Map&Make

Schema Guided Text to Table Generation

*Naman Ahuja[1], *Fenil Bardoliya[1],
Chitta Baral[1], †Vivek Gupta[1]

[1]*School of Computing and Augmented Intelligence, Arizona State University*

Ira A. Fulton Schools of
Engineering
Arizona State University

ACL 2025
VIENNA

The Goal: Unstructured Text to Structured Tables

Input Text

The Hornets (5 - 15) snapped a 10 - game winning streak and sent the Knicks (4 - 17) to their seventh straight loss with Friday's 103 - 102 victory.
………………
The Hornets were led by 22 points from Gerald Henderson and 16 points and 13 boards from Al Jefferson. Lance Stephenson added 16 points (7 - 15 FG) to go with six assists and six rebounds,
………………
The Knicks were out - rebounded 41 - 34 but committed only six turnovers, three of which were charged to Samuel Dalembert. Up next, the Knicks host the Blazers on Sunday. Charlotte is off until Wednesday, when they face the Celtics.

	Wins	Losses	Field goal percentage	Three-point percentage	Three-point shots made	Three-point shots attempts	Total Rebounds	Turnovers
Hornets	5	15	48	22	4	18	41	None
Knicks	4	17					34	6

	Three-point shots made	Points	Rebounds	Field goals made	Field goals Attempted	Assists	Turnovers
J.R. Smith	1	14	4			5	
Carmelo Anthony	2	32	7			4	
Kemba Walker		11					
Gerald Henderson		22					
Al Jefferson		16	13				
Lance Stephenson		16	6	7	15	6	
Gary Neal		13					
Tim Hardaway, Jr.		13					
Samuel Dalembert							3

Evaluation Aspects

Correctness: All the information in the tables can be attributed back to the document.

Completeness: All the relevant information in the document is incorporated into the table.

Motivation

The Core Challenge: The Schema Bottleneck

Current methods for complex generation assumes of table schemas.

Either specified in instruction for prompt-based approaches or in training data for fine-tuning based approaches.

How can we generate tables without knowing the schema beforehand?

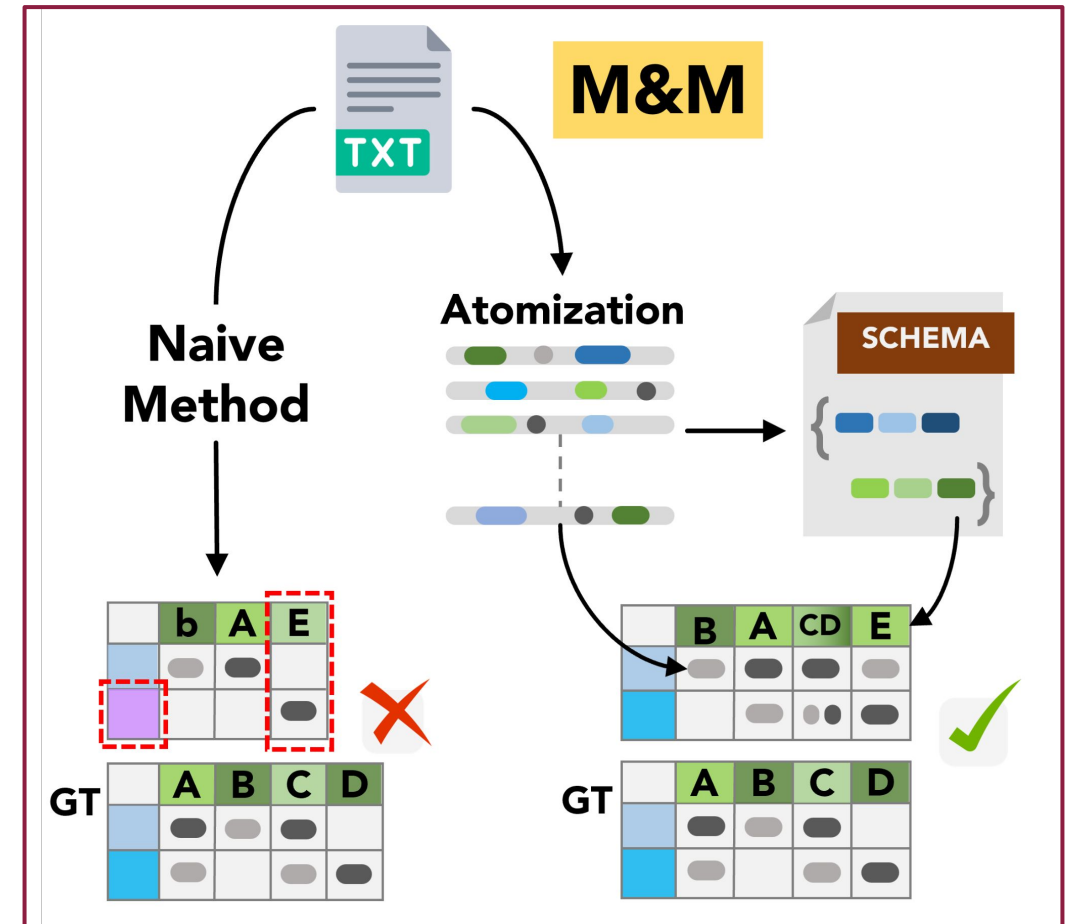## Incorporate tasks that require inference going beyond simple extraction
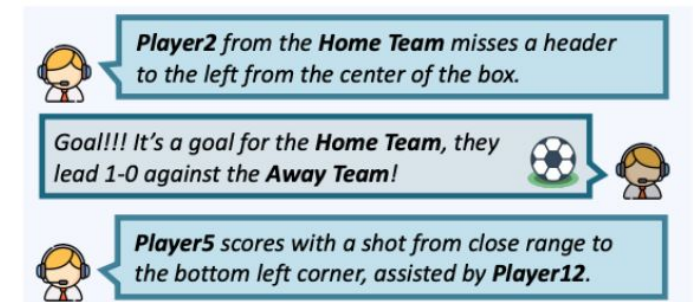
Map&Make Framework

