EG UK Computer Graphics & Visual Computing (2025)
Y. Sheng and A. Slingsby (Editors)

Mind-Mapping Data Analysis with LLMs: From Vision to First Steps

Radu Jianu[1], Maeve Hutchinson[1], Natalia Andrienko[1], Gennady Andrienko[1], Mai Elshehaly[1], and Aidan Slingsby[1]

[1]City, St George's, University of London, United Kingdom

Abstract

We explore how large language models (LLMs) can support real-time visual mapping of data analysis workflows. Building on an earlier vision, we investigate if and how LLMs can decompose analytic dialogues into "analysis maps" that capture key semantic units such as questions, datasets, tasks, and findings. Using two exemplar analyses, we test both post-hoc and interactive strategies for generating these maps and experiment with prompting techniques for structuring and updating them. Results, documented in Observable notebooks, suggest that LLMs can scaffold analysis-as-network meaningfully—laying the groundwork for user-facing systems and moving beyond purely textual forms of LLM-mediated analysis.

CCS Concepts

• ***Computing methodologies*** → *Collision detection;* • ***Hardware*** → *Sensors and actuators; PCB design and layout;*

1. Introduction

Data analysis is rarely linear. Analysts refine questions, test ideas, inspect visualizations, adjust assumptions, and pivot methods in response to emerging insights. The growing use of LLMs in analysis creates new opportunities to trace and structure this process. Because analysts articulate intent, observations, and reflections as natural language prompts, the LLM interaction stream becomes a rich source of provenance, capturing not only analysis content but also its logic, evolution, and rationale. In our recent vision paper [EJS*25], we argued that this positions LLMs as not just computational assistants but also reflective partners that help construct, track, and visualize the analytic process itself. We proposed a model in which visual representations capture the state of knowledge, analytic trajectory, and semantic structure, updated dynamically through LLM interaction. Such representations can support sensemaking, communication, and collaboration in human–LLM teams.

This paper takes first steps toward that vision. We examine whether LLMs can decompose and interpret analysis workflows as structured networks of *semantic units of analysis*—research questions, datasets, tasks, methods, findings. We test two complementary strategies—*post-hoc* and *interactive*—each with distinct advantages for experimentation and application, and evaluate how well they extract, revise, and maintain "analysis maps" that evolve with the dialogue.

Our contributions are: (i) **A novel application** of LLMs for analysis mind-mapping, showing that structured visual representations of analytic reasoning can be generated dynamically from natural language; (ii) **A methodology** for testing and comparing models, prompting strategies, and interventions, including a post-hoc mode for repeatable experimentation on archived analyses; (iii) **An interactive proof-of-concept** demonstrating real-time mind-map construction alongside active analysis.

2. Related Work

Our work builds on long-standing goals in visual analytics (VA) to capture and externalize analytic reasoning. Since Thomas and Cook's foundational call to represent the *process* of analysis, researchers have developed ways to model analytic provenance [TC05]. Provenance includes both low-level interaction histories and higher-level semantic information such as questions, hypotheses, and findings. Ragan et al. [RESC15] provide a comprehensive framework for organizing such provenance to support recall, reproducibility, and collaboration. Andrienko et al. [ALA*18] further frame VA as a model-building activity centered on evolving semantic structures.

From this, researchers have explored how structured visual representations support cognitive offloading, sensemaking, and knowledge transfer. Federico et al. [FWR*17] and Andrienko et al. [ALA*18] highlight how semantic structure and visual externalization enhance interpretability and communication. Zhao et al. [ZGI*17] show how knowledge-transfer graphs support collaborative handoff, while Shrinivasan and van Wijk [SVW08] advocate for systems that scaffold VA through externalized semantic forms. This motivation also underlies systems that help users spatially organize analytic content. Jigsaw [SGLS07] supports investigative analysis through spatial arrangement of evidence, while Cook et

DOI: 10.2312/cgvc.20251221

al. [CCI*15] combine structured workspaces with task-driven system recommendations.

