

# Input Matters: Evaluating Input Structure's Impact on LLM Summaries of Sports Play-by-Play

Barkavi Sundararajan and Somayajulu Sripada and Ehud Reiter  
 Department of Computing Science, University of Aberdeen, UK  
 {b.sundararajan.21, yaji.sripada, e.reiter}@abdn.ac.uk



**Motivation:** To generate NBA game summaries from rich and complex play-by-play logs and study how input structure affects factual accuracy (i.e., how faithfully the summaries reflect the input data).

## Play-by-Play Input Formats

### Unstructured

Time	Lakers	<No header>	Score	...
	Jump ball: A. Davis vs. O. Okongwu (J. Johnson gains possession)			
4:49			119-121	...
4.40	Defensive rebound		119-121	...

### Row Structure

Time	Period	Team	Primary player	Event type	Action type	...
5:00	OT1	Lakers	Anthony Davis	Jump ball	Jump ball	
4:49	OT1	Hawks	Onyeka Okongwu	Shot	Layup	...
4.40	OT1	Lakers	Rui Hachimura	Rebound	Defensive	...

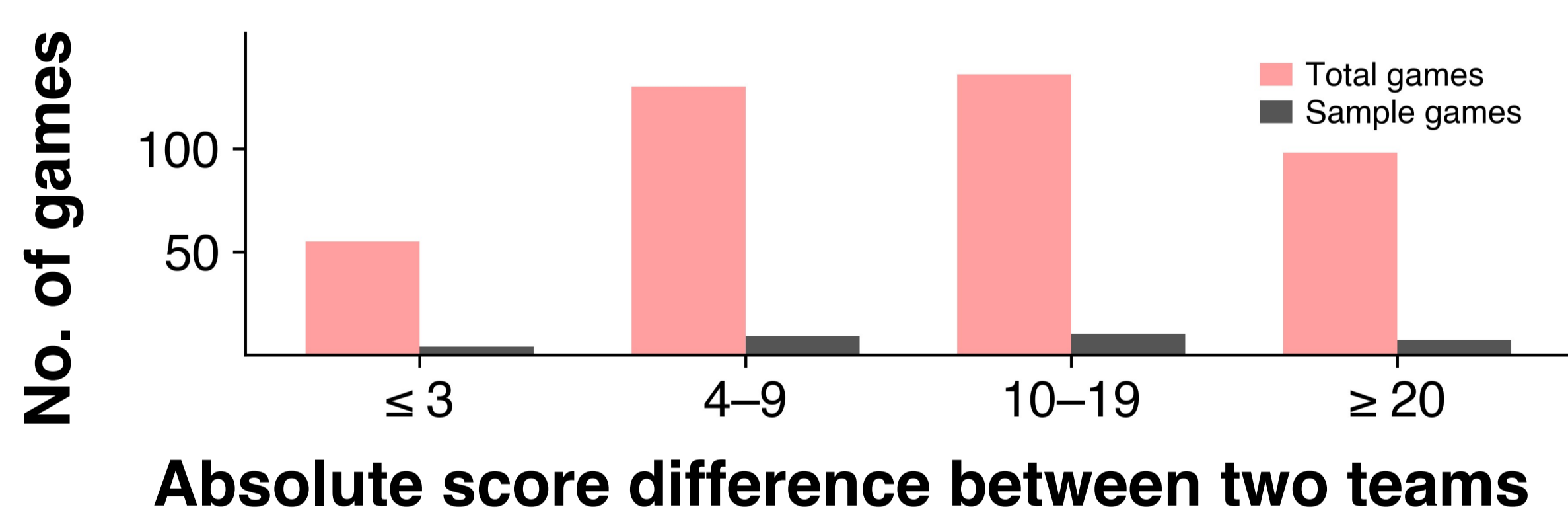
### JSON Structure (single play)

```

"time": {
  "period": "OT1",
  "clock": "4:49.0"
},
"team": "Atlanta Hawks",
"play_details": {
  "description": "Okongwu makes 2-pt layup..",
  "event": {
    "type": "Shot",
    "action": "Layup"
  },
  "players": {
    "primary_player": "Onyeka Okongwu",
    "secondary_player": "Trae Young"
  }
},
"score": {
  "points_scored": "ATL: 2",
  "cumulative": { "LAL": 119, "ATL": 121 }
}
    
```

## Game Selection and Evaluation Scope (Dec 2024 to Jan 2025, Contamination-Controlled Data)

## Manual Error Annotation Protocol and an Example (from a total of 180 annotations = 30 games \* 3 Inputs \* 2 LLMs)



Number: U, Name: N, Word-Objective: WO, Word-Subjective: WS, Context: C, Not Check: X, Other: O