Breaks down Table Generation into three parts

1. Atomize Document maintaining while maintaining Temporal Consistency.

2. Planning and Defining Table Schemas.

3. Table Population: Filling cell values on a predefined tabular schema.

Propositional Atomization

- Well-formed: Grammatically correct.

- Atomic: Smallest meaningful unit.

- Self-contained: Understandable without context.

- Supported: Directly based on input.

- Comprehensive: Covers all input claims.

Input Text

The Hornets (5 - 15) snapped a 10 - game winning streak and sent the Knicks (4 - 17) to their seventh straight loss with Friday's 103 - 102 victory.
....
The Hornets were led by 22 points from Gerald Henderson and 16 points and 13 boards from Al Jefferson. Lance Stephenson added 16 points (7 - 15 FG) to go with six assists and six rebounds,
....
The Knicks were out - rebounded 41 - 34 but committed only six turnovers, three of which were charged to Samuel Dalembert. Up next, the Knicks host the Blazers on Sunday. Charlotte is off until Wednesday, when they face the Celtics.

Propositional Breakdown

ATOMIC STATEMENTS

1. The Hornets' record is 5 wins.
2. The Hornets' record is 15 losses.
3. The Knicks' record is 4 wins.
4. J.R. Smith scored 14 points for the Knicks.
5. J.R. Smith had five assists for the Knicks.
6. The Knicks committed six turnovers.
7. Samuel Dalembert was charged with three turnovers.

Iterative Schema Extraction

ATOMIC STATEMENTS

1. The Hornets' record is 5 wins.
2. The Hornets' record is 15 losses.
3. The Knicks' record is 4 wins.
4. J.R. Smith scored 14 points for the Knicks.
5. J.R. Smith had five assists for the Knicks.
6. The Knicks committed six turnovers.
7. Samuel Dalembert was charged with three turnovers.

SCHEMA EXTRACTION from ATOMIC STATEMENTS
With every statement processed, the schema gets updated and returns the final state after processing all the statements.

Schema Extraction
```
{
    "Team": {
    "row_headers": ["Hornets", "Knicks"],
    "column_headers": ["Wins", "Losses",  ……, "Total Rebounds", "Turnovers"]
    },
    "Player": {
        "row_headers": ["J.R. Smith", "Carmelo Anthony", ….. , "Samuel Dalembert"],
        "col_headers": ["Three-point shots made", "Points",….., "Turnovers"]
    }
```

Steps

- Plans number of tables.

- Iteratively extracting relevant Table titles, Row headers, and Column headers based on Atomic Statements.

Iterative Table Generation

ATOMIC STATEMENTS

1. The Hornets' record is 5 wins.
2. The Hornets' record is 15 losses.
3. The Knicks' record is 4 wins.
4. J.R. Smith scored 14 points for the Knicks.
5. J.R. Smith had five assists for the Knicks.
6. The Knicks committed six turnovers.
7. Samuel Dalembert was charged with three turnovers.

