Using Large Language Models to Enhance Visual Analytics

L·E·V·A

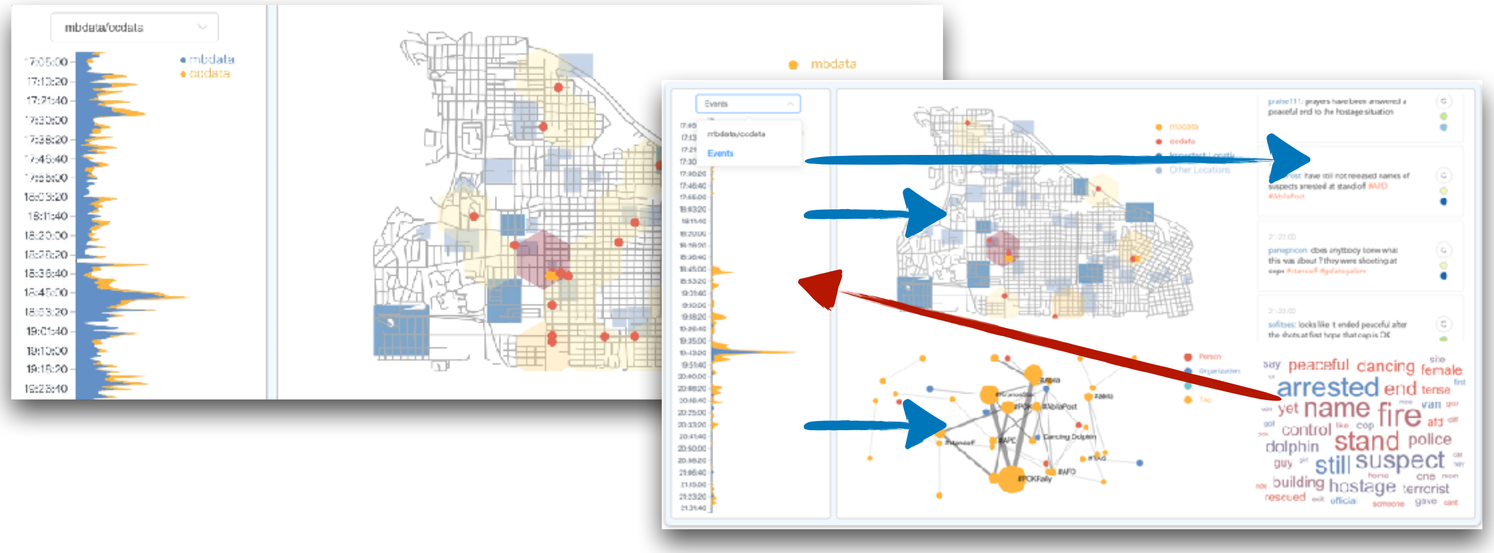
Yuheng Zhao[1], Yixing Zhang[1], Yu Zhang[2], Xinyi Zhao[1], Junjie Wang[1], Zekai Shao[1], Cagatay Turkay[3], Siming Chen[1]

[1] School of Data Science, Fudan University
[2] Department of Computer Science, University of Oxford
[3] Centre for Interdisciplinary Methodologies, University of Warwick

IEEE Transactions on Visualization and Computer Graphics, 2024

Motivation

👨 Are there any challenges when conducting visual analytics?



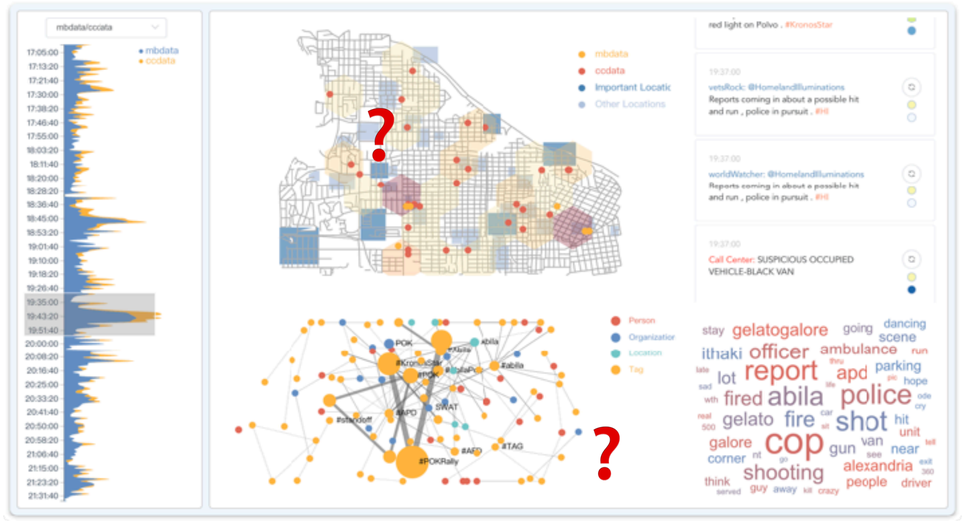
An example system from VAST Challenge 2021 MC3

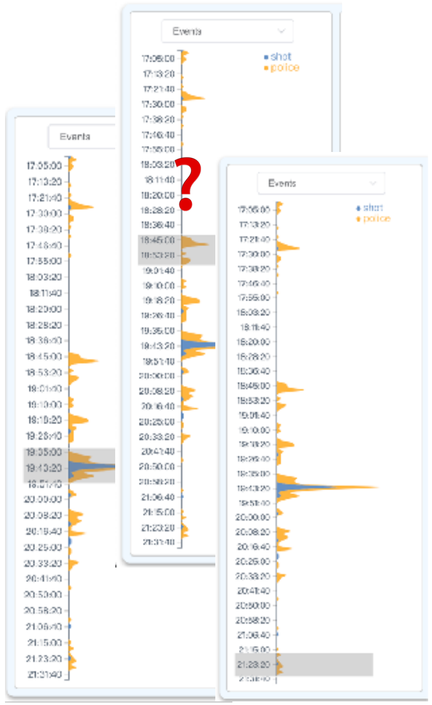
1. Onboarding stage:

Users may encounter challenges with complex systems, particularly when they are unfamiliar with visualization and interaction design principles.

Motivation

👨 Are there any challenges when conducting visual analytics?



