

# Using Large Language Models to Enhance Visual Analytics

L·E·V·A

Yuheng Zhao<sup>1</sup>, Yixing Zhang<sup>1</sup>, Yu Zhang<sup>2</sup>, Xinyi Zhao<sup>1</sup>, Junjie Wang<sup>1</sup>,  
Zekai Shao<sup>1</sup>, Cagatay Turkay<sup>3</sup>, Siming Chen<sup>1</sup>



<sup>1</sup> School of Data Science, Fudan University

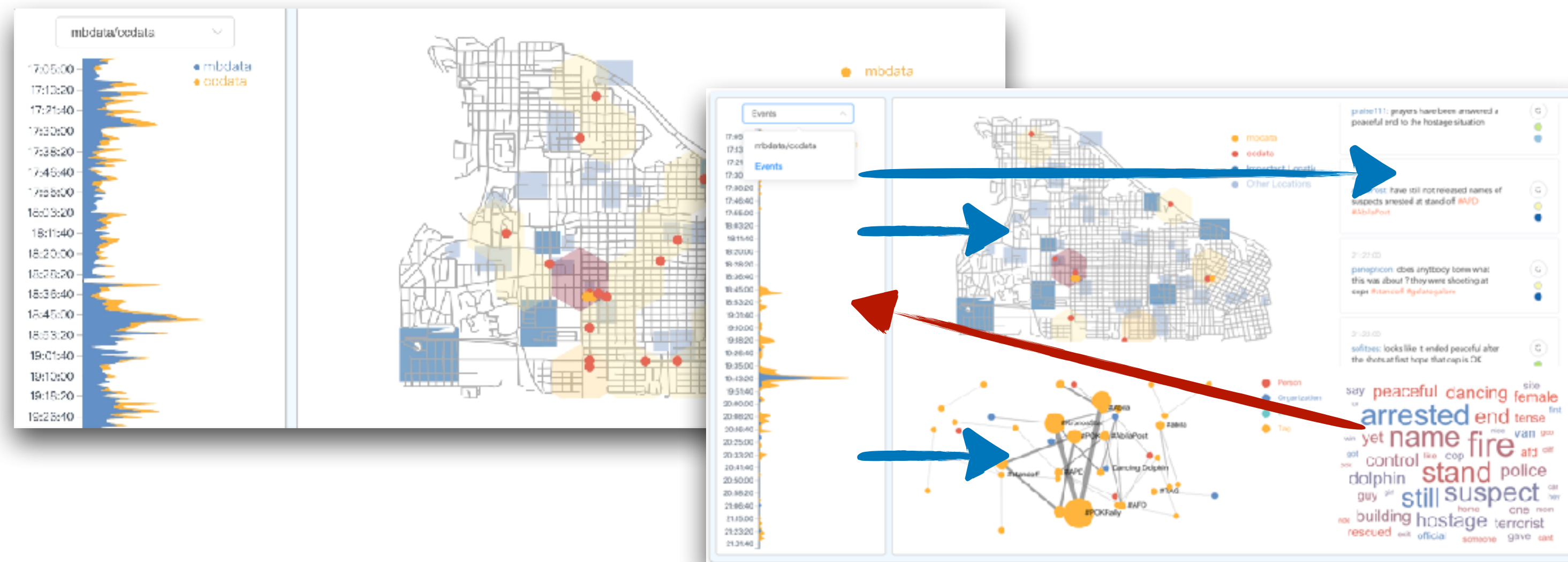
<sup>2</sup> Department of Computer Science, University of Oxford

<sup>3</sup> Centre for Interdisciplinary Methodologies, University of Warwick

*IEEE Transactions on Visualization and Computer Graphics, 2024*

# Motivation

👤 Are there any challenges when conducting visual analytics?



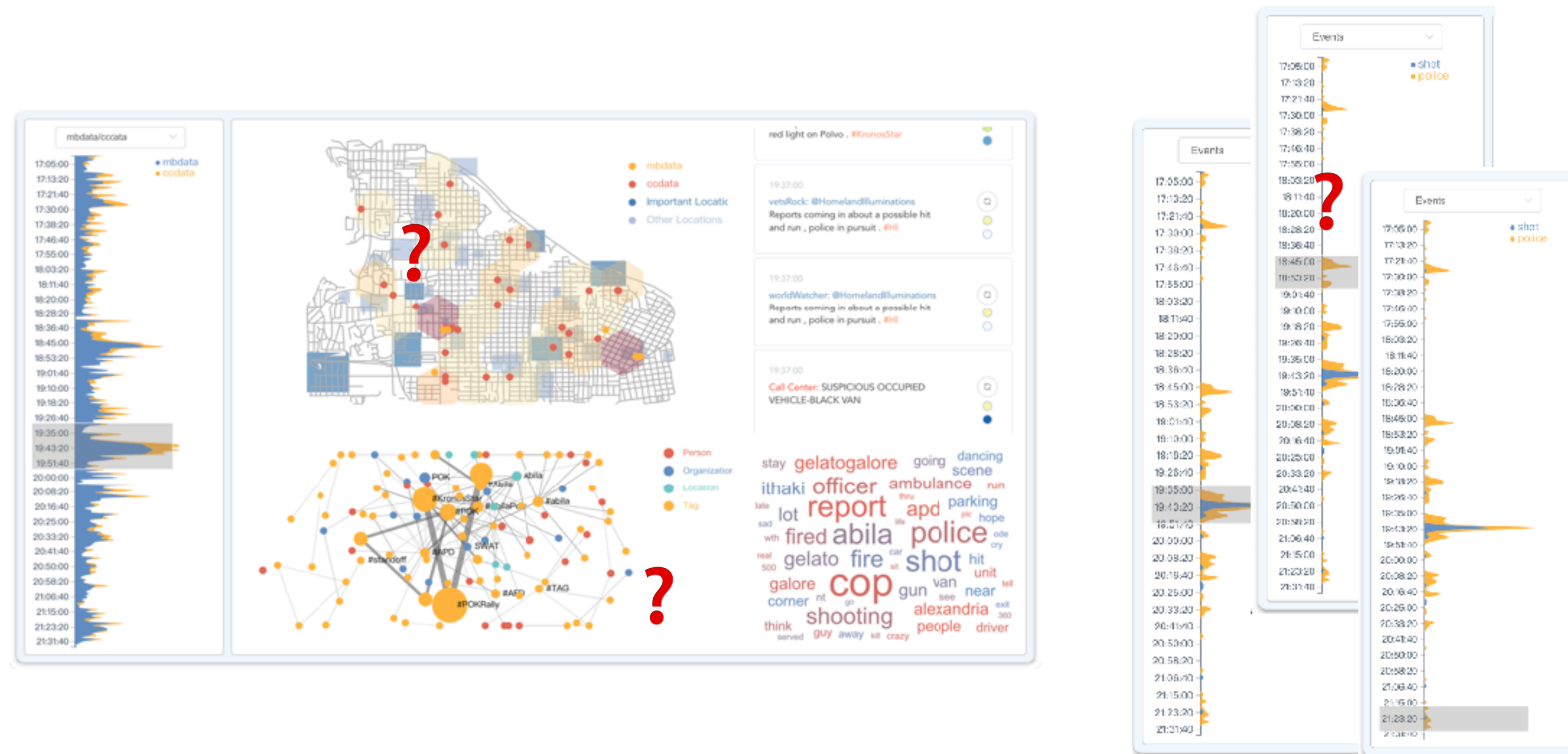
An example system from VAST Challenge 2021 MC3