Schema Extraction

```
{
    "Team": {
        "row_headers": ["Hornets", "Knicks"],
        "column_headers": ["Wins", "Losses", ......, "Total Rebounds", "Turnovers"]
    },
    "Player": {
        "row_headers": ["J.R. Smith", "Carmelo Anthony", ..... , "Samuel Dalembert"],
        "col_headers": ["Three-point shots made", "Points",....., "Turnovers"]
    }
}
```

	Wins	Losses	Field goal percentage	Three-point percentage	Three-point shots made	Three-point shots attempts	Total Rebounds	Turnovers
Hornets	5	15	48	22	4	18	41	None
Knicks	4	17	48				34	6

	Three-point shots made	Points	Rebounds	Field goals made	Field goals Attempted	Assists	Turnovers
J.R. Smith	1	14	4			5	
Carmelo Anthony	2	32	7			4	
Kemba Walker		11					
Gerald Henderson		22					
Al Jefferson		16	13				
Lance Stephenson		16	6	7	15	6	
Gary Neal		13					
Tim Hardaway, Jr.		13					
Samuel Dalembert							3

ASU

Experimental Setup - Benchmarks

1. Rotowire ([Wu et al. 2022](#))

- Dense post-game NBA summaries (2014–2017), mapped to **player & team** tables.
- Multi-table, multi-column, long texts tests schema induction & sparse fact extraction.

2. Livesum ([Deng et al. 2024](#))

- Live football commentary (Premier League)
- Generates **team-level summary** tables by aggregating events (goals, fouls, offsides, etc.).
- Each column has a difficulty rating based on linguistic variation.

3. Wiki 40B ([Parag et al. 2024](#))

- Wikipedia articles spanning across multiple domains and languages.
- Evaluates generalization to open-domain (outside sports) and **multi-table generation.**

Rotowire Cleaner Benchmark

Manual Correction to add missed facts, remove hallucinated values.

Table	Cell		Row		Col	
	H	**MI**	**H**	**MI**	**H**	**MI**
Original to Strucbench						
Team	1219	1271	8	8	627	626
Player	1390	1270	68	62	135	129
Original to Corrected						
Team	613	1137	21	50	329	528
Player	7310	1752	85	82	1000	188
Strucbench to Corrected						
Team	721	1247	21	50	385	585
Player	8104	2666	140	143	1077	271

Table 1: Total counts of corrected rows, columns, and cells across error types for Rotowire. **H** denotes Hallucination, and **MI** represents Missing Information. Here, a row or a column is flagged as hallucinated/missing if it contains at least one erroneous entry.

Experimental Setup - Baselines

- **Chain-of-Thought (CoT):** Directly organizes and fills tables using LLM step-by-step reasoning .

- **Text-Tuple-Table*:** (text → tuples, tuple integration via code, tuple→table)

    - T^3: 3-step sequential
    - T^3D: Unified, combining all steps into one LLM call..

- **Map&Make (M&M) Variants:**

    - M&M-3S: Three-step sequential.
    - M&M-U: Unified, combining all steps into one LLM call.

Models Used for Baselines:
- Closed-Source LLMs: GPT-4o, Gemini-2.0 Flash.
- Open-Source LLMs: Llama-3.3 70B Instruct.

Experimental Setup - Evaluations

| Metric	How It Works	Used For
EM / ChrF / BERT Score	Flatten both tables into (row, col, val) triples -► greedy align gold→predicted -► score exact, chrF, or embedding-F1.	Rotowire
RMSE / Error Rate	For every mapped numeric cell compute take (RMSE) or % cells wrong (ER).	Livesum
TabEval (Ramu et al.)	Linearise each table into factual sentences *(table unrolling)*, run NLI entailment set-vs-set → Correctness/Completeness/F1	Rotowire, Livesum
AutoQA (Jain et Al. 2024)	LLM auto-writes Q-A pairs from source text, then generates answers using generated table; accuracy = % answers correct.	Rotowire, Wiki40B

Why so many Metrics?

No singular metric can provide a comprehensive picture, which makes for an interesting problem to explore.