Recently, MindMap [WWS24] shows that LLMs can reflect on dialogue and translate it into knowledge graphs, but the graphs are used mainly internally to scaffold model inference. Our aim is different: we treat the evolving graph as a visual boundary object, co-created with and interpretable by analysts. In VA, there is growing acceptance that analyst–LLM chats should materialize into manipulable visual artifacts. Systems like LEVA [ZZZ*25] and PhenoFlow [KLJ*25] record and curate the exploration process and its outputs (e.g., rounds/steps, screenshots, cohorts) to aid reporting and iterative exploration, but they stop short of modeling the analysis itself. Closest to our work is InsightLens [WWL*25], which makes the conversation navigable via insight-centric cards with linked evidence, organized by data attributes and a two-level topic hierarchy. We differ by representing analysis as a more general, semantically typed network of questions, hypotheses, datasets, tasks, and findings intended for human co-editing and coordination.

3. Methods

Our goal was to evaluate whether LLMs can extract and maintain structured representations of data analysis as it unfolds. We adopted a two-part methodology. First, we collected exemplar analyses conducted with ChatGPT and subjected them to post-hoc decomposition, testing how different prompting strategies could reconstruct and evolve an "analysis map" from the transcript. By analysis map we mean a semantically typed network capturing key elements of an analysis (e.g., research questions, datasets, tasks, methods, findings) and their evolving relationships. Second, we implemented interactive versions of this process, where analysis and mapping co-occurred in real time. This allowed us to compare the strengths and limitations of both modes and assess the feasibility of live mapping.

3.1. Creating and Capturing Analysis Samples

Two authors independently conducted exploratory data analyses using ChatGPT. One examined the effects of COVID-19 on the rise of populism; the other focused on stop-and-search practices in London, with particular attention to racial disparities. Each session, conducted entirely through natural language, lasted about 4 hours and resulted in 96 and 57 utterances (analyst requests and LLM responses). This approach matches that used in InsightLens [WWL*25] and aligns with our broader vision of interactive, iterative co-creation of analysis flows in which evolving maps are constructed from analyst–LLM dialogues about data.

Both analyses involved iterative refinement of goals and questions, dataset identification, and preliminary exploration. Tasks included filtering, faceting, and visualization for hypothesis generation and interpretation. The LLM acted as an active assistant, transforming data, suggesting methods, generating plots, and synthesizing early insights. Each session was fully captured, including transcript, artifacts (datasets, charts), and referenced materials. These sessions form the corpus for both post-hoc and interactive decomposition.

3.2. Post-hoc Decomposition and Experimentation

In our first experiment, we treated each analysis as a fixed object and tested whether an LLM could reconstruct its analytic structure after the fact. The system prompt specified: (i) a list of semantic units to extract (e.g., research questions, datasets, analytic tasks, visualizations, insights); (ii) the network schema (node and edge fields); and (iii) the expected output format—a JSON object describing network updates via `newNodes`, `updatedNodes`, `newLinks`, and `referredNodes`.

We experimented with two prompting dimensions. First, three *transcript segmentation strategies*: (i) full transcript in one prompt (**whole**), (ii) one utterance at a time (**single**), or (iii) each request–response pair (**paired**). Second, two *concept refinement strategies*: while new semantic concepts always resulted in an added node, refinements to existing concepts could be **divergent** (new linked node) or **in-place** (update existing node with edit history). These strategies were specified in the prompts.

Experiments were run iteratively: for each analysis, we fed in one or more pre-recorded utterances (per segmentation strategy), received the LLM's response, and applied updates to the network. Each prompt included the latest utterance(s), the ten prior utterances for context, and the current network state. We automated this process via the OpenAI API.

Finally, we tested whether the LLM could revise mappings based on corrective feedback. When we felt interpretation was particularly suboptimal, we added correction requests (e.g., merge or relabel nodes), logging these along with the triggering exchange and including them in future prompts. This allowed us to examine responsiveness to feedback and incorporation of corrections.

All prompts, LLM responses, and evolving networks produced during this stage can be browsed in an interactive Observable notebook: `https://observablehq.com/@rdjianu/mind-mapping-analyses`.

3.3. Interactive Analysis Mapping

To test whether live analysis mapping could work in practice, we implemented an interactive setup where analysis and interpretation occurred concurrently. The system responded to analyst requests as part of an ongoing data workflow while simultaneously mapping the exchange into a semantic network.