## 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:

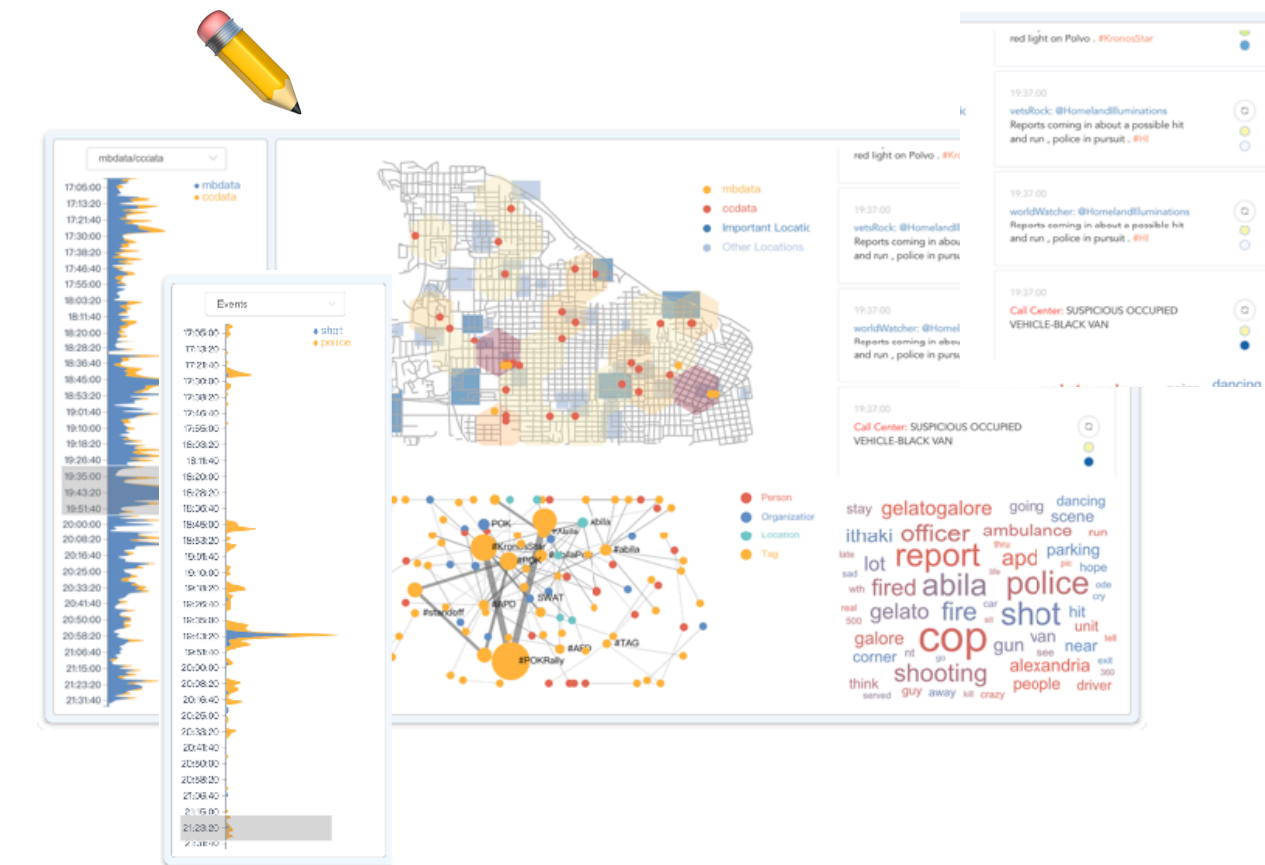
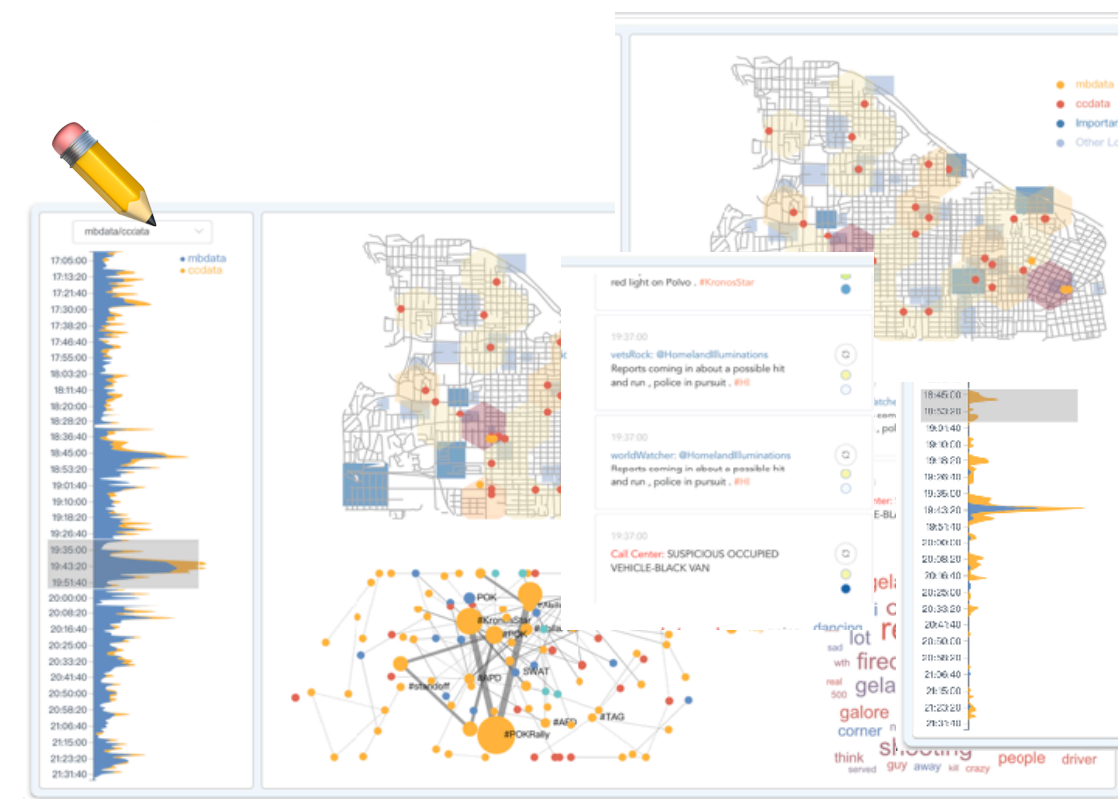
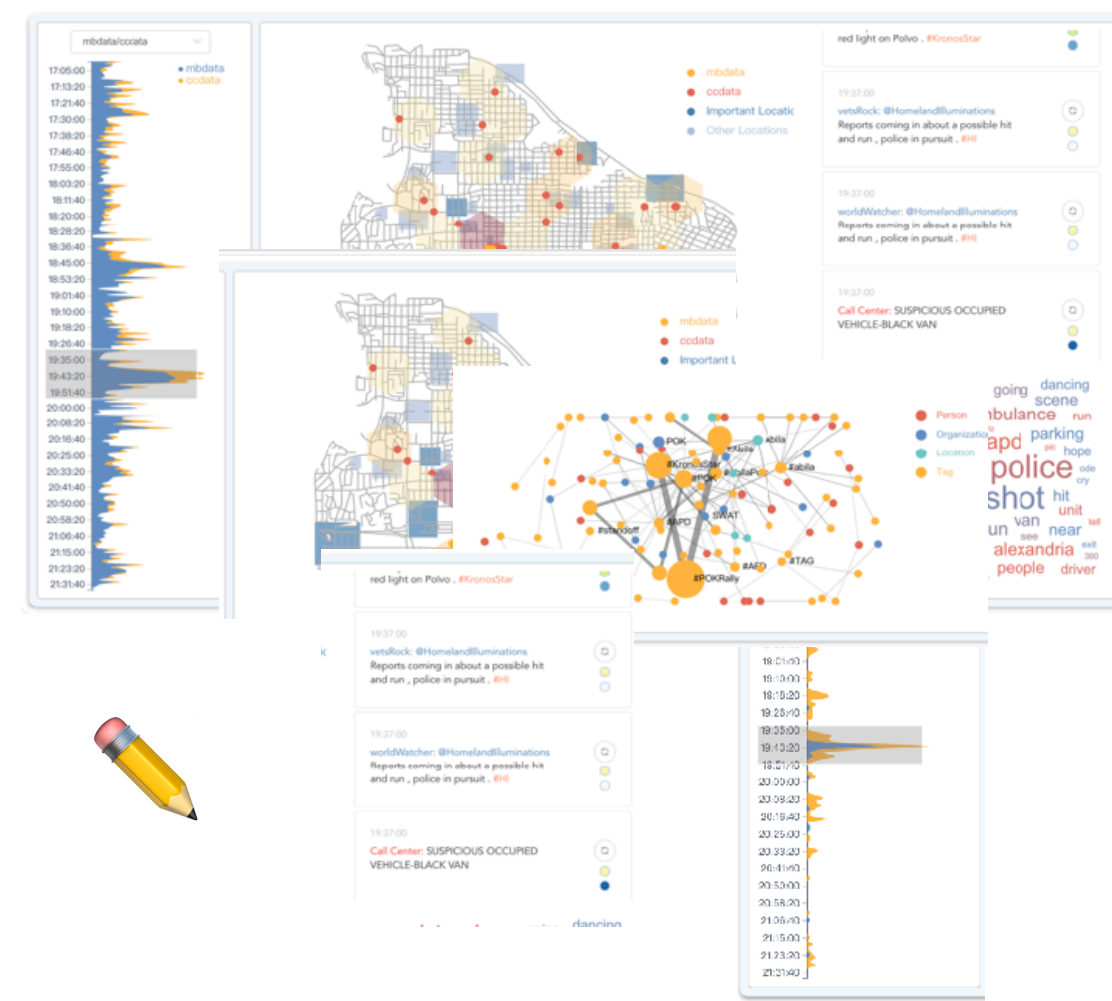
A lot of information needs to be observed and processed, resulting in low exploration efficiency



# Motivation



Are there any challenges when conducting visual analytics?