M&M Results - Complex Extraction (Rotowire)

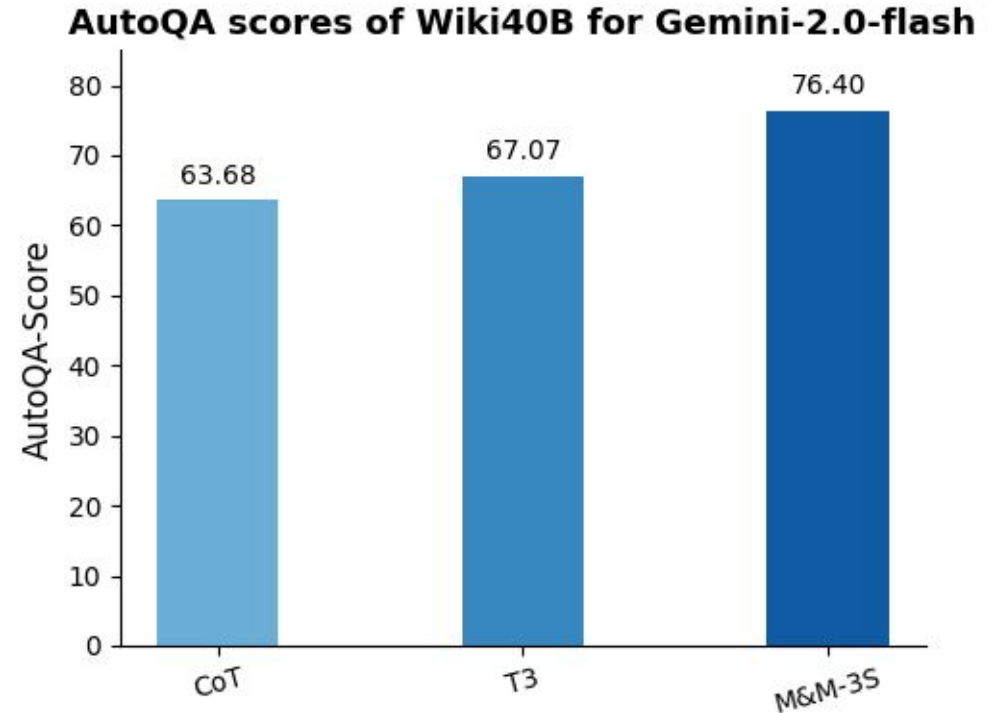Gemini-2.0 One-Shot — Completeness (TabEval) and Auto-QA

