

# Multimodal DeepResearcher: Generating Text-Chart Interleaved Reports From Scratch with Agentic Framework

Zhaorui Yang<sup>1</sup>, Bo Pan<sup>1</sup>, Han Wang<sup>1</sup>, Yiyao Wang<sup>1</sup>, Xingyu Liu<sup>1</sup>, Luoxuan Weng<sup>1</sup>

Yingchaojie Feng<sup>2</sup>, Haozhe Feng<sup>3</sup>, Minfeng Zhu<sup>1</sup>, Bo Zhang<sup>1</sup>, Wei Chen<sup>1</sup>



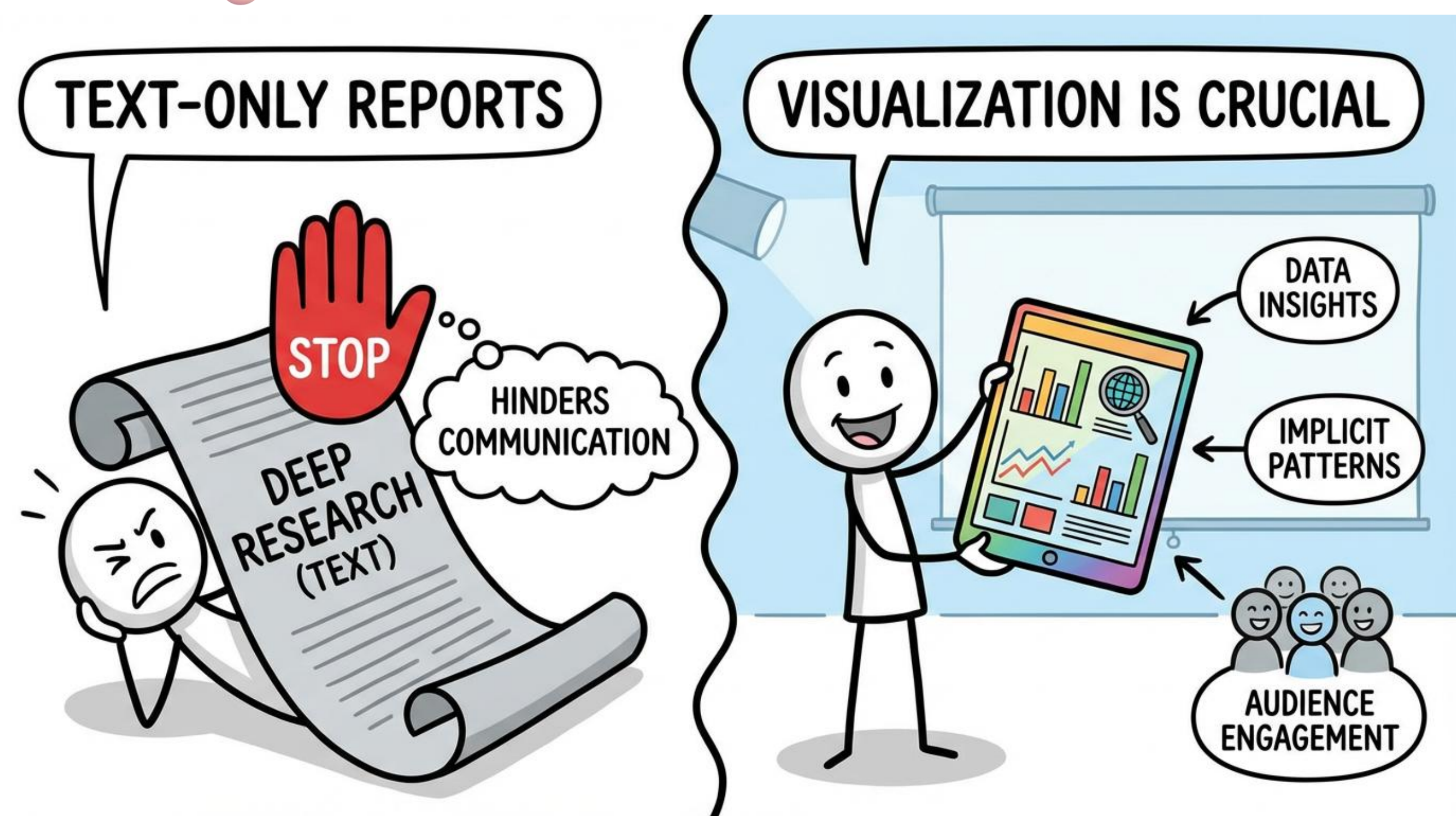