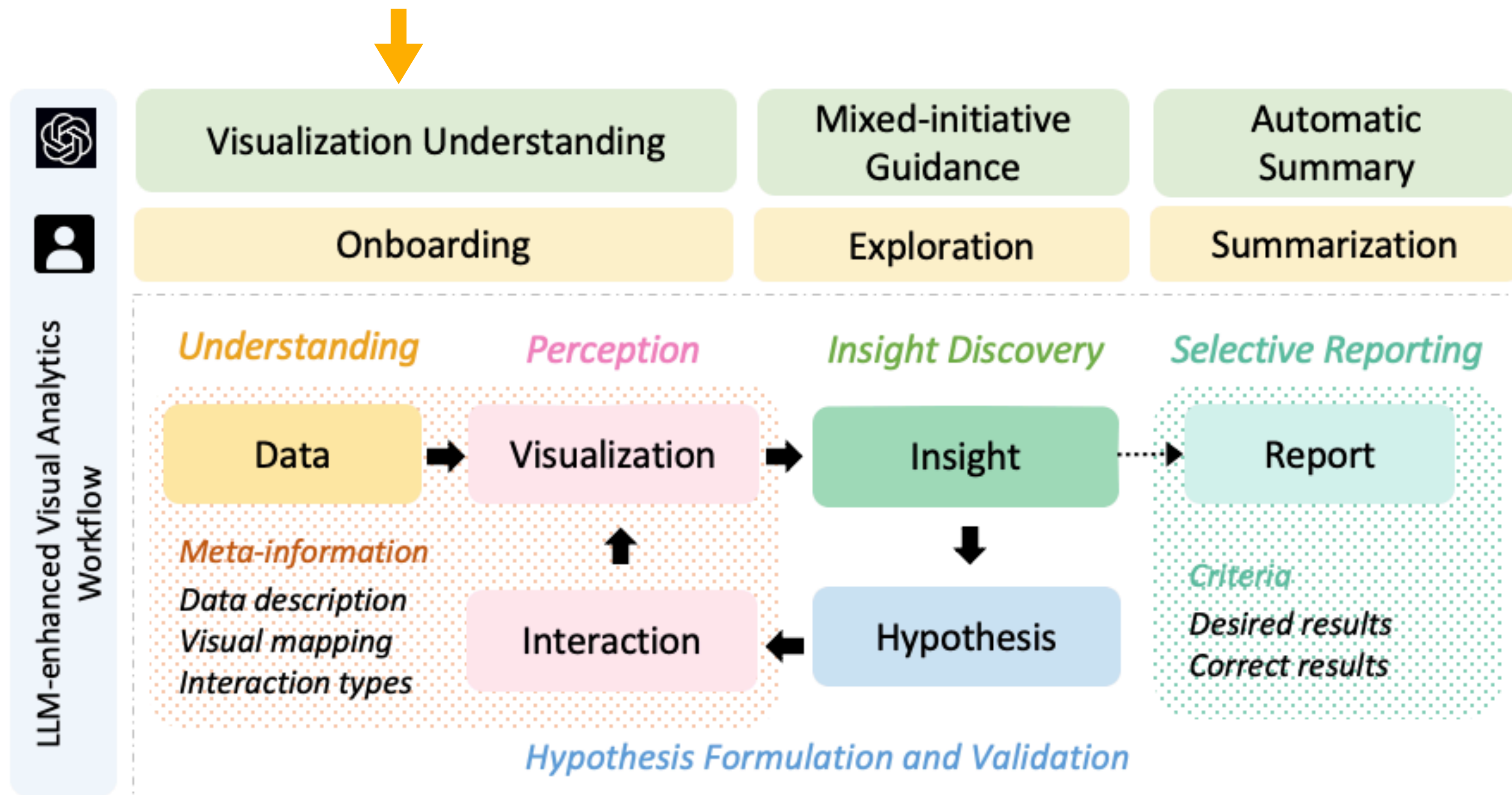


### 3. Summary stage:

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

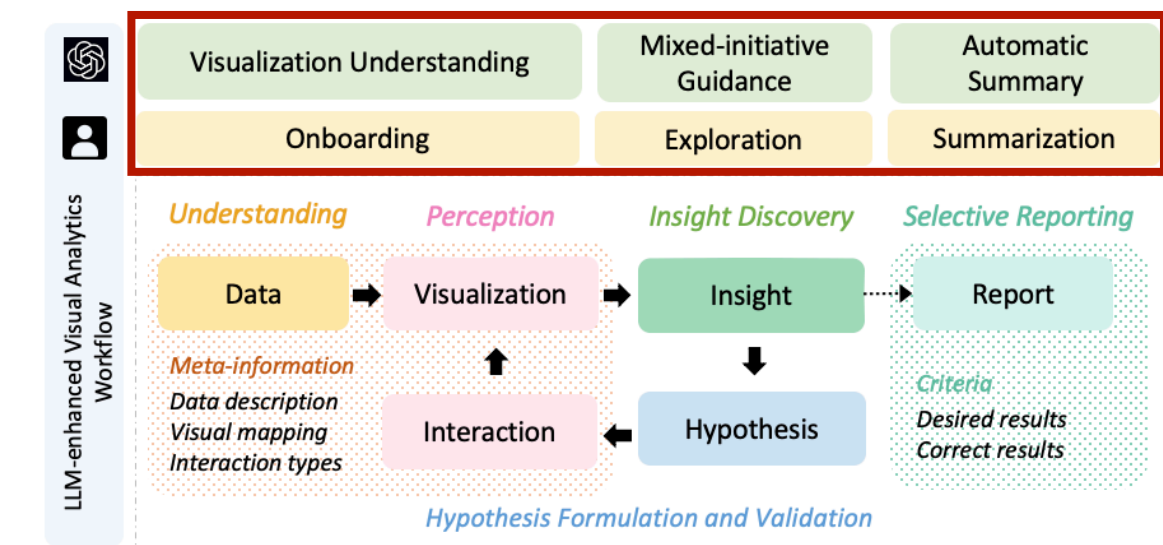
# Motivation

Where to enhance User's VA workflow?