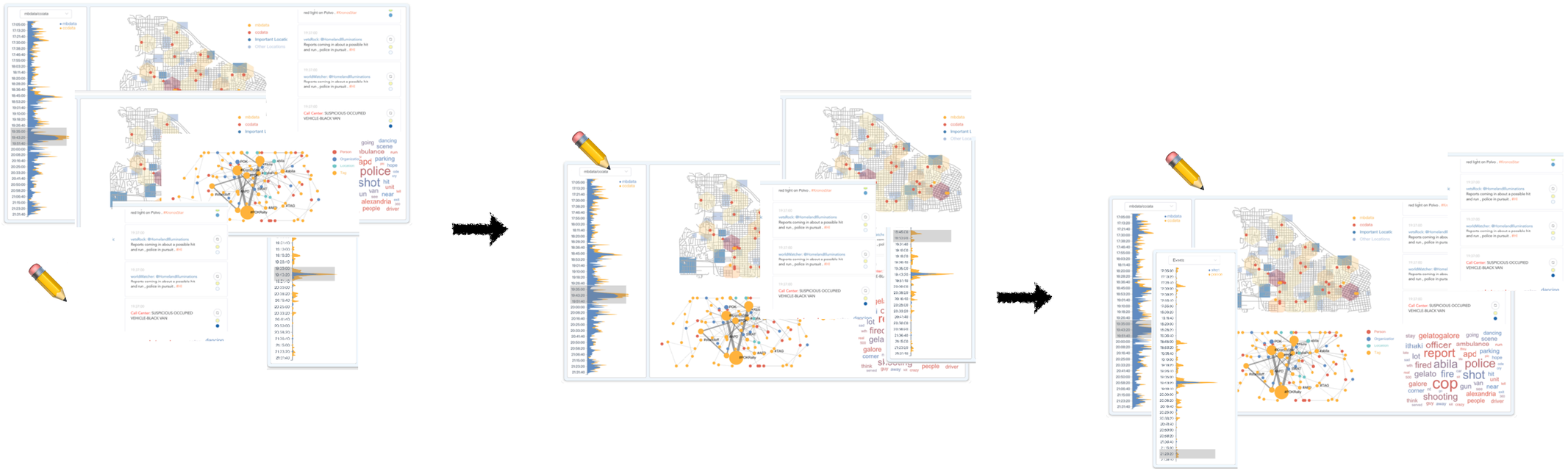


2. Exploration stage:

A lot of information needs to be observed and processed, resulting in low exploration efficiency

Motivation

👨🏻 Are there any challenges when conducting visual analytics?

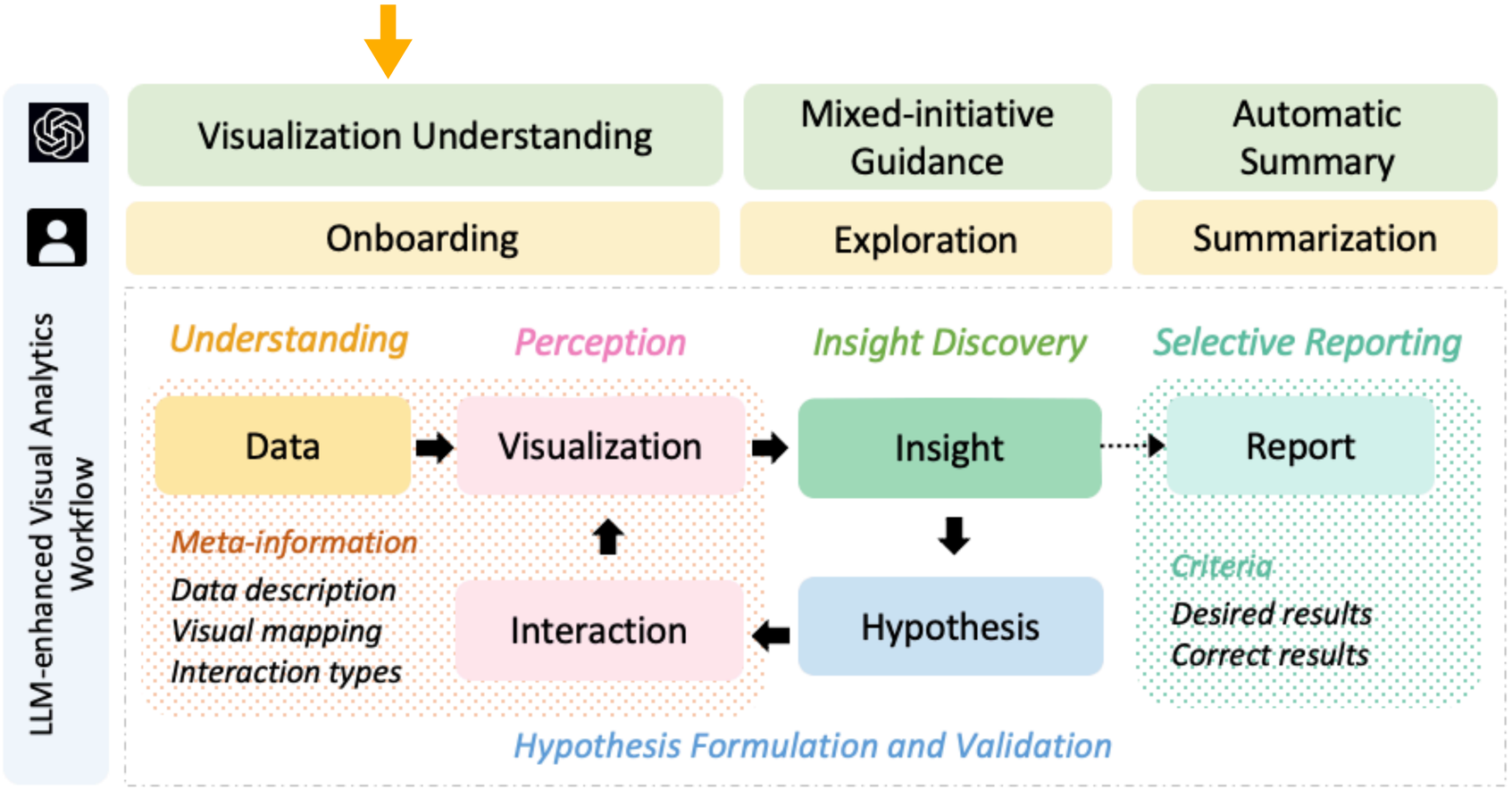


3. Summary stage:

To summarize the exploration results, users need to take notes and screenshots manually.

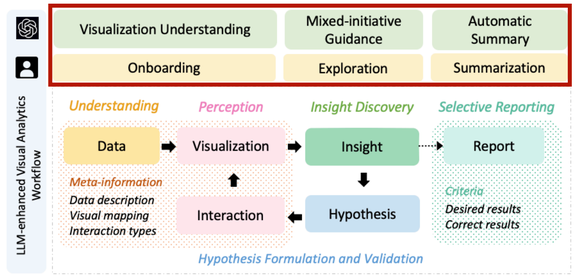
Motivation

Where to enhance User's VA workflow?



Motivation

Why integrate LLMs into the VA pipeline?