AAAI-26 / IAAI-26 / EAAI-26

Oral Presentation

Background: Current Deep Research works are text-only

Existing works focus on **text-only** content, which

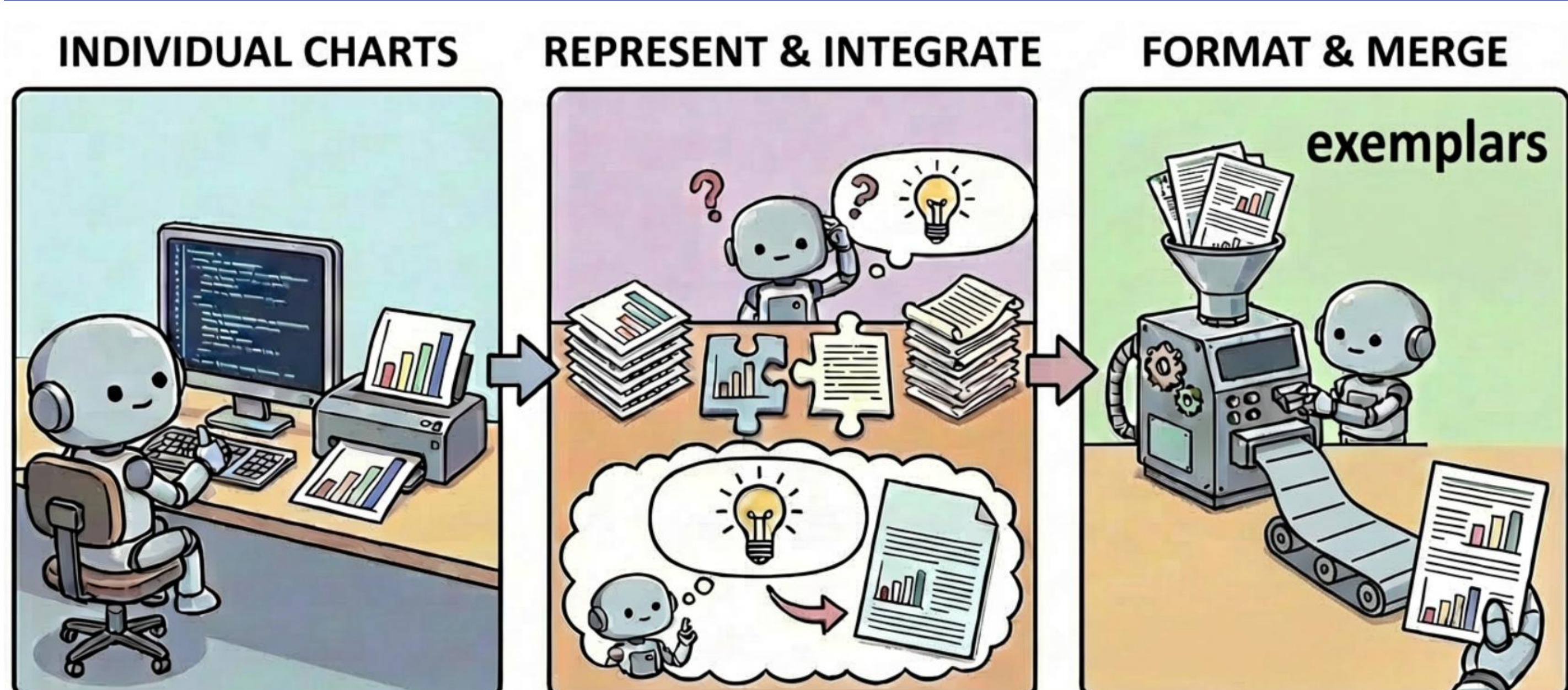
hinders 🖐 effective communication. Needs **Visualizations**



Human create coherent reports with **interleaved texts and visualizations** 📊. They integrate charts within appropriate textual context and maintain consistency.