ASU
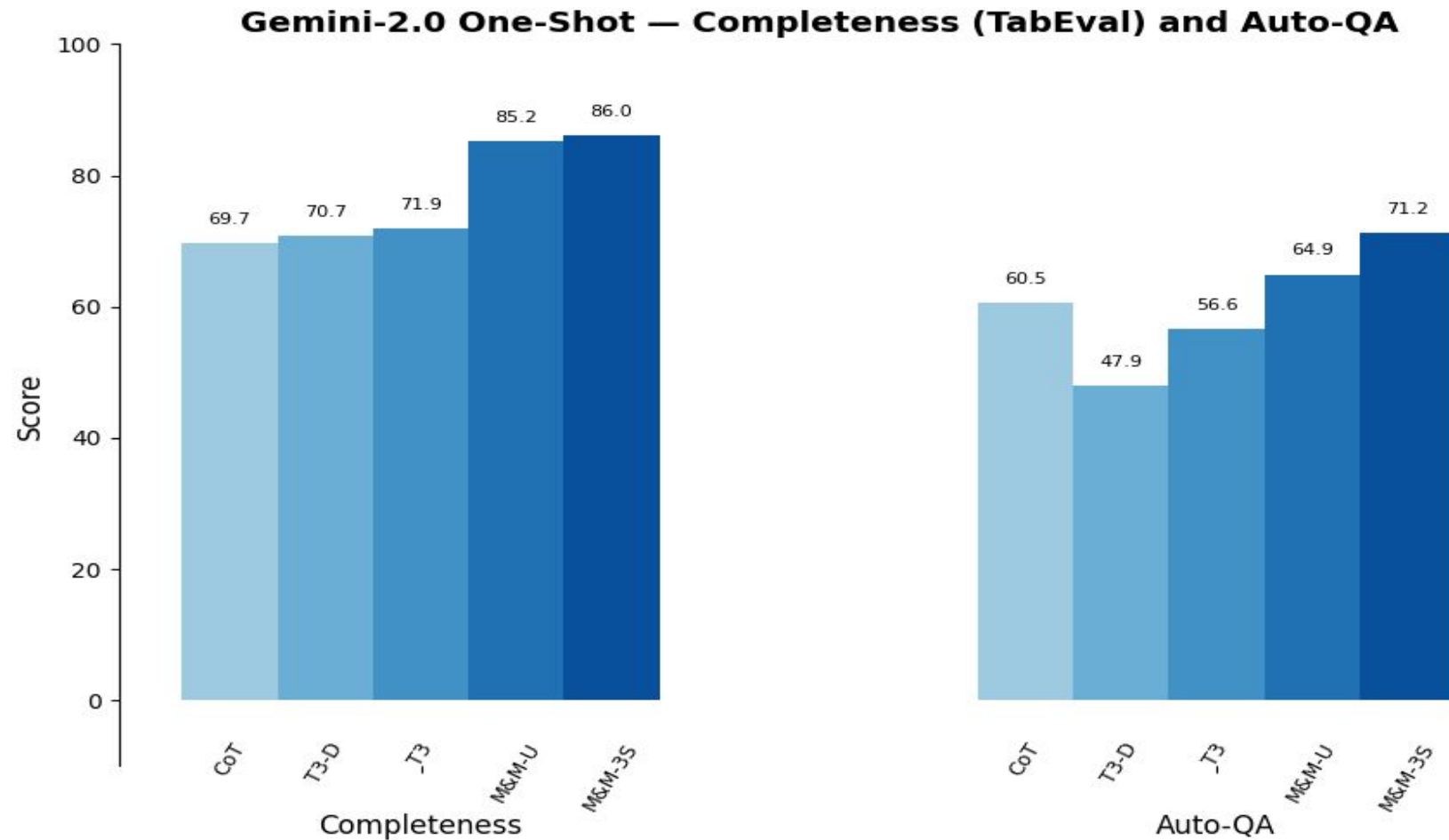
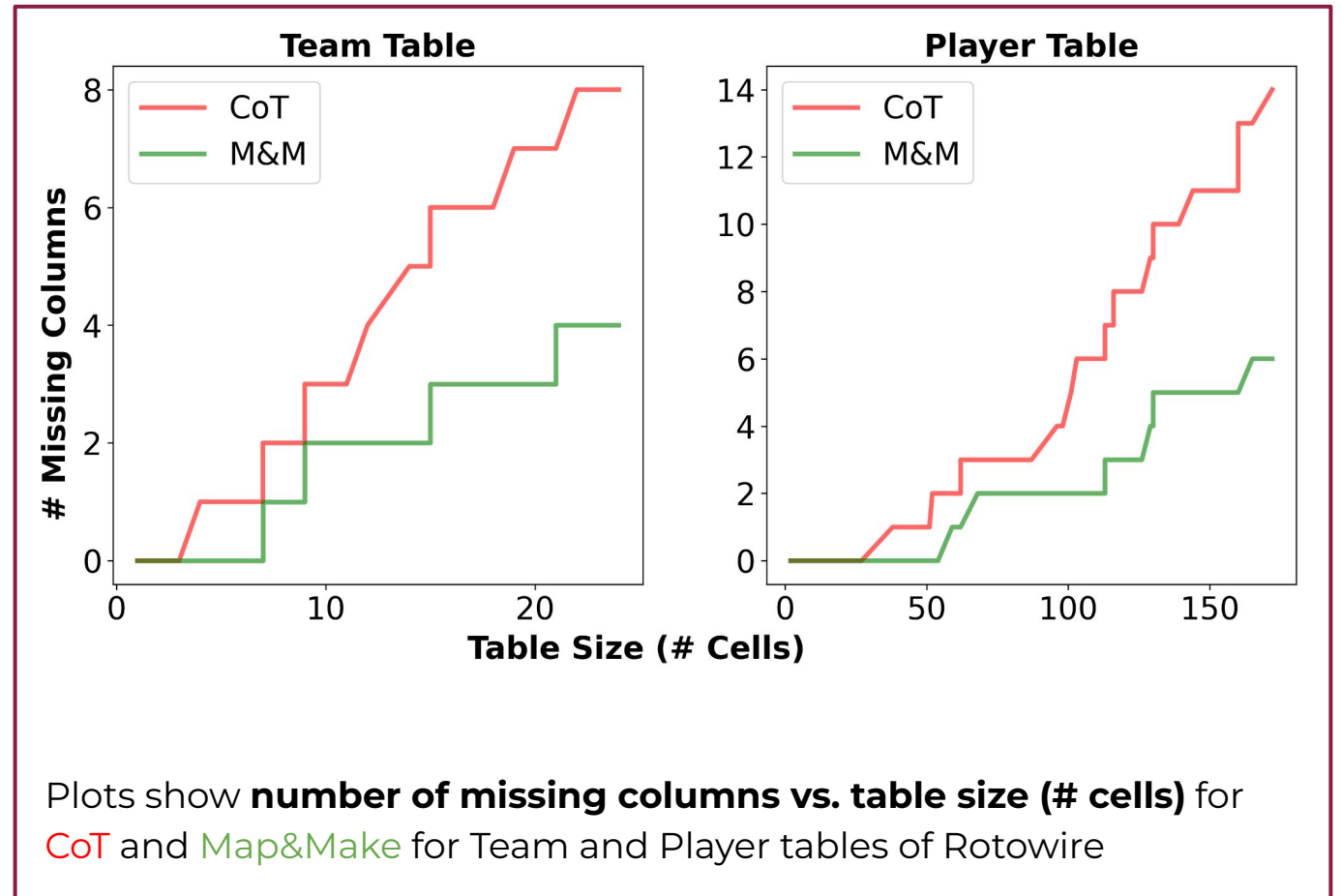# M&M