Intelligent VA requires the assistant:

Generate tutorials for VA system

↓

Recommend tasks and analyze data

↓

Summarize exploration results and generate reports

➡

Capabilities the model should have:

Understand the design of VA system

↓

Write code, or use tools to analyze data

↓

Story summarization, slides generation (e.g. LaTeX)

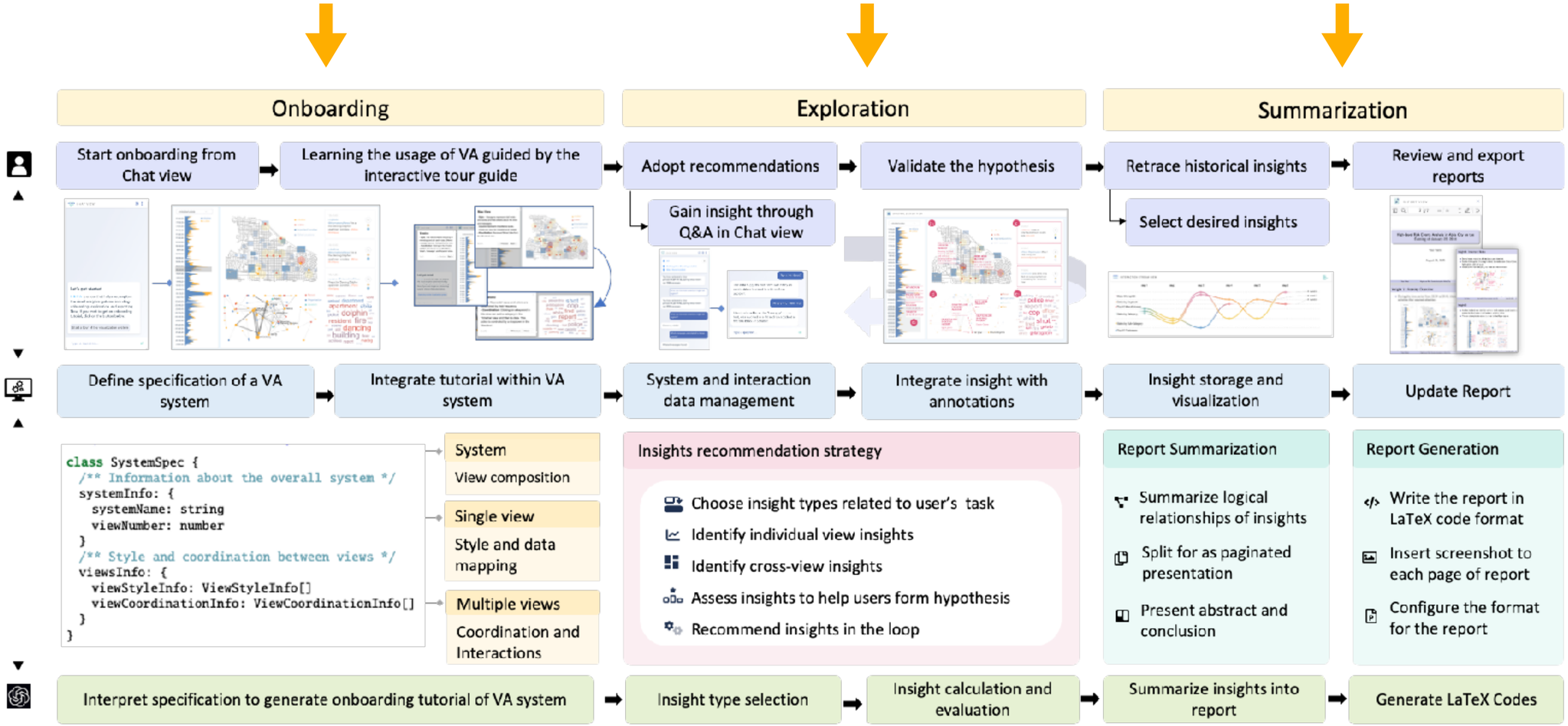
⬅

💡 **Why use LLM?**

The LLM, as a model with **generalized capabilities**, can support these tasks while **avoiding the cost for specific model training and splicing**.

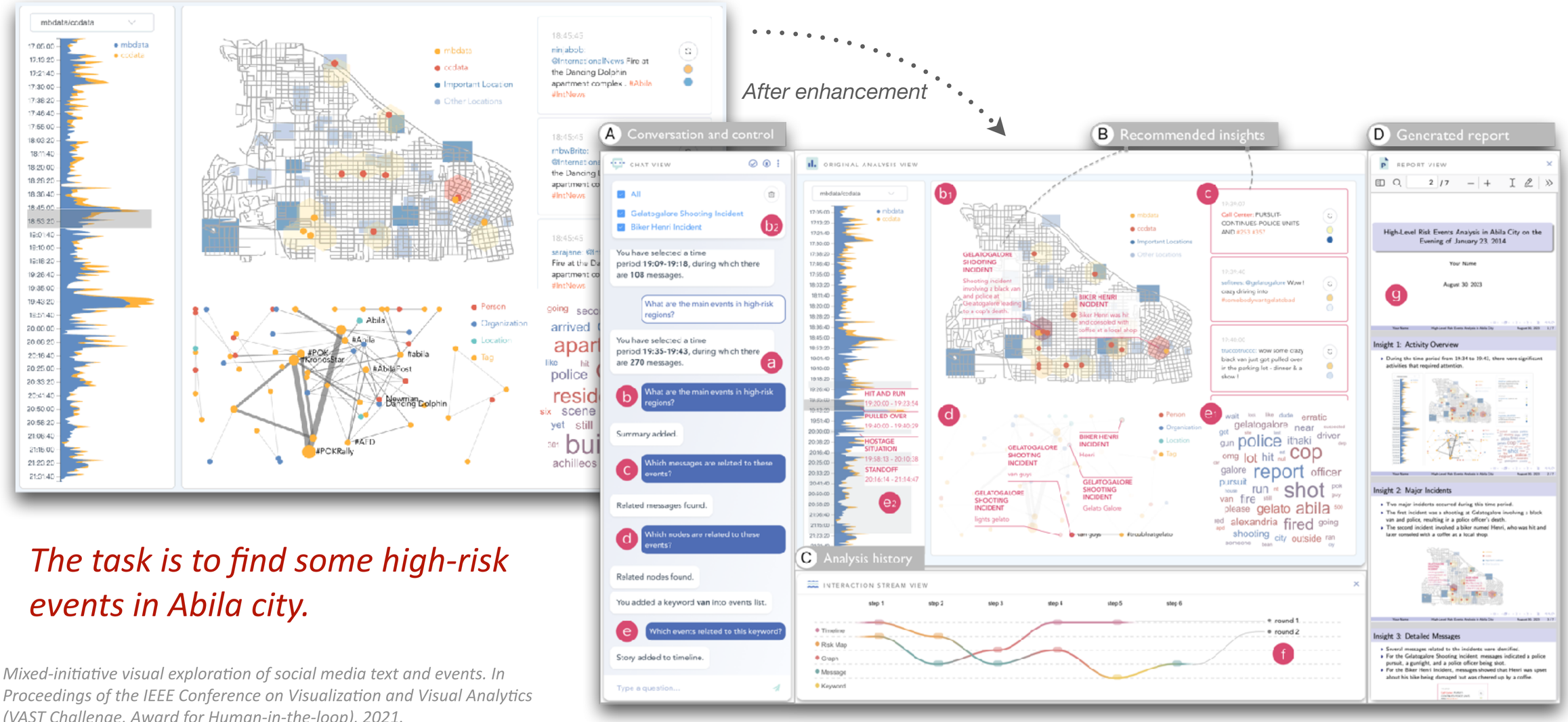
LEVA Pipeline

How we integrate LLMs into the VA pipeline?