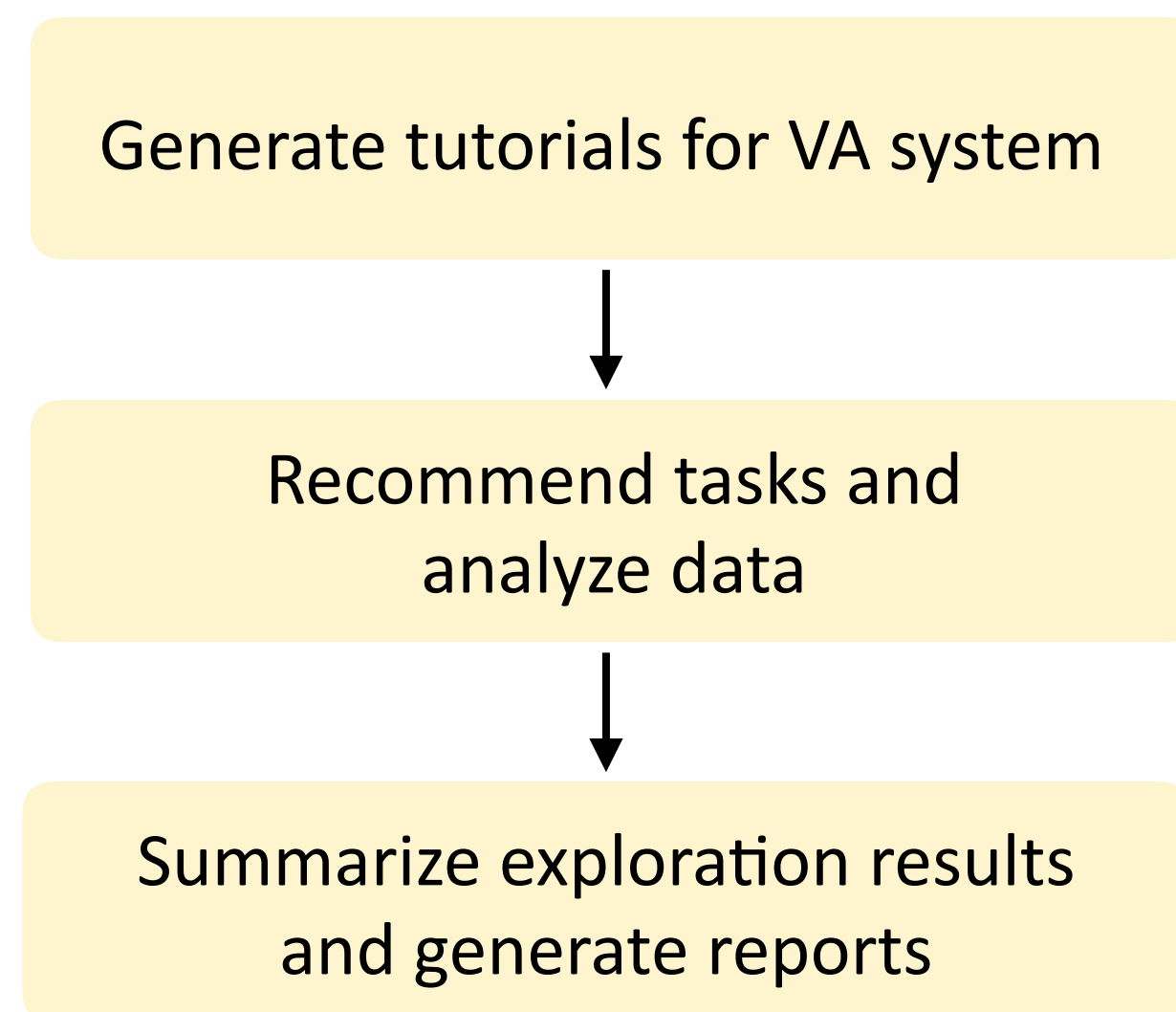


# Motivation

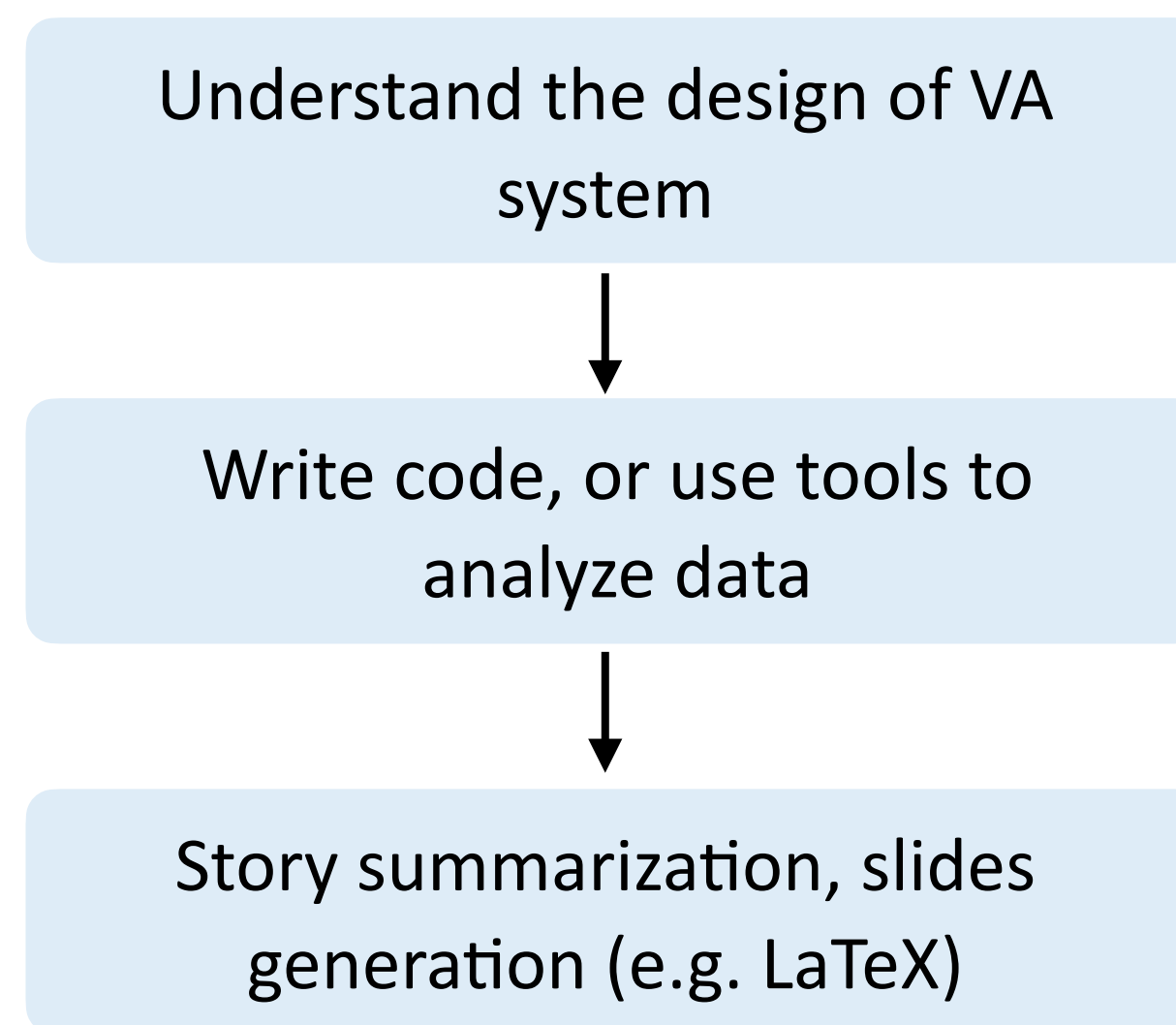
Why integrate LLMs into the VA pipeline?



Intelligent VA requires the assistant:



Capabilities the model should have:



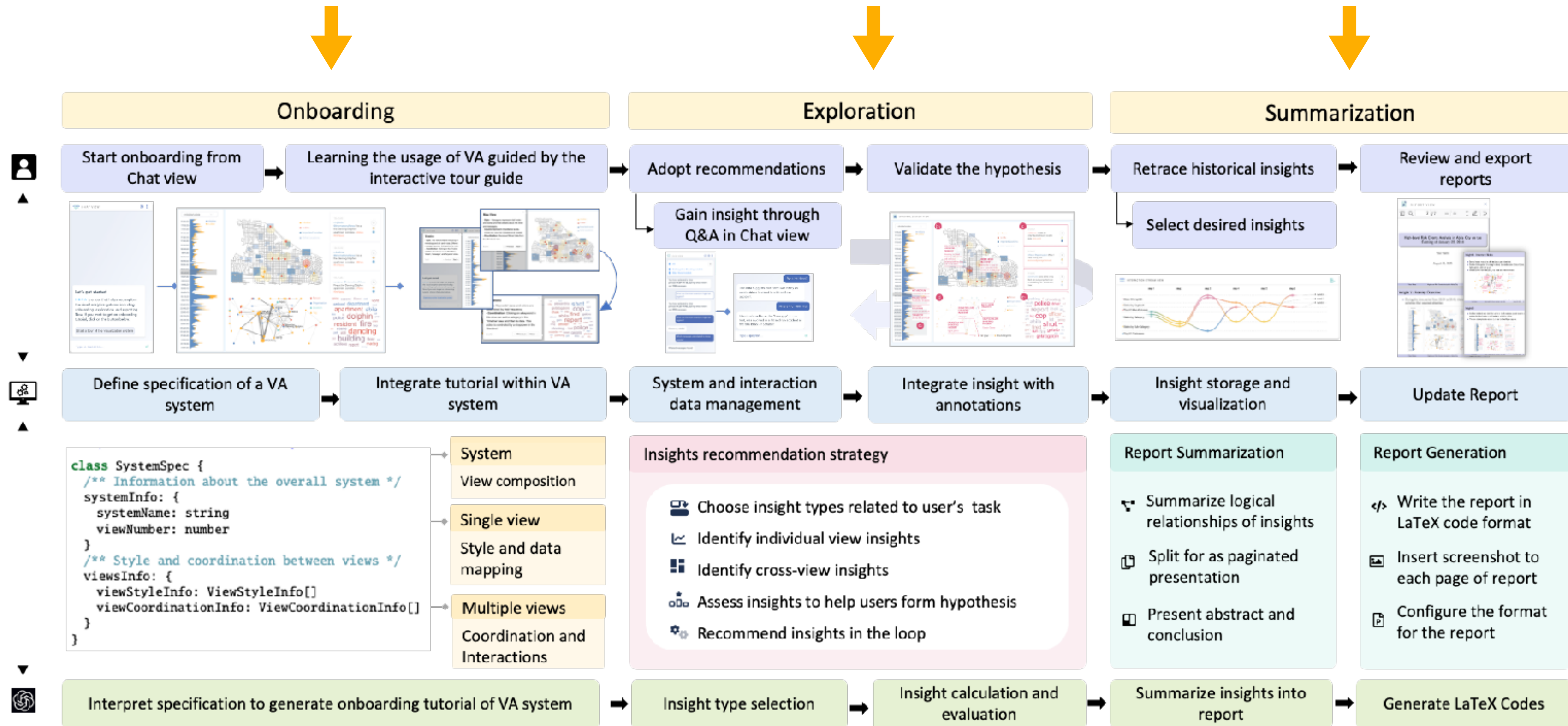
💡 **Why use LLM?**

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?