Results - Numerical Aggregation (Livesum)
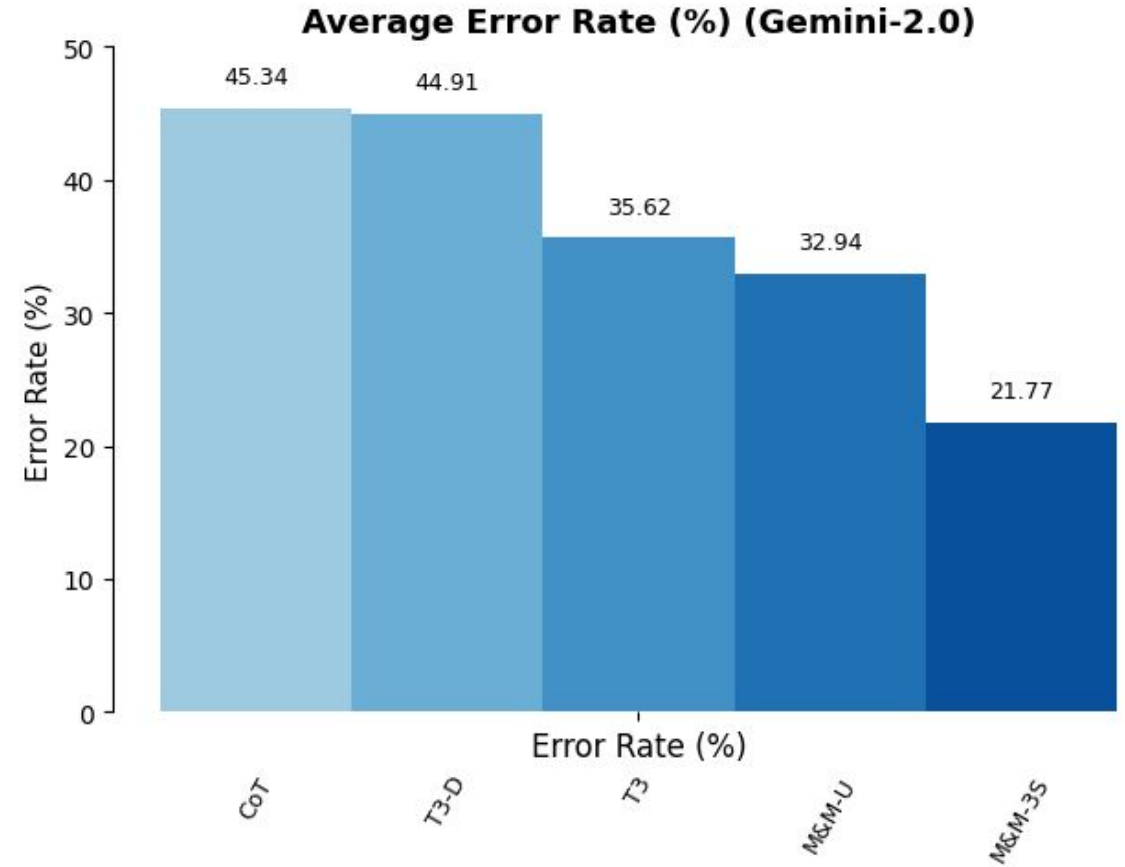