An Example

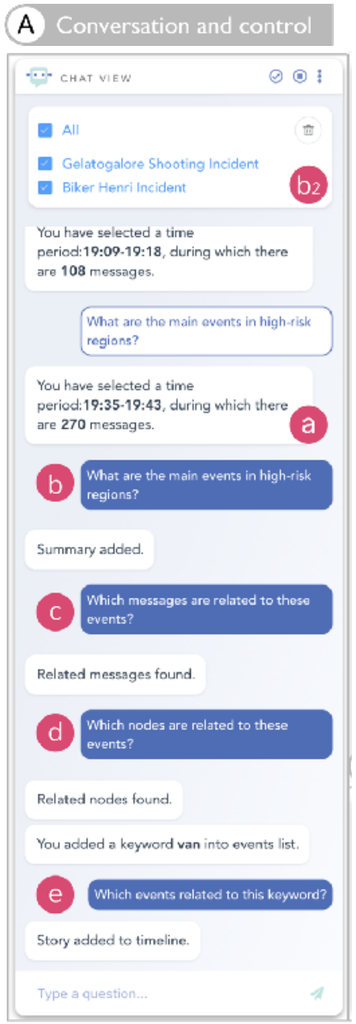
The system is from a winning entry of VAST Challenge 2021 · Mini-Challenge 3



The task is to find some high-risk events in Abila city.

Mixed-initiative visual exploration of social media text and events. In Proceedings of the IEEE Conference on Visualization and Visual Analytics (VAST Challenge, Award for Human-in-the-loop), 2021.

