessly pushing forward, setting the stage for one of the most dramatic comebacks in World Cup history.

Thank you