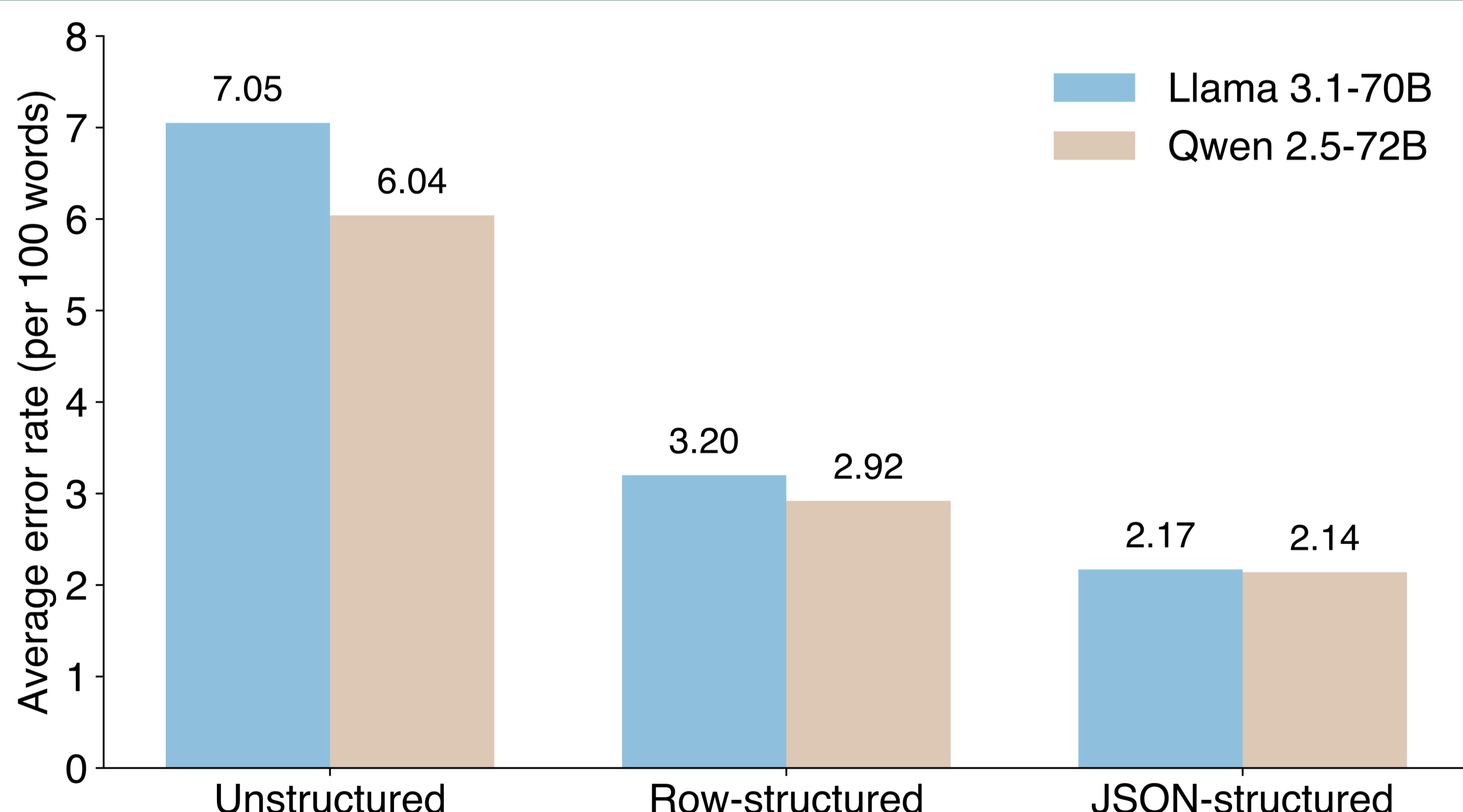
### Partial summary (unstructured)

The quarter ended with Cleveland **leading**<sup>WO</sup> 32-**36**<sup>U</sup>, thanks to **a flurry of 3-pointers from both sides**<sup>WO</sup>, including a buzzer-beater by **Christian Wood**<sup>C</sup>, who hit a 3-pointer to cap off a strong first quarter for Houston.

### List of errors.

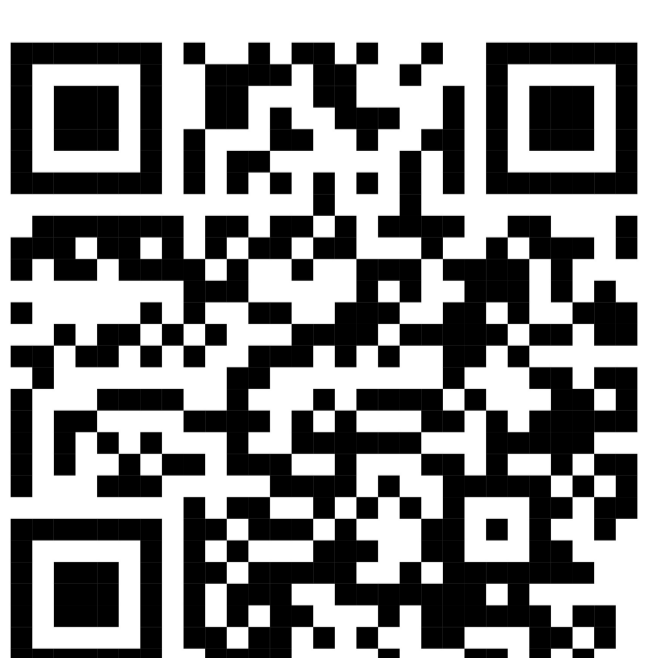
- **leading**<sup>WO</sup>: incorrect word, Cleveland was trailing at the end of Q1.
- **36**<sup>U</sup>: incorrect score (number), the correct score was 39.
- **a flurry of 3-pointers from both sides**<sup>WO</sup>: word objective error, there were only a couple of three-pointers by the Rockets.
- **Christian Wood**<sup>C</sup>: incorrect context, he did not play for Rockets or NBA in the 2025 season. Cam Whitmore is the correct player.

## Structured Inputs Yield Fewer Errors: JSON < Row << Unstructured (180 summaries)



This annotation protocol achieved 80% inter-annotator agreement between two annotators.

## Dataset



## Results

- **Input Structure strongly impacts factual accuracy:** JSON reduces error rates by 69% for Llama and 65% for Qwen compared to *unstructured*, while row-structured reduces by 54% for Llama and 51% for Qwen.
- A two-way repeated measures ANOVA shows **input structure explains > 80%** of variance (Tukey HSD: all pairwise differences are significant).
- **Error-type patterns:** Word-objective errors dominate across models; Llama made more Number errors, while Qwen made more Name and Word-subjective errors.