LLM-Enhanced VA System Interface



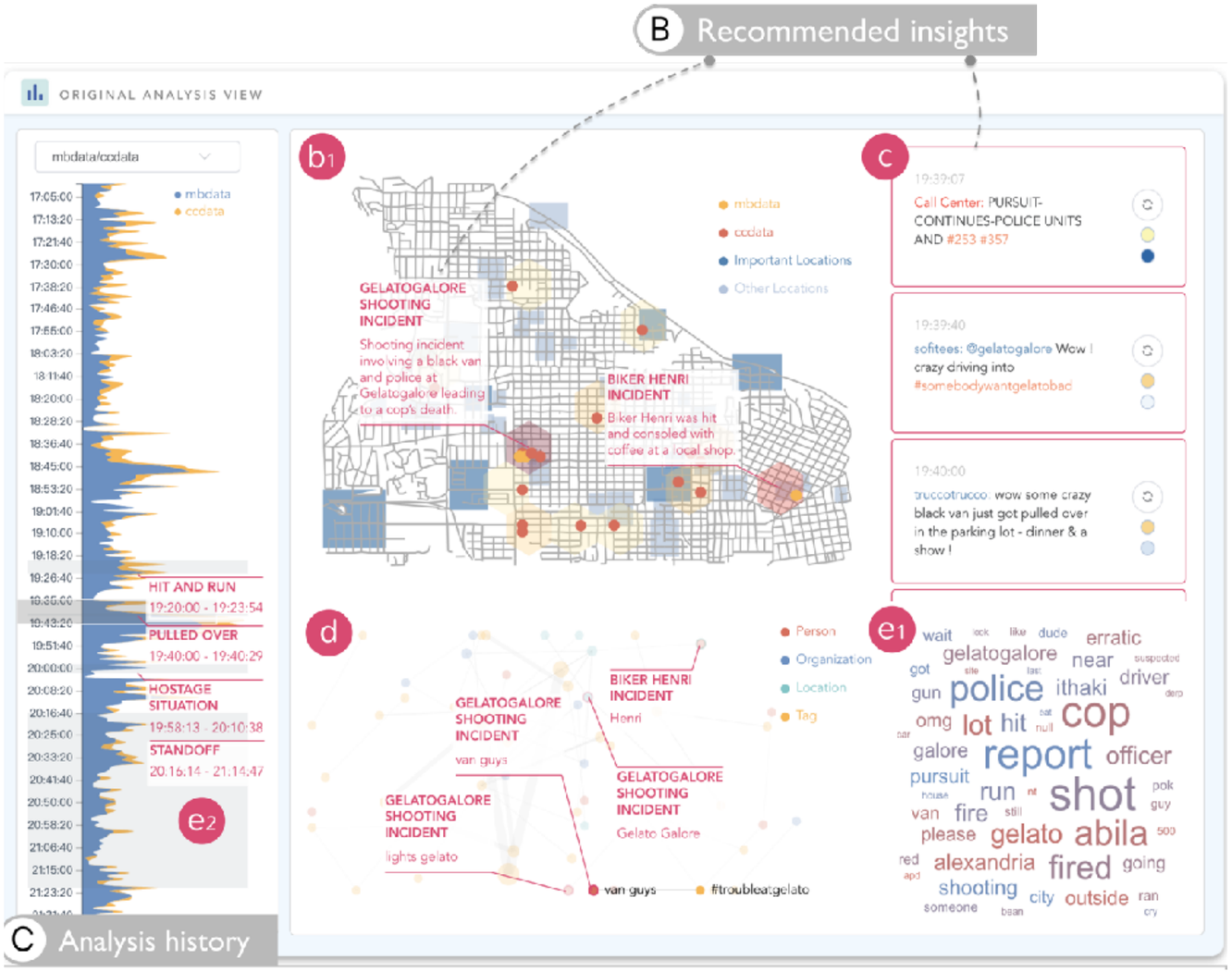
→ **Chat View**

receive user's questions and update LLMs' feedback

LLM-Enhanced VA System Interface

Original System View

Recommended insights ->Annotations

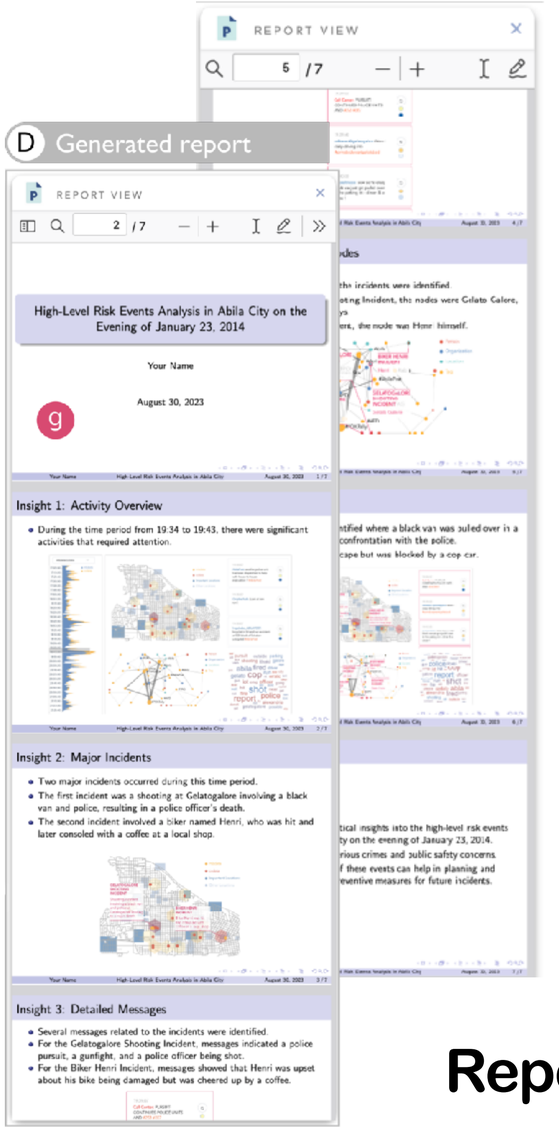


LLM-Enhanced VA System Interface

Interaction Stream View

Retrace the history and generate reporting

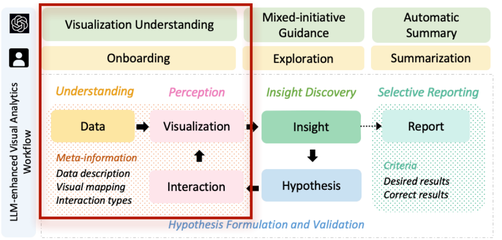
Click->



Report View

Methods

Onboarding tutorial generation



VA system
Propose a grammar of VA system

→

LLM
generate a HTML style tutorial

→

VA system
Add an interactive tutorial using Driver.js

- **Grammar: system info, view style info, view coordination info**

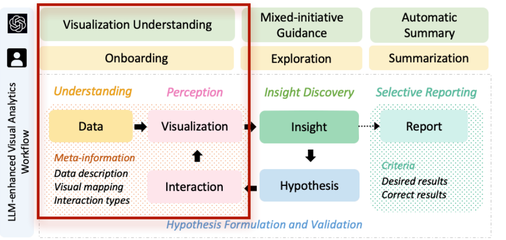
```
class SystemSpec {
  /** Information about the overall system */
  systemInfo: {
    systemName: string
    viewNumber: number
  }
  /** Style and coordination between views */
  viewsInfo: {
    viewStyleInfo: ViewStyleInfo[]
    viewCoordinationInfo: ViewCoordinationInfo[]
  }
}
```

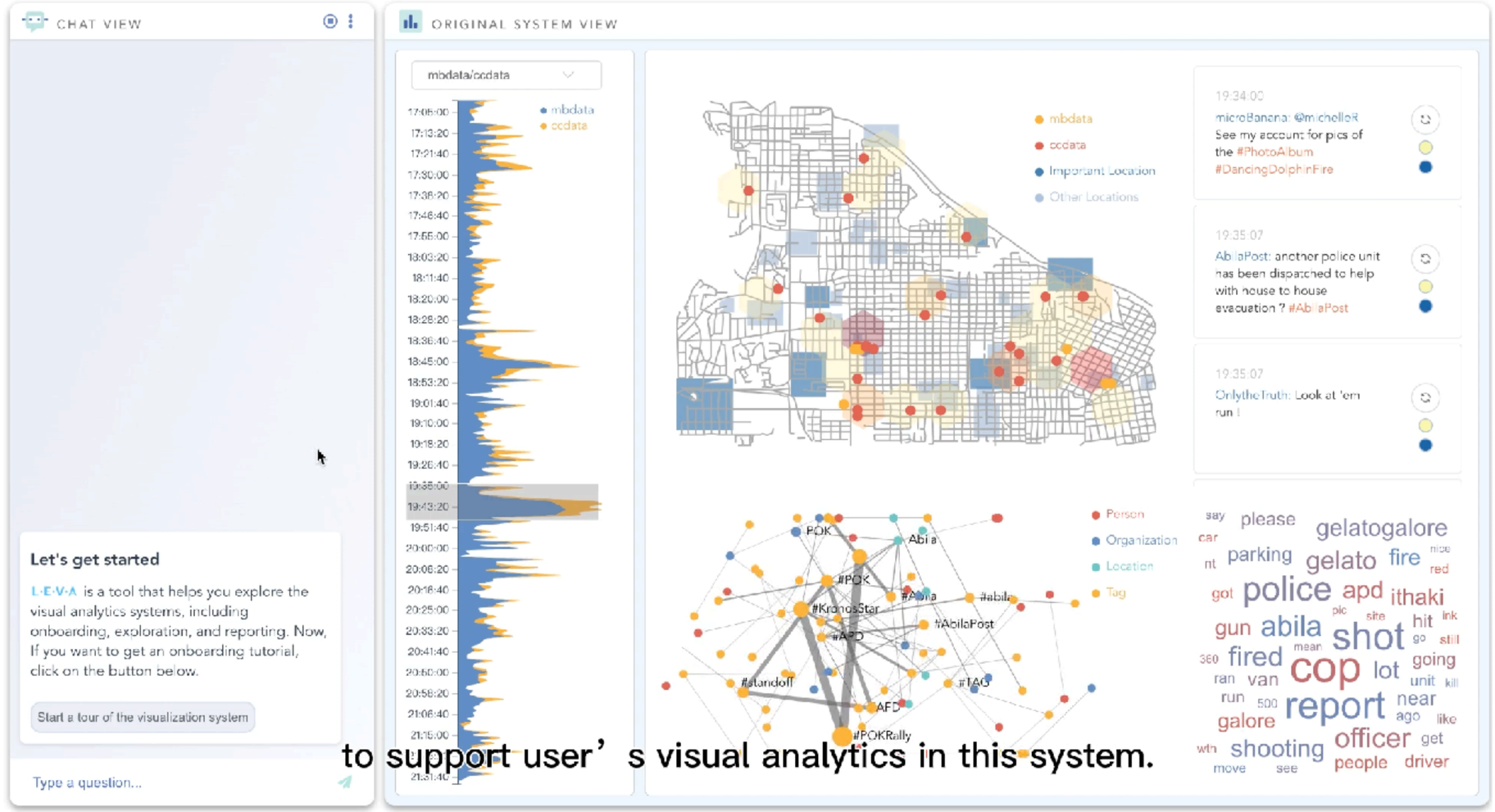
```
class EncodingInfo {
  field: string;
  type: string;
  description: string;
}
class ViewStyleInfo {
  viewName: string
  /** Information about styles of multiple encodings */
  layers: {
    markType: string
    encoding: {
      x?: EncodingInfo[];
      y?: EncodingInfo[];
      color?: EncodingInfo[];
      size?: EncodingInfo[];
      /** Detail level data field */
      lod?: EncodingInfo[];
    }
    /** Display prompt information for mouse hover.*/
    tooltip?: EncodingInfo[];
  }[]
}
```

```
class ViewCoordinationInfo {
  sourceViewName: string
  targetViewName: string | string[]
  /** Coordination type */
  coordinationType: string
  interaction: {
    /** Type of interaction */
    type: string
    /** Interaction's effect on target */
    effect?: {
      /** Action type */
      action: string
      targetViewName: string
      /** Data category for action */
      category: string
      /** Control the result */
      changeby: string
    }
  }[]
}
```