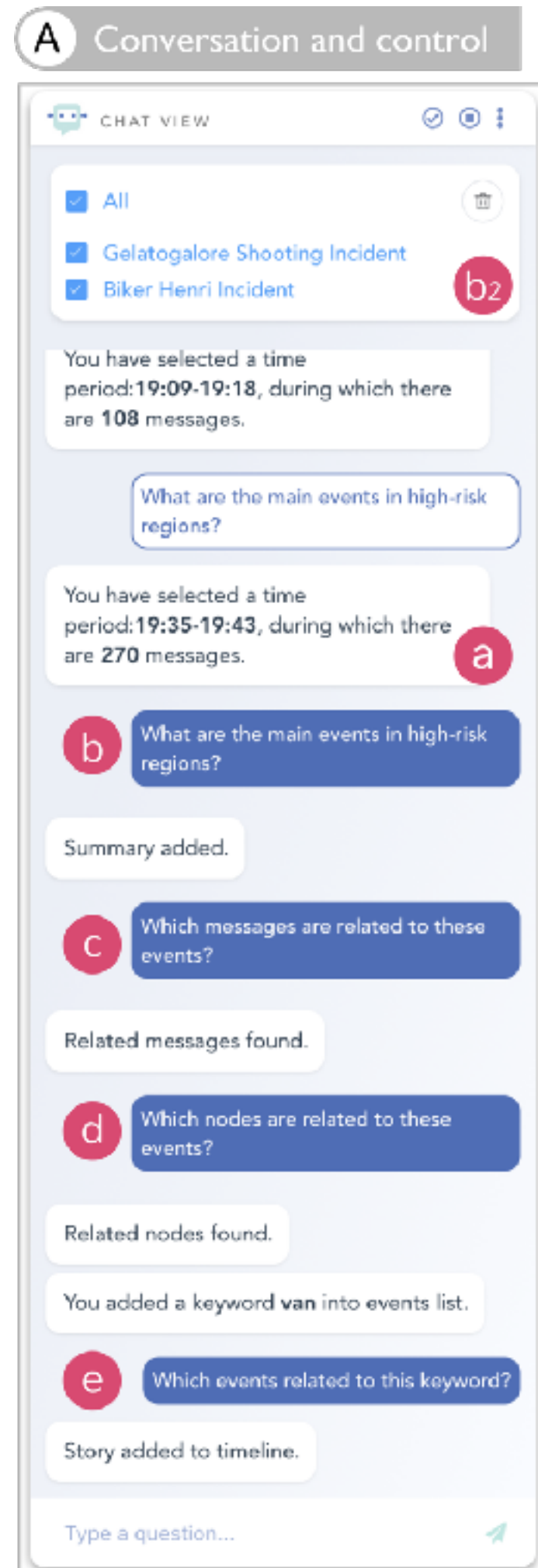
# An Example

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





# LLM-Enhanced VA System Interface



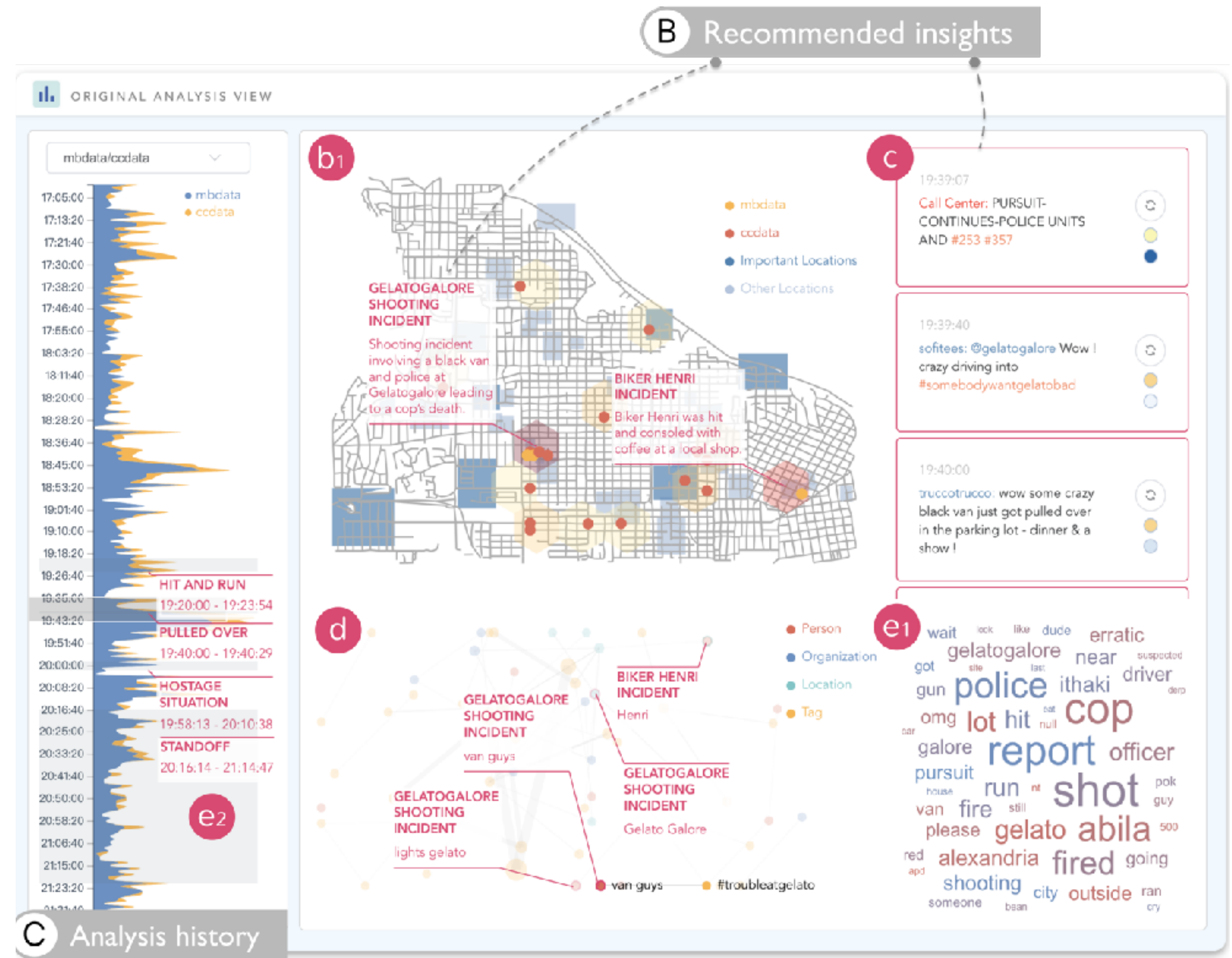
—————→ **Chat View**

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

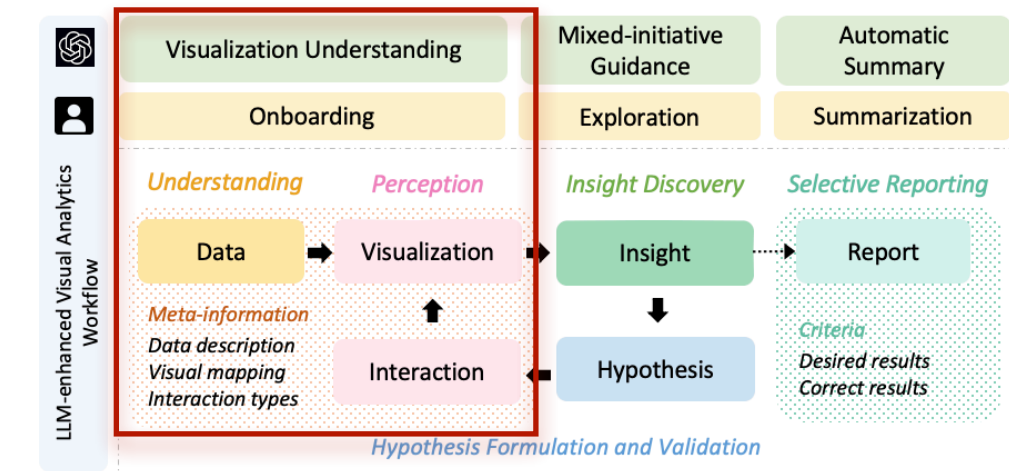
Retrace the history and generate reporting



Report View

# Methods

## Onboarding tutorial generation



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

```

1 class SystemSpec {
2   /** Information about the overall system */
3   systemInfo: {
4     systemName: string
5     viewNumber: number
6   }
7   /** Style and coordination between views */
8   viewsInfo: {
9     viewStyleInfo: ViewStyleInfo[]
10    viewCoordinationInfo: ViewCoordinationInfo[]
11  }
12 }
  
```

```

1 class EncodingInfo {
2   field: string;
3   type: string;
4   description: string;
5 }
6 class ViewStyleInfo {
7   viewName: string
8   /** Information about styles of multiple encodings */
9   layers: {
10    markType: string
11    encoding: {
12      x?: EncodingInfo[];
13      y?: EncodingInfo[];
14      color?: EncodingInfo[];
15      size?: EncodingInfo[];
16      /** Detail level data field */
17      lod?: EncodingInfo[];
18    }
19    /** Display prompt information for mouse hover.*/
20    tooltip?: EncodingInfo[];
21  }[]
22 }
  
```

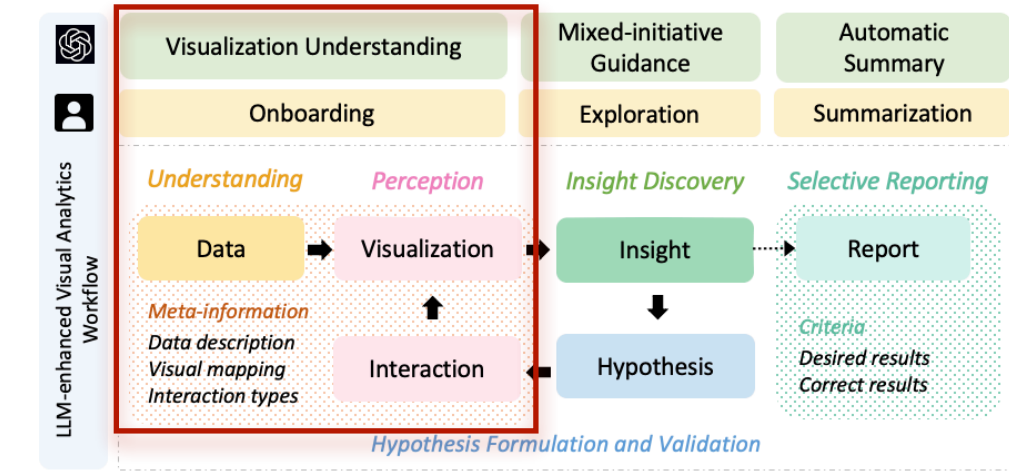
```

1 class ViewCoordinationInfo {
2   sourceViewName: string
3   targetViewName: string | string[]
4   /** Coordination type */
5   coordinationType: string
6   interaction: {
7     /** Type of interaction */
8     type: string
9     /** Interaction's effect on target */
10    effect?: {
11      /** Action type */
12      action: string
13      targetViewName: string
14      /** Data category for action */
15      category: string
16      /** Control the result */
17      changeby: string
18    }
19  }[]
20 }
  
```



# Methods

Onboarding tutorial generation: demo



CHAT VIEW

ORIGINAL SYSTEM VIEW

mbdata/ccdata

17:05:00, 17:13:20, 17:21:40, 17:30:00, 17:38:20, 17:46:40, 17:55:00, 18:03:20, 18:11:40, 18:20:00, 18:28:20, 18:36:40, 18:45:00, 18:53:20, 19:01:40, 19:10:00, 19:18:20, 19:26:40, 19:35:00, 19:43:20, 19:51:40, 20:00:00, 20:08:20, 20:16:40, 20:25:00, 20:33:20, 20:41:40, 20:50:00, 20:58:20, 21:06:40, 21:15:00, 21:23:20, 21:31:40

mbdata, ccdata

mbdata, ccdata, Important Location, Other Locations

Person, Organization, Location, Tag

19:34:00, 19:35:07, 19:35:07

microBanana: @michelleR See my account for pics of the #PhotoAlbum #DancingDolphinFire

AbilaPost: another police unit has been dispatched to help with house to house evacuation ? #AbilaPost

OnlytheTruth: Look at 'em run !

say please gelatogalore car parking gelato fire nice red got police apd ithaki gun abila shot hit link fired mean shot go still 360 ran van cop lot unit kill run 500 report near galore report ago like with shooting officer get move see people driver

Let's get started

L·E·V·A is a tool that helps you explore the visual analytics systems, including onboarding, exploration, and reporting. Now, If you want to get an onboarding tutorial, click on the button below.