Can **agents** generate such multimodal reports? 🤖

## Challenges of Generating Interleaved Reports



individual charts can be generated through *coding*, **but**

➤ How to **represent and integrate** them with texts?

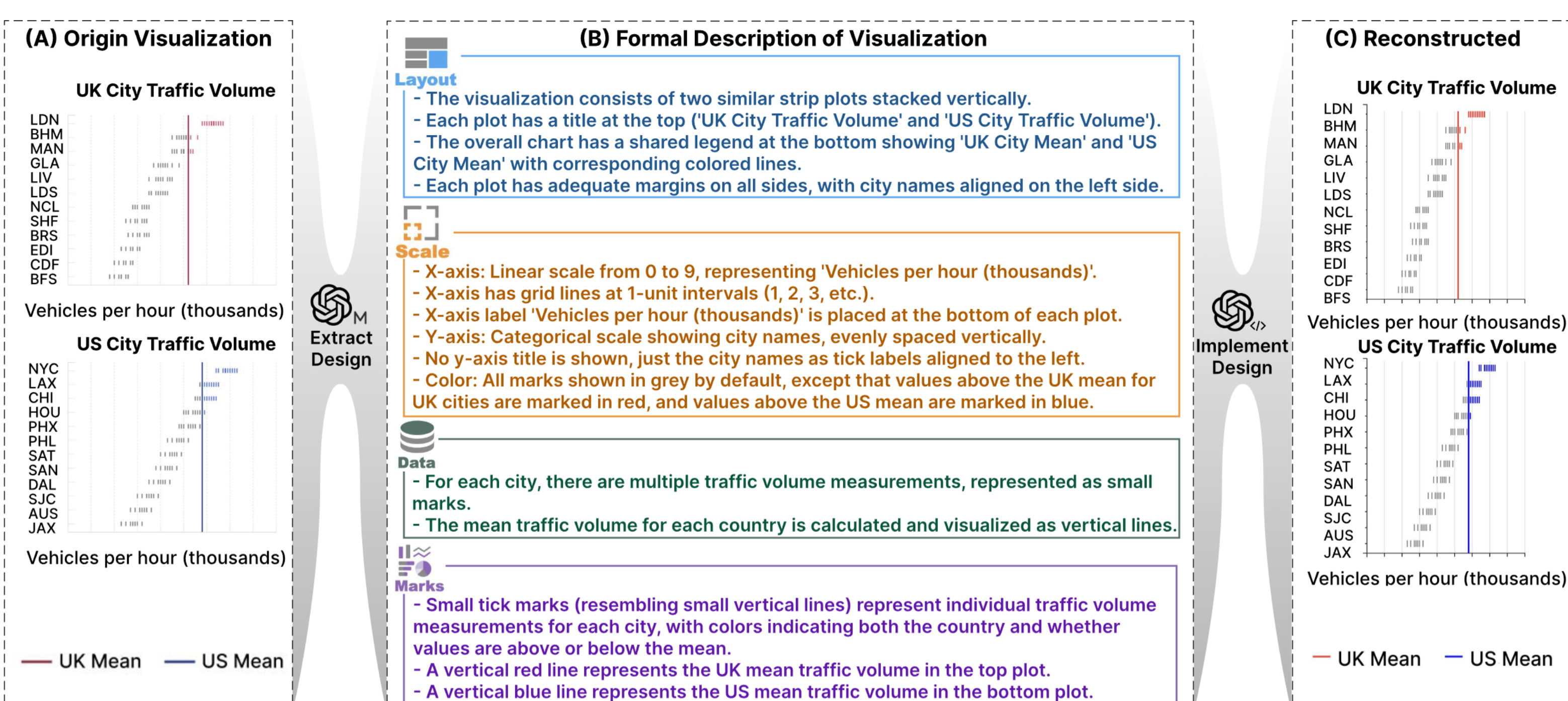
➤ How to maintain **consistency**? (chart x texts / charts)