Methods

Onboarding tutorial generation: demo

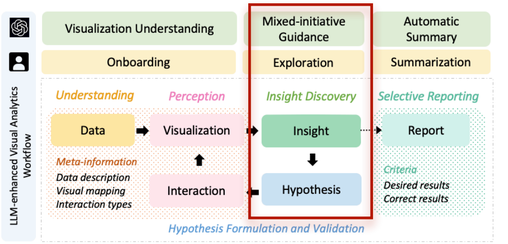


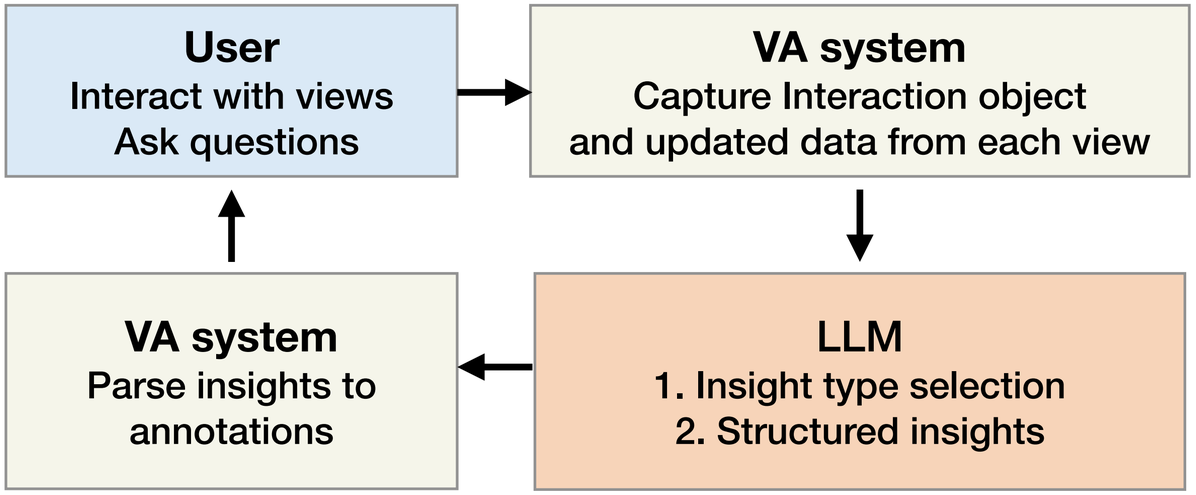


to support user's visual analytics in this system.

Methods

Insight recommendation (mixed-initiative process)

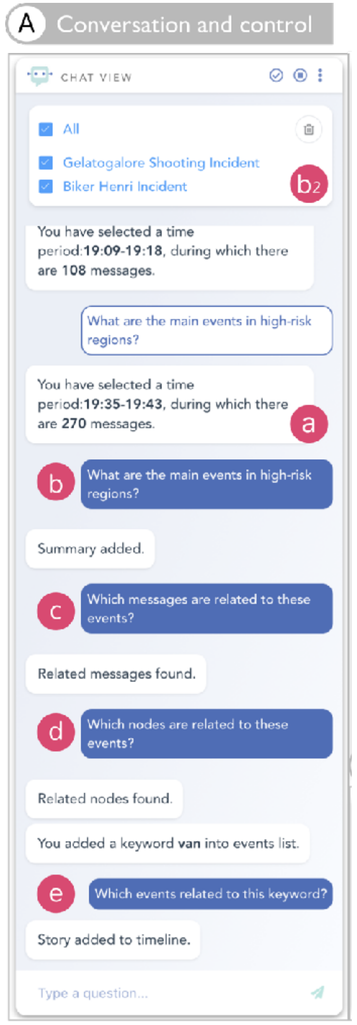




1. Insight type list

Insight type	Description
Outstanding No.1	The leading value is significantly higher than all the remaining values.
Outstanding Top 2	The leading two values are significantly higher than all the remaining values.
Outstanding Last	The value is remarkably lower than all the remaining values.
Attribution	The leading value dominates (accounting for $\geq 50\%$) the group.
Change Point	A specific point in time where there is a significant change or shift in the underlying data-generating process.
Outlier	An observation or data point that significantly deviates from the rest of the data.
Seasonality	A regular and predictable pattern of fluctuations or variations that occur at specific intervals of time.
Trend	A time series has an obvious trend (increase or decrease) with a certain turbulence level (steadily/ with turbulence).
Correlation	The statistical relationships between random variables, multidimensional data or time series.
Difference	The similarity or difference between two or more datasets
Aggregation	The descriptive statistical indicators (e.g., average, sum, count, etc.) based on the data attributes.
Value	The the exact value of data attribute(s) under specific criteria.
Text summary	The core ideas of a text dataset. The summary might have spatial or temporal features.
Important nodes or links	The important nodes or links in a graph under specific criteria.
Important text or keywords	The important original texts or keywords under specific criteria.

Propose questions

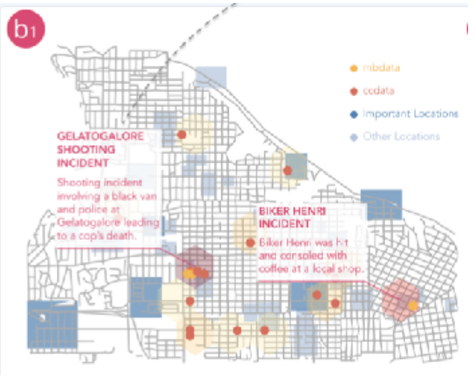


2. Calculate insights

Tool using

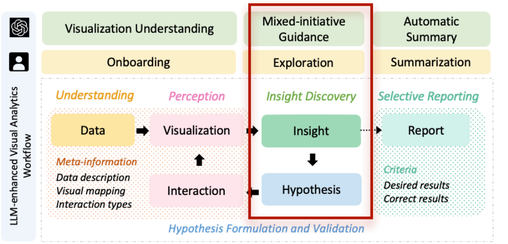
```
"functions": [
  {
    "name": "get_change_point",
    "description": "Get the change point in a time series
        dataset"
  },{
    "name": "get_seasonality",
    "description": "Get the seasonality in a time series
        dataset"
  },{
    "name": "get_trend",
    "description": "Get the trend in a time series dataset
        "
  },{
    "name": "get_outlier",
    "description": "Get outliers in a dataset",
  },{
    "name": "get_correlation",
    "description": "Get the correlation of two time series
        datasets"
```