We tested two architectural variants. The first used two agents: an assistant that addressed analysis queries and a *scribe* that interpreted each request–response pair and returned structured updates (nodes and links) to the map. The second used a single dual-role agent, prompted to both answer the query and extract a map update. It returned a Markdown response for the user and a JSON update processed in the background to evolve the network.

Unlike the post-hoc mode, this setup reflects practical integration: how such a system could support analysts without interrupting their flow. However, it also introduces variability: because the LLM generates new responses each time, results are harder to compare or reproduce.

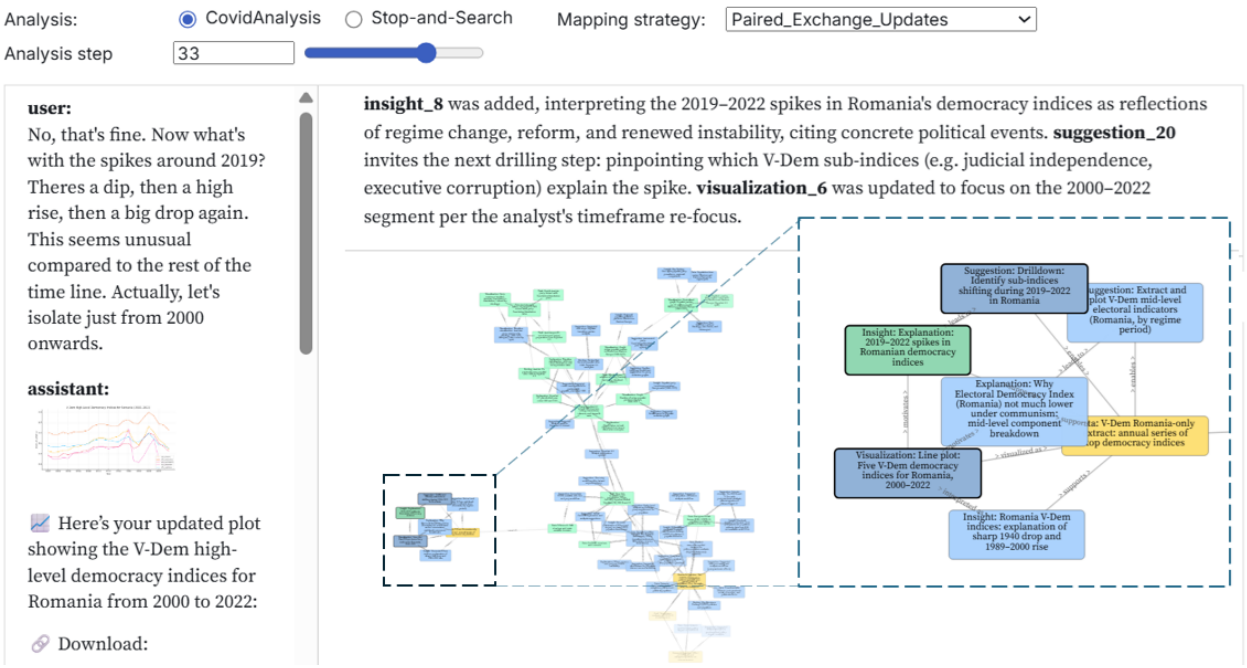

Figure 1: *Our Observable notebook captures the effectiveness of different prompts in LLM-assisted data analysis mapping. Experimenters can step through pre-recorded analysis exchanges (left) and see how they are gradually mapped into an analysis network (right). Explanations of the changes made to the network at each step are also captured (upper-right).*

3.4. Technical Details

All experiments were conducted using OpenAI's GPT-4.1 model via the API. The semantic networks were visualized using a custom D3-based force-directed layout implemented in JavaScript.

4. Results

We discuss the effectiveness of LLM-generated analysis maps under two experimental settings, post-hoc and interactive, focusing on how different prompting strategies influence quality. We assess the model's ability to decompose analysis dialogue into structured maps, refine and update concepts over time, and respond to corrective feedback. We note that evaluation was qualitative, by the authors, without quantitative measures of semantic faithfulness. We reflect on notable tradeoffs in granularity, interpretability, and reproducibility.