In-context learning seems promising:

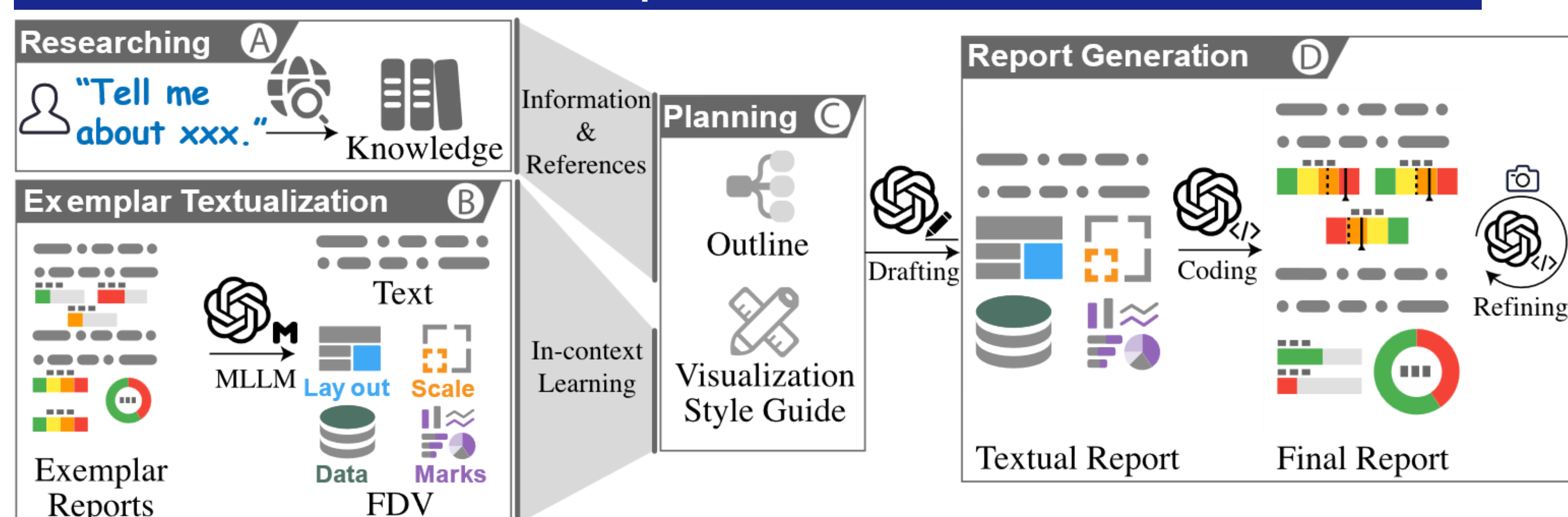
Both exemplars & outputs should be **multimodal**.

➤ Need a unified **representation** for exemplars & outputs

## Representation: Formal Description of Visualization (FDV)



## Multimodal DeepResearcher Framework



## Experimental Settings

**Input:** 100 real-world topics; **Baseline:** DataNarrative

**Criteria:** Score & Compare; **Eval:** MLLM judge & Human

## Experiment Results: Report-Level & Chart-Level

Multimodal DeepResearcher consistently **outperforms** Ours vs DataNarrative

Evaluation Metrics	Ours Win	Ours Lose	Tie
w. Claude 3.7 Sonnet			
Informativeness and Depth	75%	25%	0%
Coherence and Organization	76%	21%	3%
Verifiability	86%	5%	9%
Visualization Quality	80%	16%	4%
Visualization Consistency	78%	17%	5%
Overall	82%	16%	2%

Evaluation Metrics	Ours	DataNarrative
w. Claude 3.7 Sonnet		
Readability	8.97	8.52
Layout	9.23	8.48
Aesthetics	9.12	8.38
Data Faithfulness	9.83	9.59
Goal Compliance	9.75	9.24