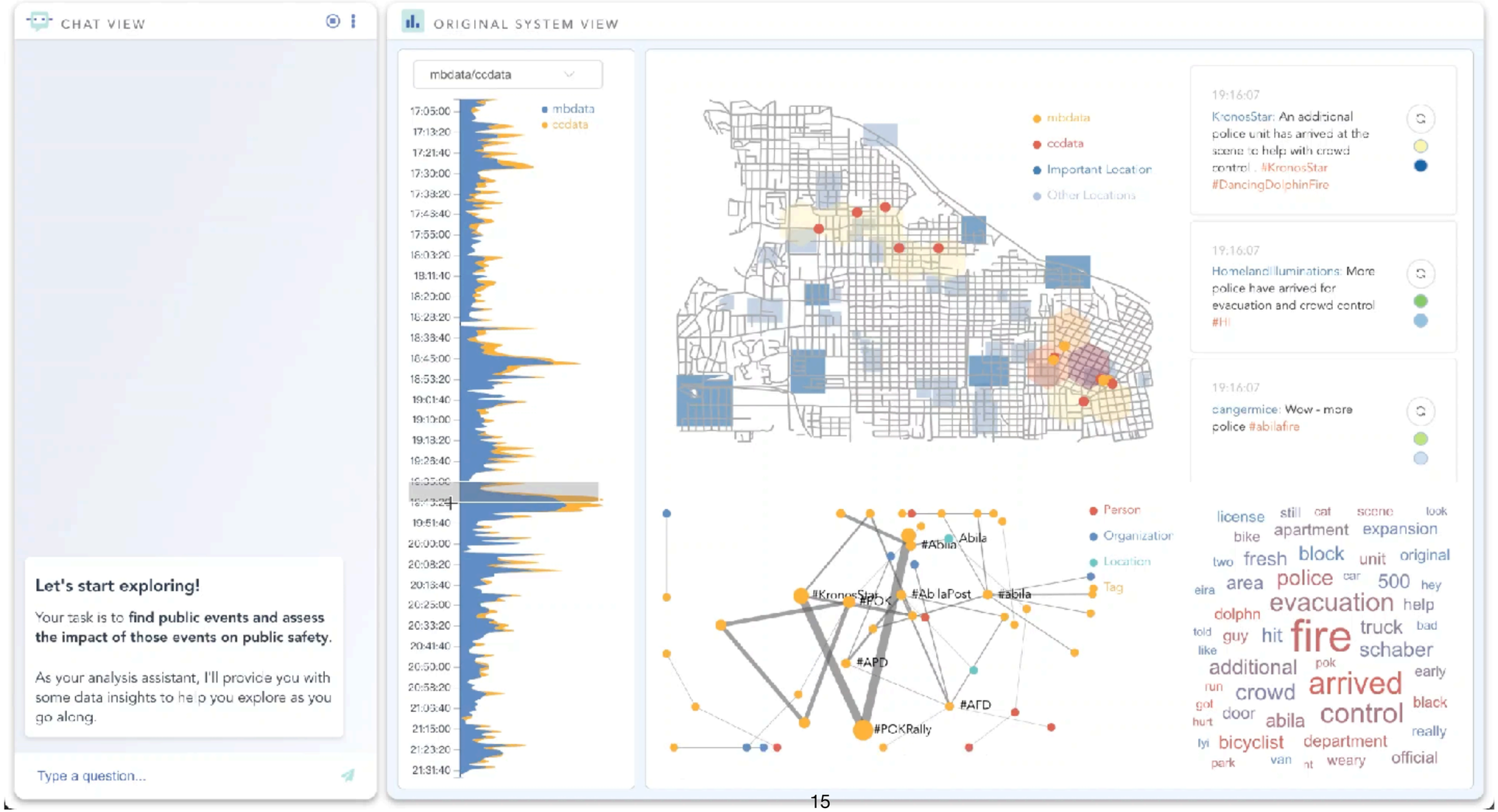
Return structured insights



Methods

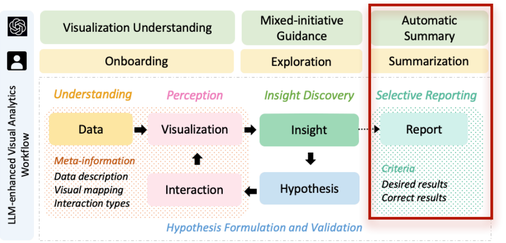
Insight recommendation: demo





Methods

Report generation



VA system
Log interaction and insights

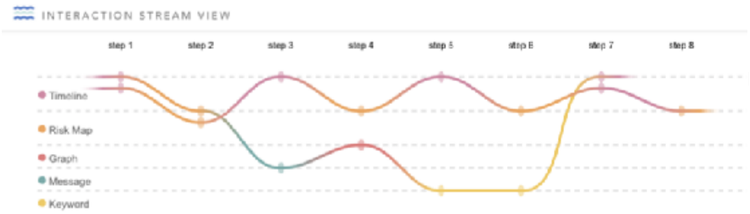
→

User
Select desired logs

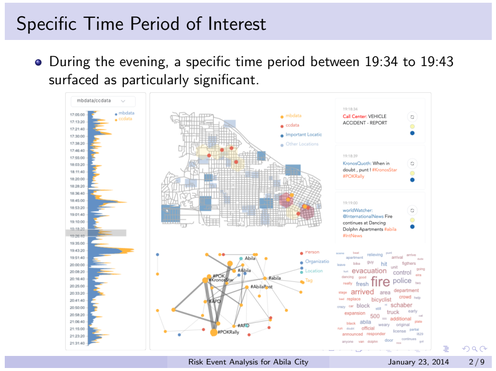
→

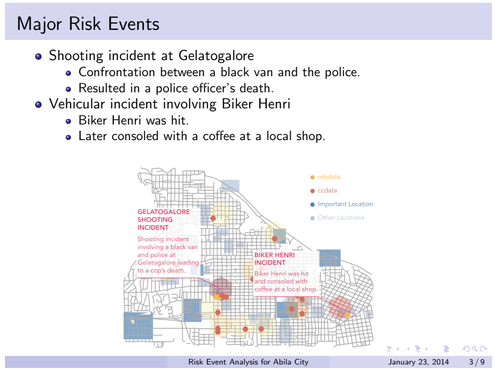
LLM
Summarize and generate LaTeX slides

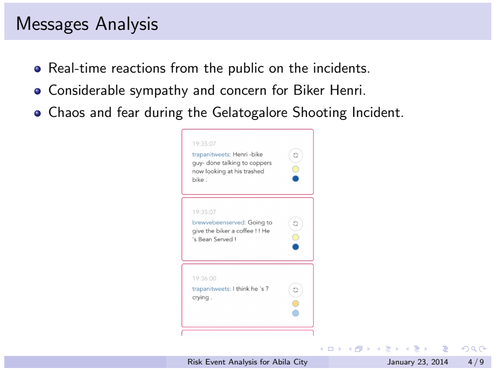
Visualize analysis history in a stream graph

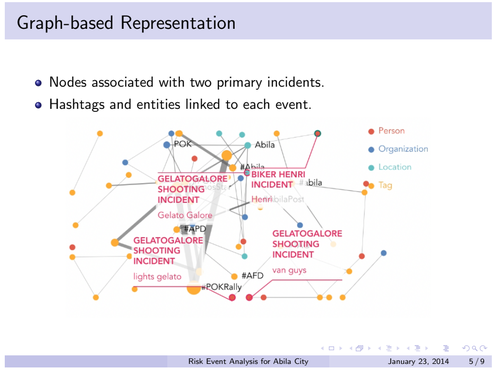


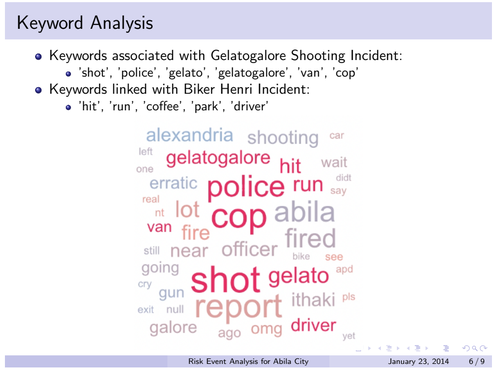
→

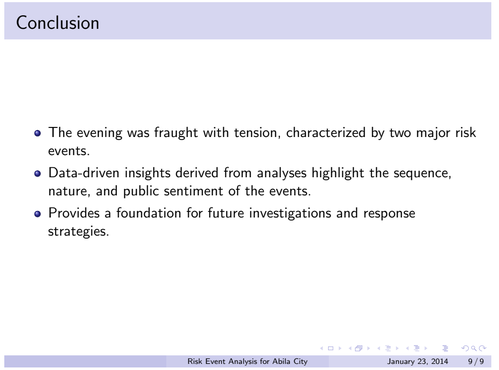












How to implement LEVA into your own system?