Start a tour of the visualization system

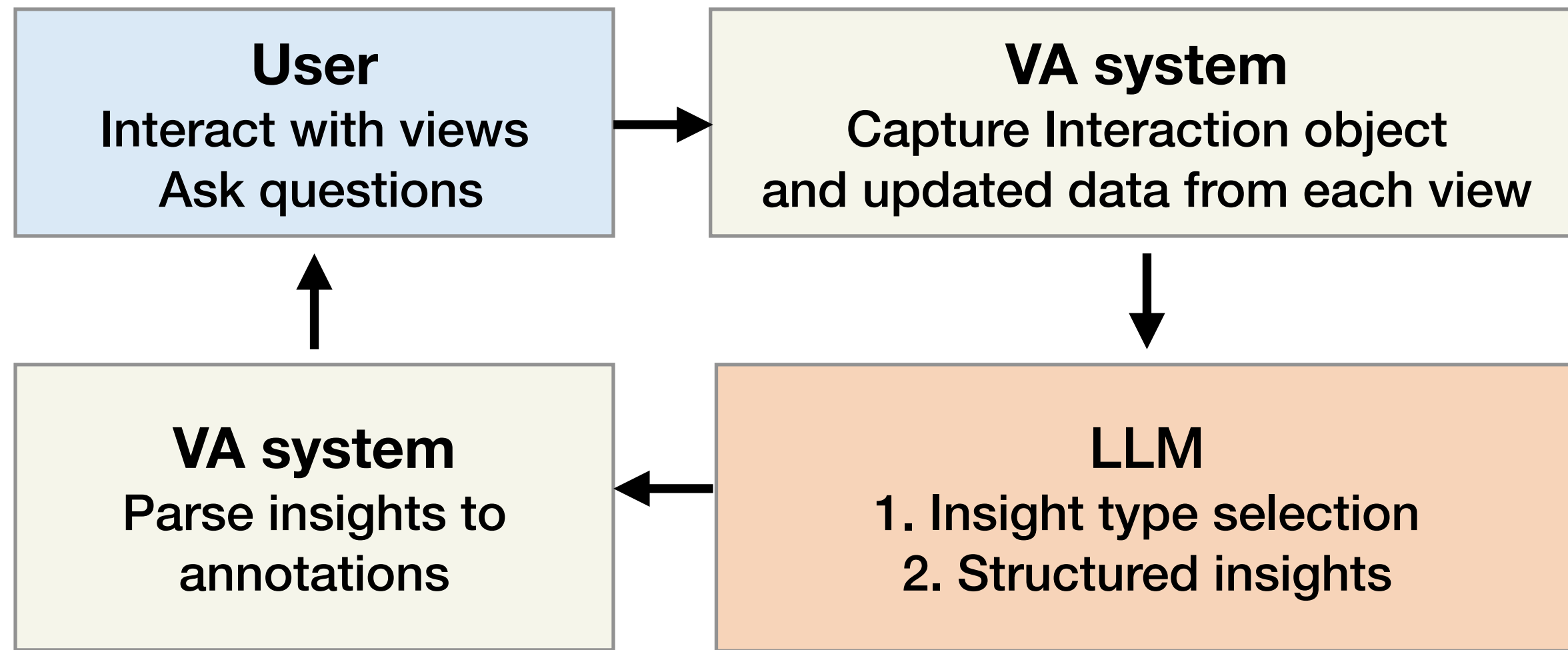
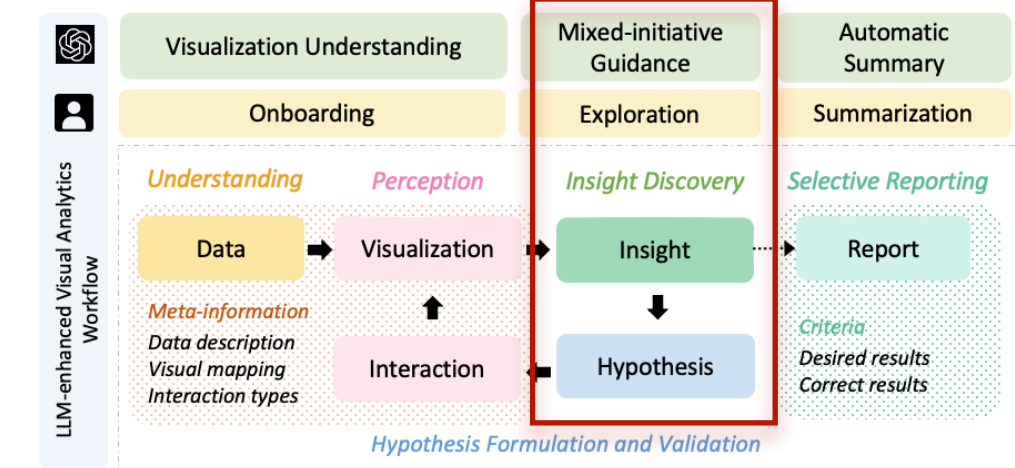
Type a question...

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

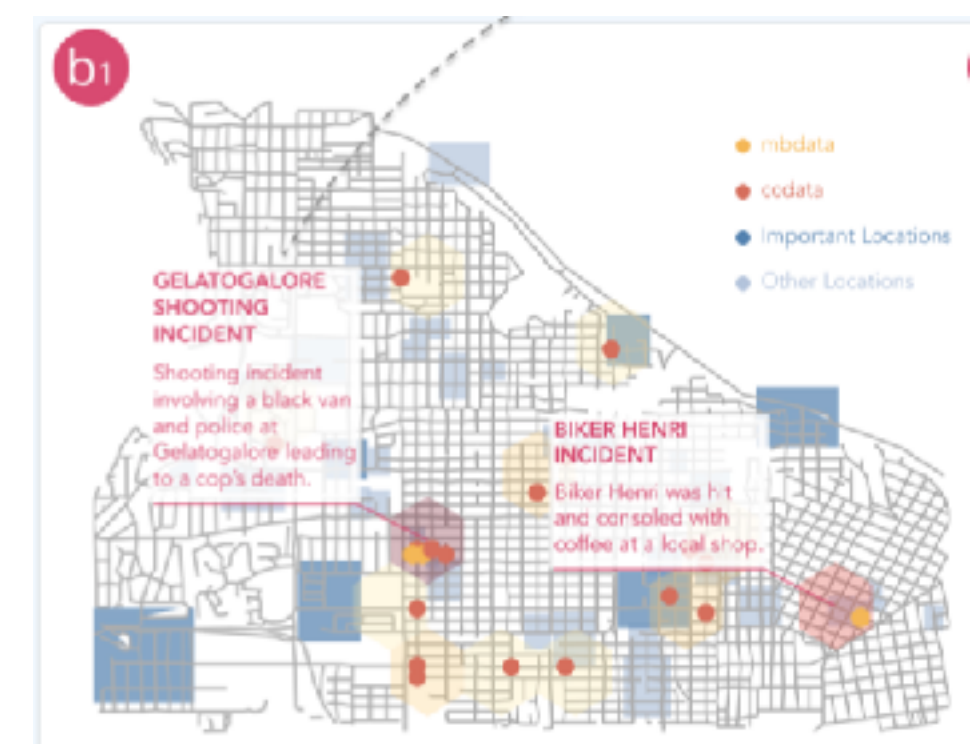
## 2. Calculate insights

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"
  }
]
    