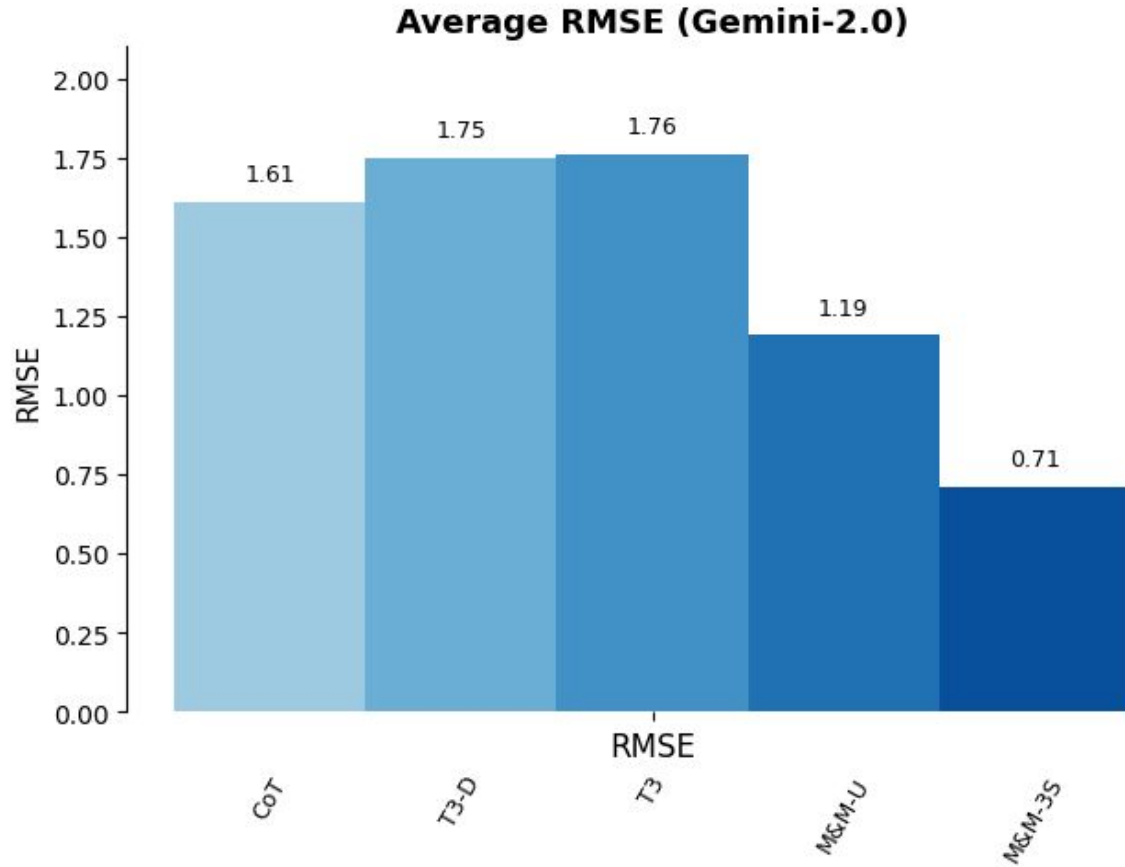
M&M Results - Open-Domain Text (Wiki 40B)