The implementation requirement of our framework includes two parts:

1. Extensions of the original VA system

➡ *configuration files and data handlers to process user selections and LLM outputs*

2. LLM-powered components

➡ *prompt handlers and presentation modules for showcasing outputs across the original system view and three new views (chat view, report view, stream view)*

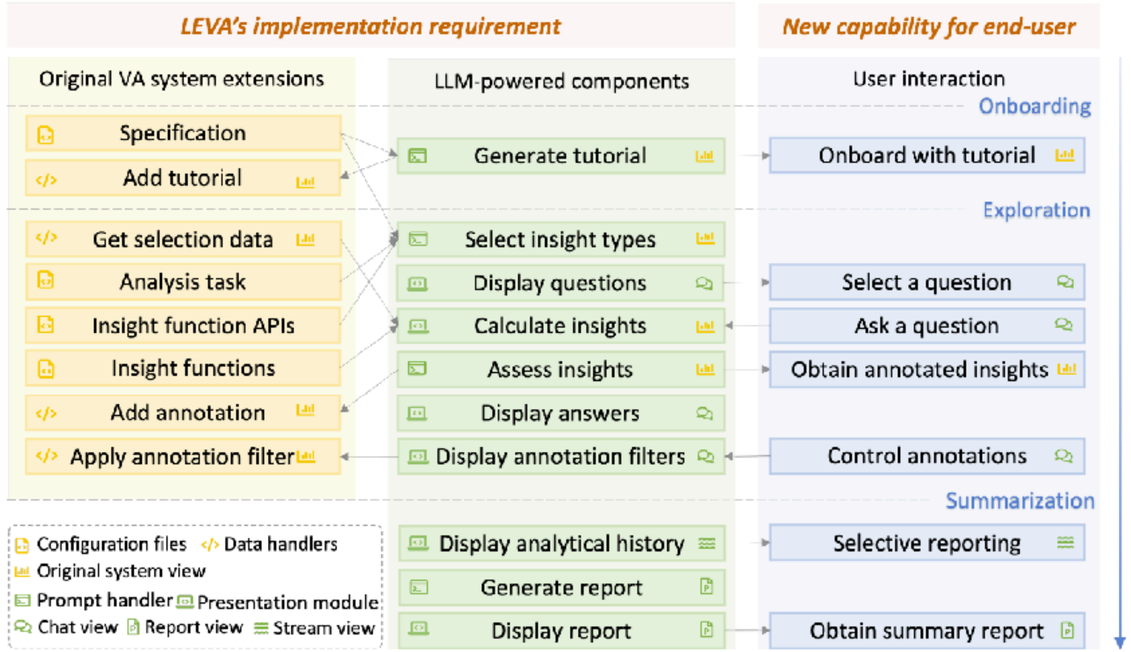
The new capability for end-users after enhancement:

3. User interactions on the enhanced interface

➡ *Onboarding <- interactively learn VA system*

Exploring <- Ask questions, be perceived, gain insight

Summarizing <- gain automatically generated reports



Future work

Extend and generate the grammar of VA systems to facilitate LLM's understanding

- *A more unified grammar can be designed to describes modules of data, visualization, interaction and models of a VA system.*

Bring domain knowledge for LLMs to solve specific tasks

- *Knowledge integration by tool learning, RAG and Fine-tuning, to bridge the gap between VA and end-users.*

Consider the relations between LLM-agents and Human

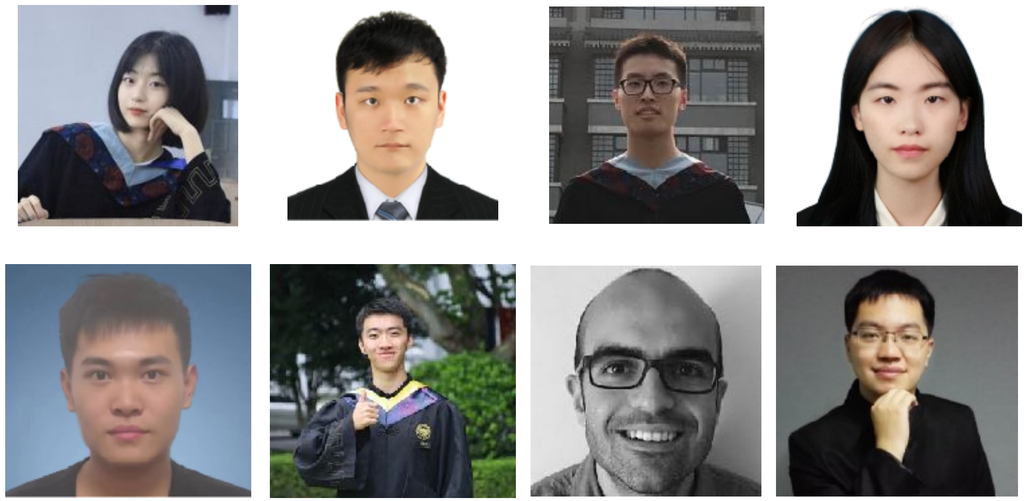
- *Support **mixed-initiative**, not only human-guide or agent-guide*
- *The **interaction modality** could be diverse.*

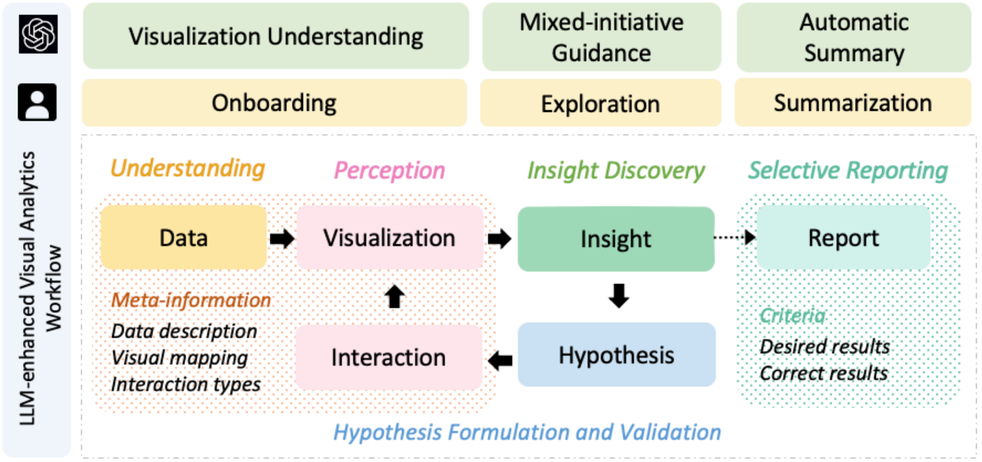
Mitigate concerns when using LLMs

- *The transparency, explainability, understandability, provenance etc*

Conclusion

Acknowledgement





LEVA: Using Large Language Models to Enhance Visual Analytics

Enhance visual analytics in three stages: onboarding, exploration and summarization.