```

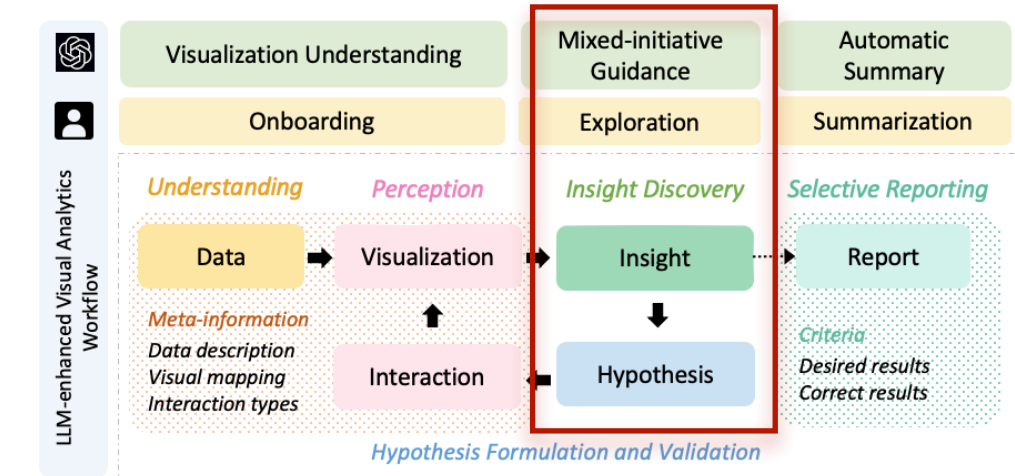
Return structured insights





# Methods

Insight recommendation: demo



**CHAT VIEW**

Let's start exploring!  
Your task is to **find public events and assess the impact of those events on public safety.**

As your analysis assistant, I'll provide you with some data insights to help you explore as you go along.

Type a question...

**ORIGINAL SYSTEM VIEW**

mbdata/ccdata

19:16:07  
KronosStar: An additional police unit has arrived at the scene to help with crowd control . #KronosStar #DancingDolphinFire

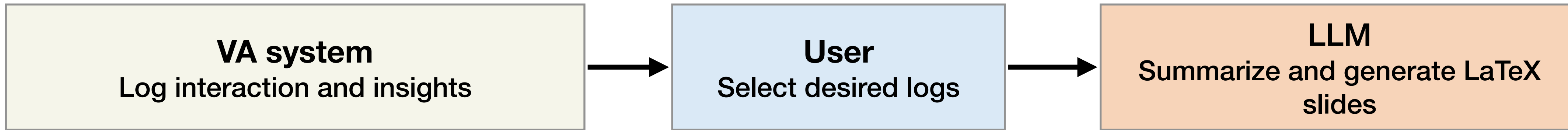
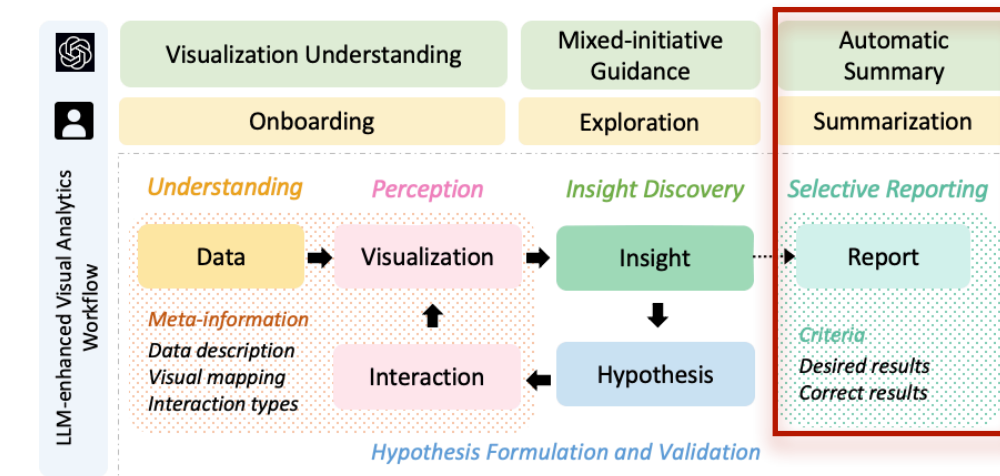
19:16:07  
Homelandilluminations: More police have arrived for evacuation and crowd control #HI

19:16:07  
cangermice: Wow - more police #abilafire

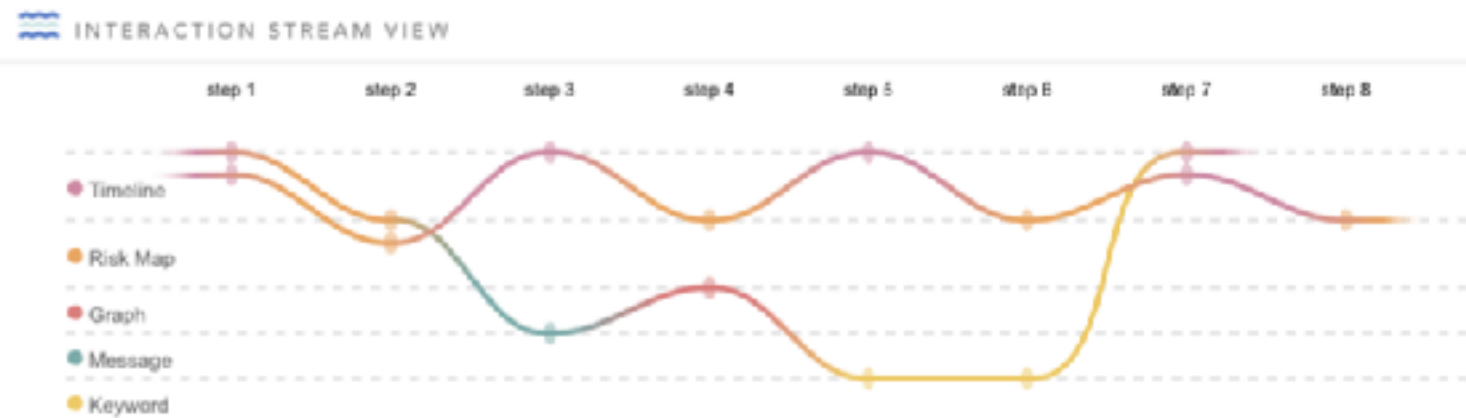


# Methods

## Report generation

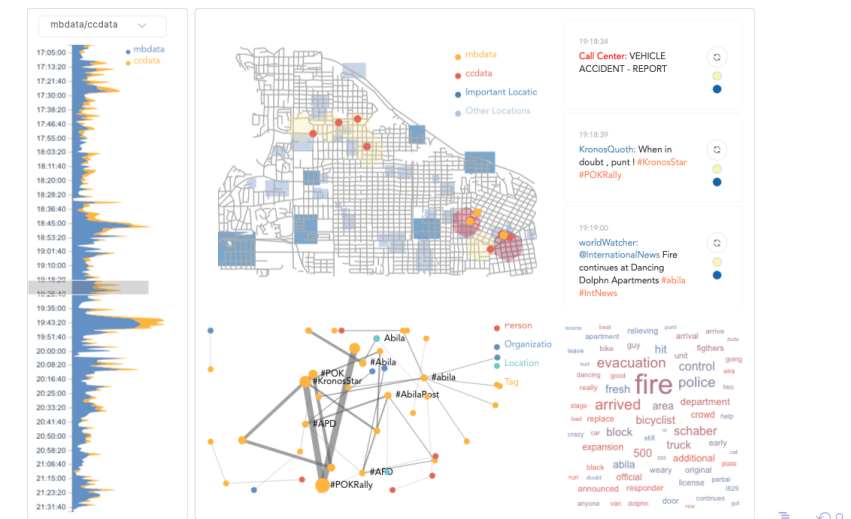


## Visualize analysis history in a stream graph



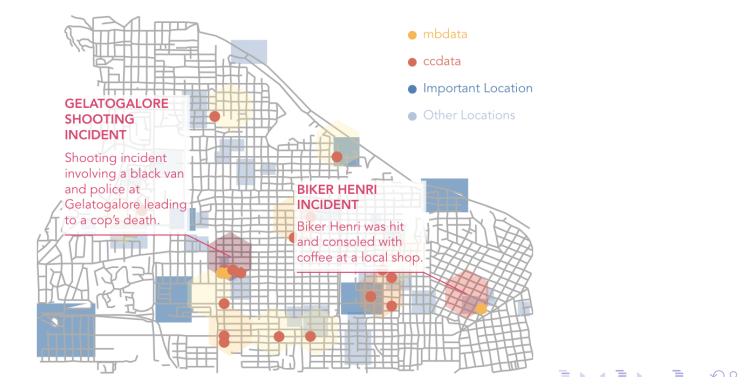
### Specific Time Period of Interest

- During the evening, a specific time period between 19:34 to 19:43 surfaced as particularly significant.



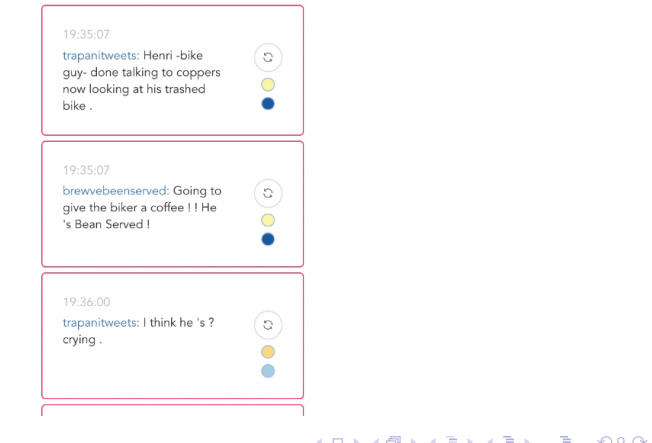
### Major Risk Events

- Shooting incident at Gelatogalore
  - Confrontation between a black van and the police.
  - Resulted in a police officer's death.
- Vehicular incident involving Biker Henri