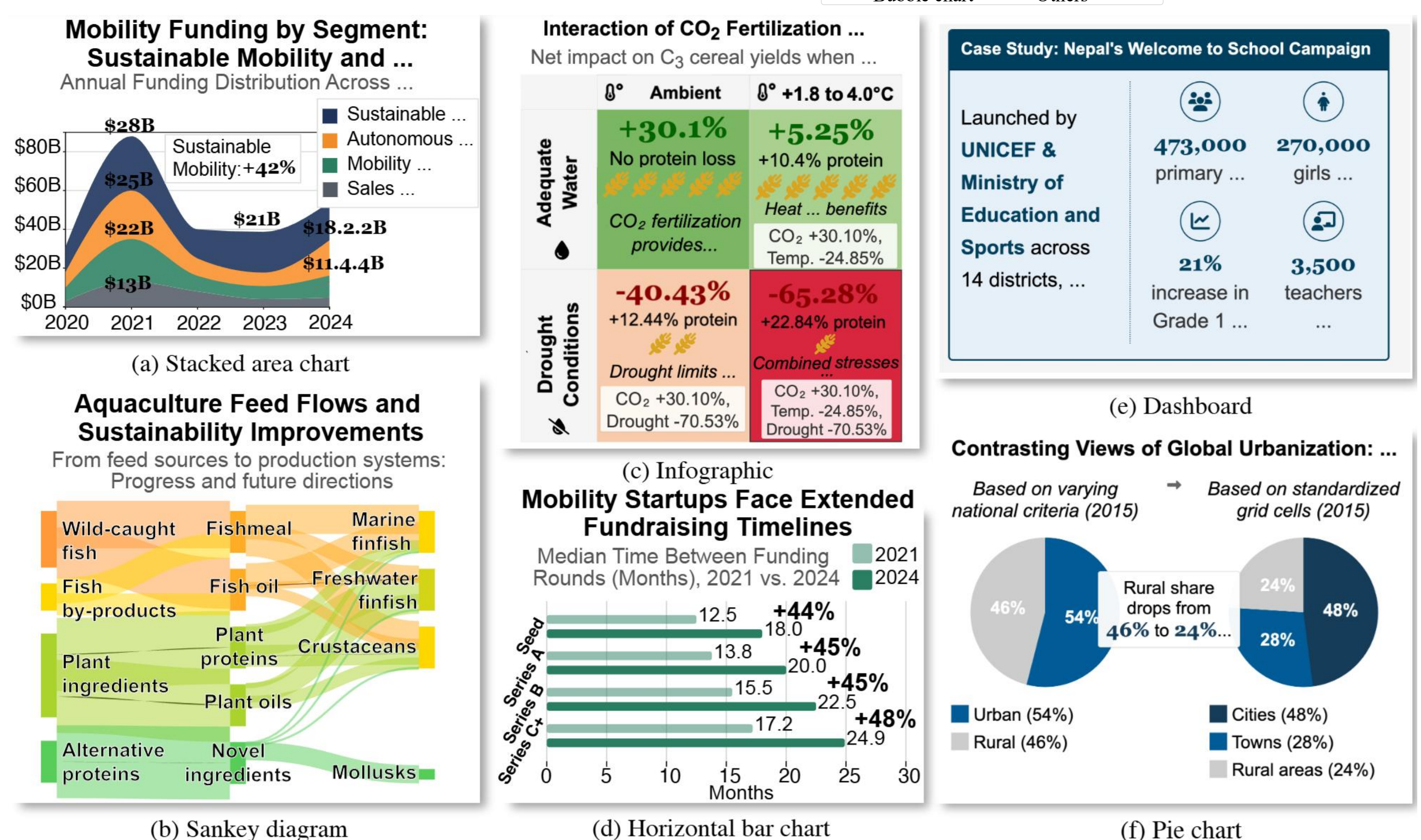
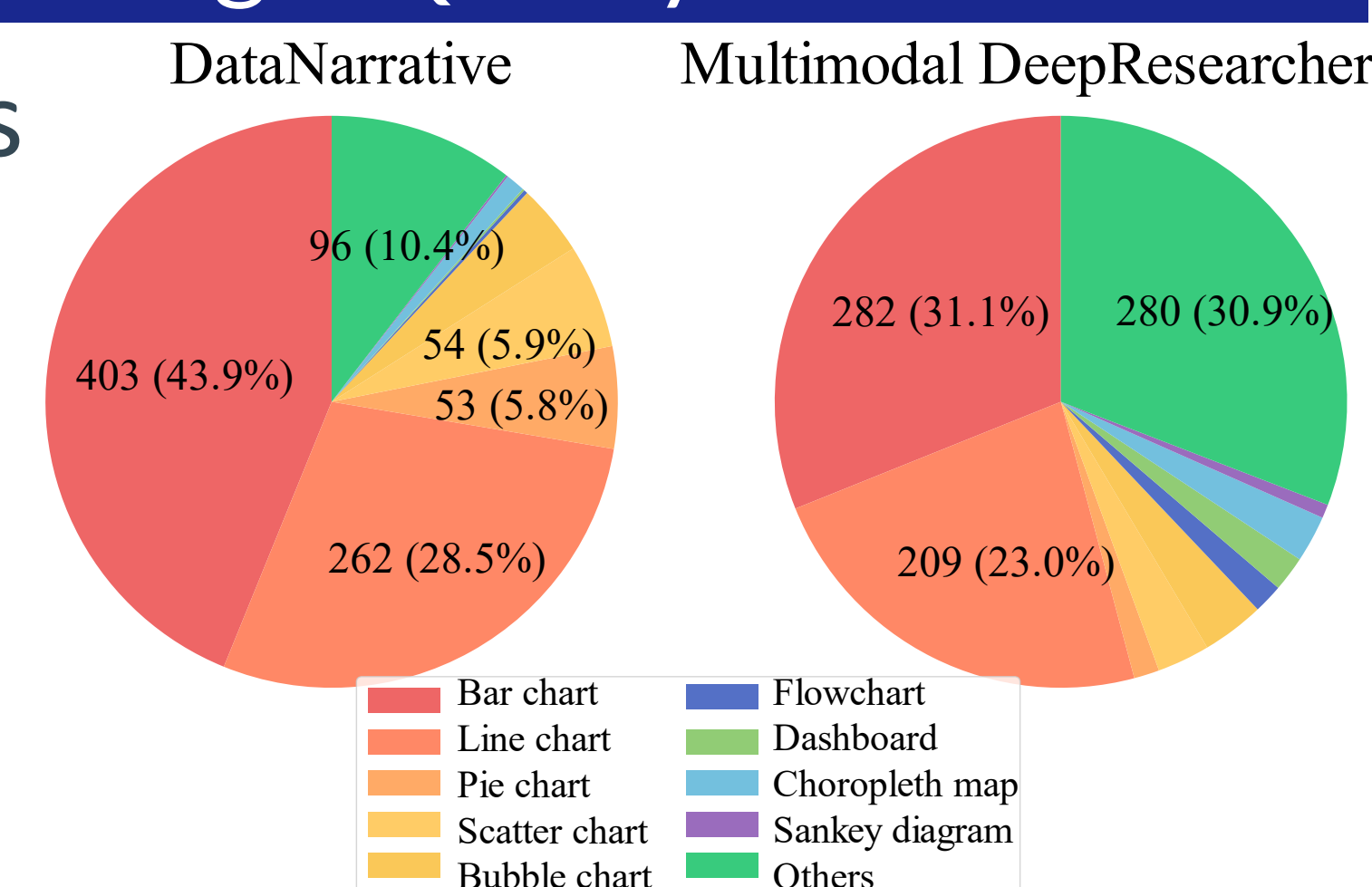
More Results in the paper

## Feature: Diverse and High-Quality Charts

Distribution of chart types

➤ Warm colors: basic

➤ More **diverse** charts



## Conclusion of Contributions

**Novel task:** Text-Chart Interleaved report generation

**Representation for visualizations:** Formal Description of Visualization (FDV)

**Framework:** end-to end agentic framework for

interleaved generation (Multimodal DeepResearcher)

Characterizes Chart → (1) Layout (2) Scale (3) Data (4) Marks

Extended Version: <https://arxiv.org/pdf/2506.02454> Presenter: Zhaorui Yang Email: [zhaorui.yang@zju.edu.cn](mailto:zhaorui.yang@zju.edu.cn)