4.1. Post-hoc Decomposition Performance

Across both analyses, LLMs were able to extract meaningful semantic structure from transcript data. The **whole** segmentation strategy, where the full transcript was processed at once, yielded broad but often overgeneralized maps, missing task-level detail and subtle shifts in research framing. In contrast, the **single** and **paired** strategies produced more faithful and coherent structures that better captured the analysis trajectory.

Among these, the **paired** strategy, where each analyst request was coupled with the corresponding response, led to more aligned and interpretable maps. Pairing preserved the relationship between prompts and replies, reducing the excessive granularity of the **single** strategy, which often produced speculative or redundant nodes.

Results were also shaped by the **concept refinement strategy**. **Divergent refinement**, where each update generated a new node, yielded detailed maps with semantically similar elements, which is useful for tracing reasoning or supporting knowledge-graph applications. In contrast, **in-place refinement** generally produced more coherent structures and let the model integrate updates more effectively, likely due to fewer, more stable nodes with richer content.

When map quality degraded due to ambiguity or misinterpretation, correction prompts proved effective. Mid-decomposition edits improved outcomes and were often generalized correctly by the model in later steps. This suggests that perfect prompting is not essential; lightweight, in-context feedback can meaningfully steer the mapping process.

4.2. Interactive Mapping Outcomes

The interactive setup confirmed that real-time analysis mapping is feasible and responsive. Of the two configurations tested, the *dual-role* approach—where one LLM handled both analysis generation and map interpretation—outperformed the two-agent setup. It was faster and more fluid, returning network updates with analysis responses without extra processing, and avoided duplication or mismatch by consolidating interpretation within a single model.

We were initially concerned that the LLM might struggle to distinguish network correction prompts from analysis requests, but

this was not the case. Corrections could be interleaved directly into the dialogue without confusion. The LLM consistently interpreted them correctly, and we found that offering feedback mid-session felt natural and not overly disruptive. These observations suggest that lightweight, embedded corrections can effectively guide mapping without breaking the flow of analysis.

5. Discussion

Across both post-hoc and interactive setups, the model interpreted naturalistic exchanges into structured representations of goals, questions, data, tasks, and findings. This supports our vision that LLMs can help externalize data analysis as a dynamic, visual process. However, the work is preliminary and comes with limitations, outlined below with directions for future research.

Post-hoc mapping as a testbed for experimentation: Post-hoc decomposition offers strong control and repeatability, supporting systematic testing of prompting strategies on fixed inputs and side-by-side comparison of network outputs in environments like Observable. It also benefits from the feature-rich ChatGPT client, which supports file handling, multimodal input, and code execution—capabilities absent from the OpenAI API. This makes post-hoc mapping well-suited for prototyping and refining prompts before full system integration.

Interactive mapping toward real-world integration: Interactive mapping mirrors real analysis and is essential for building and evaluating future user-facing tools. It enables examination of how analysts experience mapping, the cognitive overhead involved, and whether evolving visual traces aid reflection or decision-making. It also supports integrated workflows where analysts reference map elements to drive analysis (e.g., "Can you refine this [selected] observation into a testable hypothesis?"). Achieving this requires coupling the model to a persistent, manipulable network, feasible only in an interactive setup. However, the OpenAI API lacks key agent-level features such as persistent memory, file uploads, and visual rendering, so interactive experimentation requires non-trivial infrastructure to replicate ChatGPT client capabilities.

Output variability and model/prompt sensitivity: Even with identical inputs, prompts, and models, map outputs varied. Minimal in-context feedback often improved consistency. Prompt wording also mattered: beyond the broad strategies in Section 3.1, changes such as clarifying node naming, summarisation, or definitions of analytic concepts influenced results. Model choice was another factor: we began with GPT-4o, then switched to GPT-4.1 for formal testing, noting substantial improvements in structure and coherence. We did not formally or systematically study these sources of variation—a clear limitation of this exploratory work—but doing so remains an important avenue for future research.