- **14%** improvement from CoT and **9%** from T3 in information coverage

- Our framework dynamically adapts to multi-table generation (Range **1-13**; Mean: **6.02**; Std: **4.33**).

**AutoQA scores of Wiki40B for Gemini-2.0-flash**

M&M Analysis - How Does Coverage Scale?

- Paraphrase Output Tables Schemas to Match Exact Table Schema as Ground Truths (prompting)

- Extract Missing and Extra Columns

Plots show **number of missing columns vs. table size (# cells)** for CoT and Map&Make for Team and Player tables of Rotowire

M&M Analysis (Mitigating Hallucinations)

- CoT shows high overcounting errors i.e overcounting events

- M&M shows stable performance across different difficulty levels

Undercounted: Cells with value < Ground Truth

Overcounted: Cells with value > Ground Truth
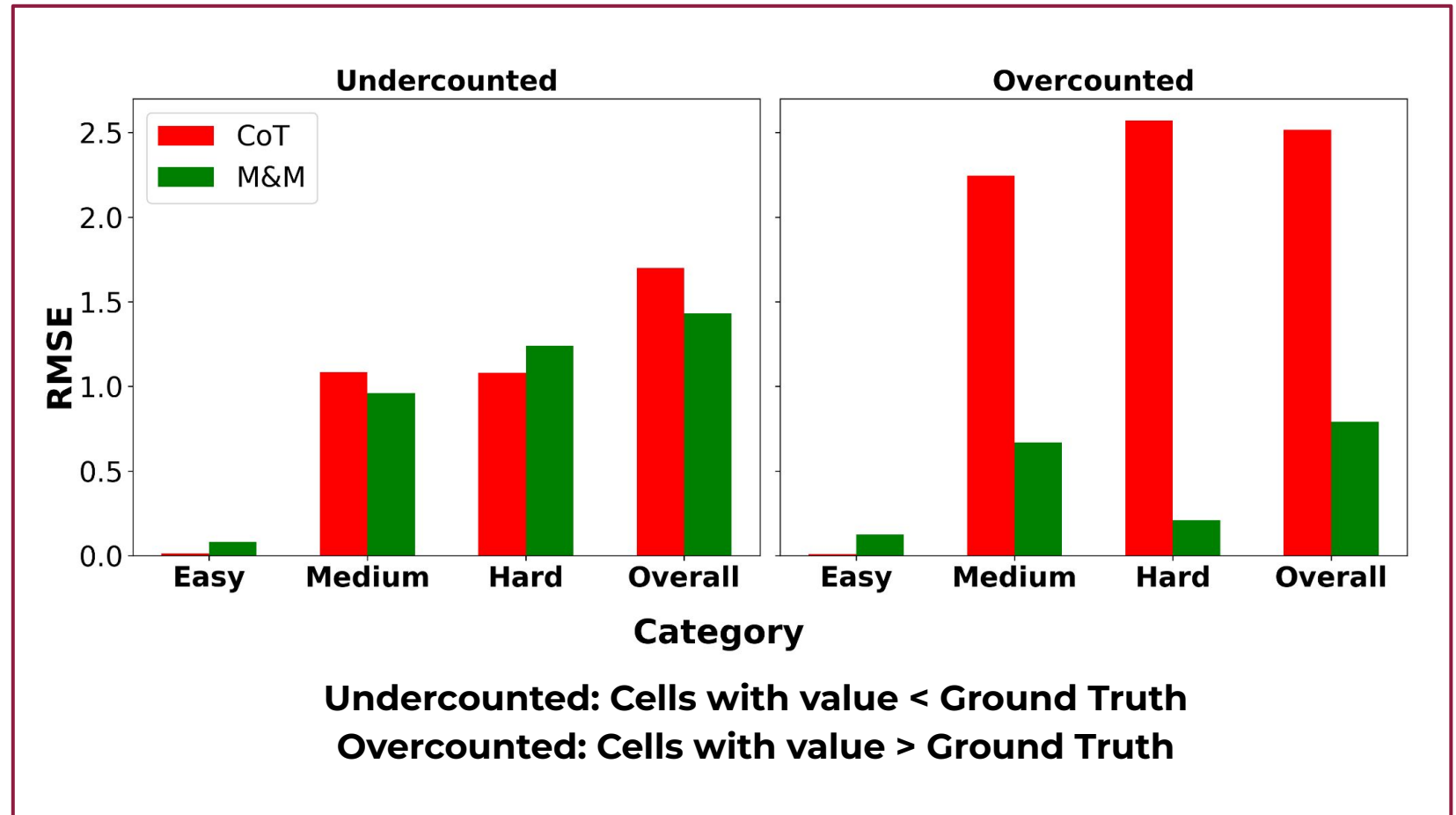
M&M Ablation Study: Are all steps necessary?

Removing any step significantly hurts performance, validating our design.

Yes! Each component is crucial.

Rotowire

M&M	Correctness		Completeness		Overall	
	Team	Player	Team	Player	Team	Player
No Ablation	**75.27**	**91.17**	**77.47**	**92.92**	**74.80**	**91.36**
- Atomization	72.91	82.80	77.37	85.61	73.28	82.42
- Iterative						
Schema	69.74	80.45	70.75	86.42	68.50	81.73
Table	70.58	83.38	69.18	89.05	67.55	84.79

Livesum

M&M	Easy		Medium		Hard		Average	
	RMSE	ER	RMSE	ER	RMSE	ER	RMSE	ER
No Ablation	0.1	4.54	0.6	23.74	0.89	33.14	0.70	21.37
- Atomization	0.05	2.90	0.64	20.36	1.59	47.89	1.01	22.96
- Iterative	0.08	3.73	1.47	37.44	2.81	65.67	1.85	36.07

ASU

Takeaways & Future Directions

- Shift from pre-defined to building dynamic schemas based on document source

- Planning Table Schema Improve Information Coverage

- Leverage Long Generation Capabilities of LLMs to improve Accuracy

Looking Ahead

Fine Tuning Smaller Models: Distilling Knowledge From Models helps but is not significant

Curation of Open-Domain Benchmarks

Referenceless Eval Metrics

SCAN FOR PAPER LINK