Analysis sample and evaluation limitations: The analysis sample was small—two author-led analyses. While these provided controlled, information-rich testbeds, the scope limits generalizability. Prompt design was also tuned on one transcript, risking overfitting to its style. Evaluation was purely qualitative, relying on author inspection rather than objective measures of semantic accuracy or reproducibility. Thus, findings should be interpreted as exploratory. Future work should use more diverse analyses and apply rigorous quantitative and qualitative evaluation to assess both analytic validity and user experience.

6. Conclusion

We take first steps toward LLM-supported analysis mapping by testing whether analytic dialogue can be structured into evolving semantic networks. Through static and interactive experiments, we show that LLMs can extract and maintain representations of analytic structure, shaped by prompt design and granularity choices. Our interactive prototype demonstrates real-time mapping during analysis, laying groundwork for user-facing systems and empirical studies. Future work will pursue user studies, bidirectional map interactions, and integration into practical analytic workflows.

References

[ALA*18] ANDRIENKO N., LAMMARSCH T., ANDRIENKO G., FUCHS G., KEIM D., MIKSCH S., RIND A.: Viewing visual analytics as model building. In *Comput. Graph. Forum* (2018), vol. 37, pp. 275–299. 1

[CCI*15] COOK K., CRAMER N., ISRAEL D., WOLVERTON M., BRUCE J., BURTNER R., ENDERT A.: Mixed-initiative visual analytics using task-driven recommendations. In *Proc. IEEE Conf. Visual Analytics Sci. Technol. (VAST)* (2015), pp. 9–16. 2

[EJS*25] ELSHEHALY M., JIANU R., SINGSBY A., ANDRIENKO G., ANDRIENKO N.: Designing for collaboration: Visualization to enable human-llm analytical partnership. *IEEE Comput. Graph. Appl. (to appear)* (2025). 1

[FWR*17] FEDERICO P., WAGNER M., RIND A., AMOR-AMOROS A., MIKSCH S., AIGNER W.: The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *Proc. IEEE Conf. Visual Analytics Sci. Technol. (VAST)* (2017), pp. 92–103. 1

[KLJ*25] KIM J., LEE S., JEON H., LEE K.-J., BAE H.-J., KIM B., SEO J.: Phenoflow: A human-llm driven visual analytics system for exploring large and complex stroke datasets. *IEEE Trans. Vis. Comput. Graph. 31*, 1 (2025), 470–480. 2

[RESC15] RAGAN E. D., ENDERT A., SANYAL J., CHEN J.: Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2015), 31–40. 1

[SGLS07] STASKO J., GORG C., LIU Z., SINGHAL K.: Jigsaw: Supporting investigative analysis through interactive visualization. In *Proc. IEEE Symp. Vis. Analytics Sci. Technol. (VAST)* (2007), pp. 131–138. 1

[SVW08] SHRINIVASAN Y. B., VAN WIJK J. J.: Supporting the analytical reasoning process in information visualization. In *Proc. SIGCHI Conf. Human Factors Comput. Syst.* (2008), pp. 1237–1246. 1

[TC05] THOMAS J. J., COOK K. A. (Eds.): *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Comput. Soc., 2005. 1

[WWL*25] WENG L., WANG X., LU J., FENG Y., LIU Y., FENG H., HUANG D., CHEN W.: Insightlens: Augmenting llm-powered data analysis with interactive insight management and navigation. *IEEE Trans. Vis. Comput. Graph. 31*, 6 (2025), 3719–3731. 2

[WWS24] WEN Y., WANG Z., SUN J.: Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models. In *Proc. Annu. Meet. Assoc. Comp. Linguist. (ACL)* (2024), pp. 10370–10388. 2

[ZGI*17] ZHAO J., GLUECK M., ISENBERG P., CHEVALIER F., KHAN A.: Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE Trans. Vis. Comput. Graph. 24*, 1 (2017), 340–350. 1

[ZZZ*25] ZHAO Y., ZHANG Y., ZHANG Y., ZHAO X., WANG J., SHAO Z., TURKAY C., CHEN S.: Leva:using large language models to enhance visual analytics. *IEEE Trans. Vis. Comp. Graph. 31*, 3 (2025). 2