  - Biker Henri was hit.
  - Later consoled with a coffee at a local shop.



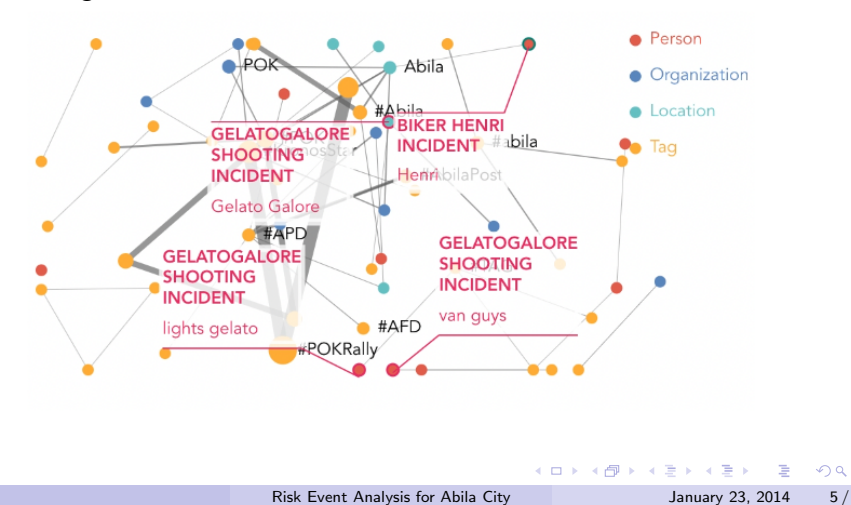
### Messages Analysis

- Real-time reactions from the public on the incidents.
- Considerable sympathy and concern for Biker Henri.
- Chaos and fear during the Gelatogalore Shooting Incident.



### Graph-based Representation

- Nodes associated with two primary incidents.
- Hashtags and entities linked to each event.



### Keyword Analysis

- Keywords associated with Gelatogalore Shooting Incident:
  - 'shot', 'police', 'gelato', 'gelatogalore', 'van', 'cop'
- Keywords linked with Biker Henri Incident:
  - 'hit', 'run', 'coffee', 'park', 'driver'



### Conclusion

- The evening was fraught with tension, characterized by two major risk events.
- Data-driven insights derived from analyses highlight the sequence, nature, and public sentiment of the events.
- Provides a foundation for future investigations and response strategies.



# 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)

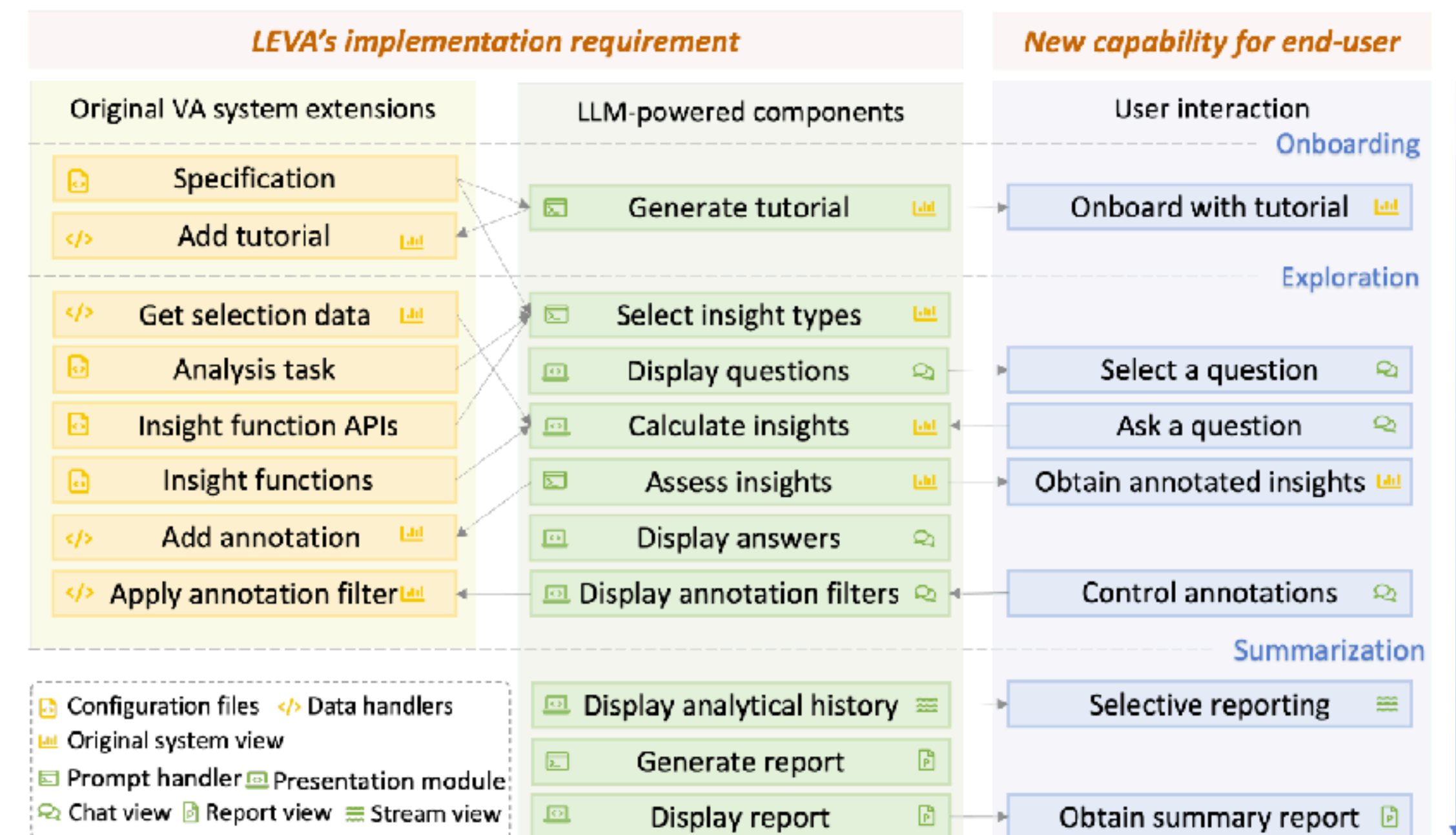
The new capability for end-users after enhancement:

## 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



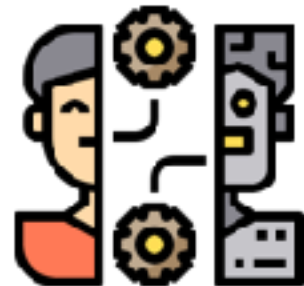
## Extend and generate the grammar of VA systems to facilitate LLM's understanding

- *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.*



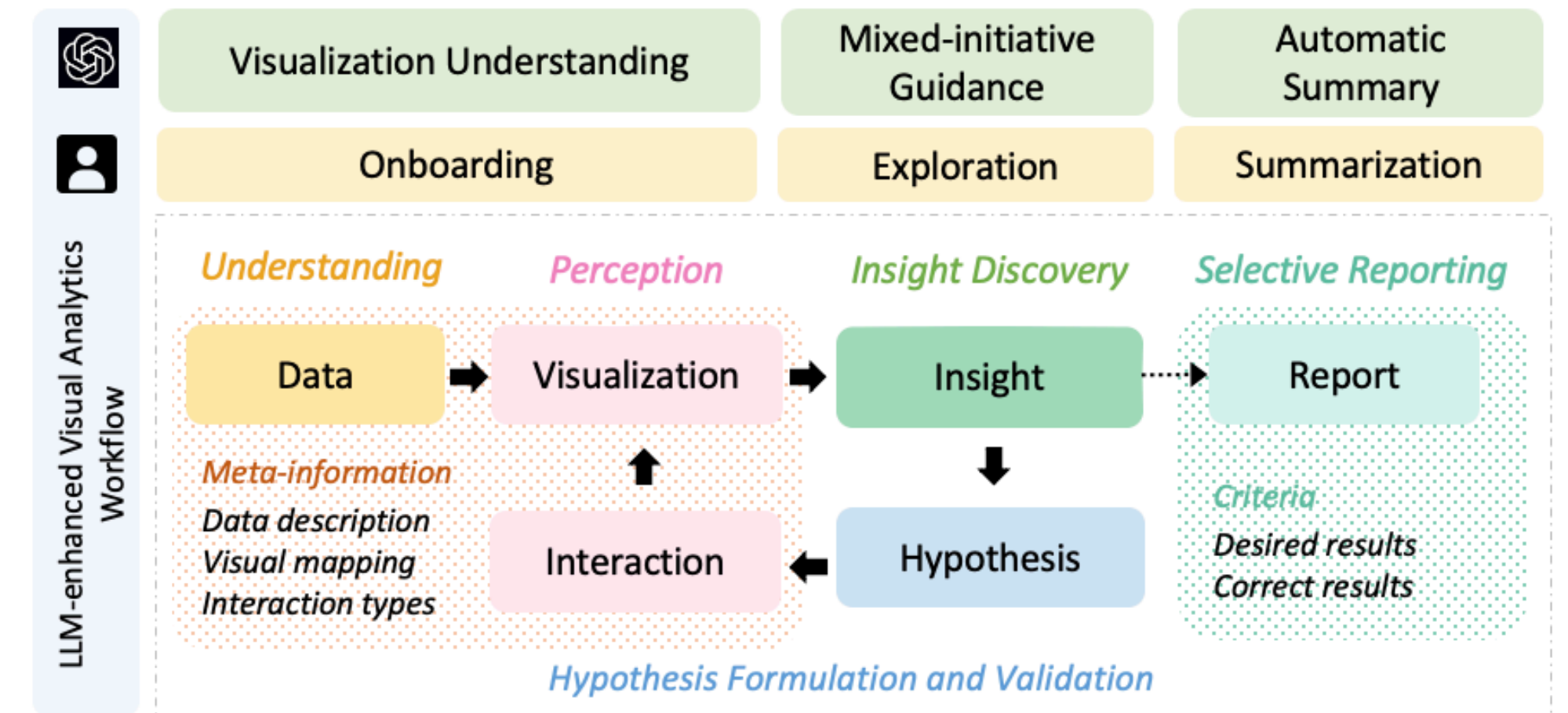
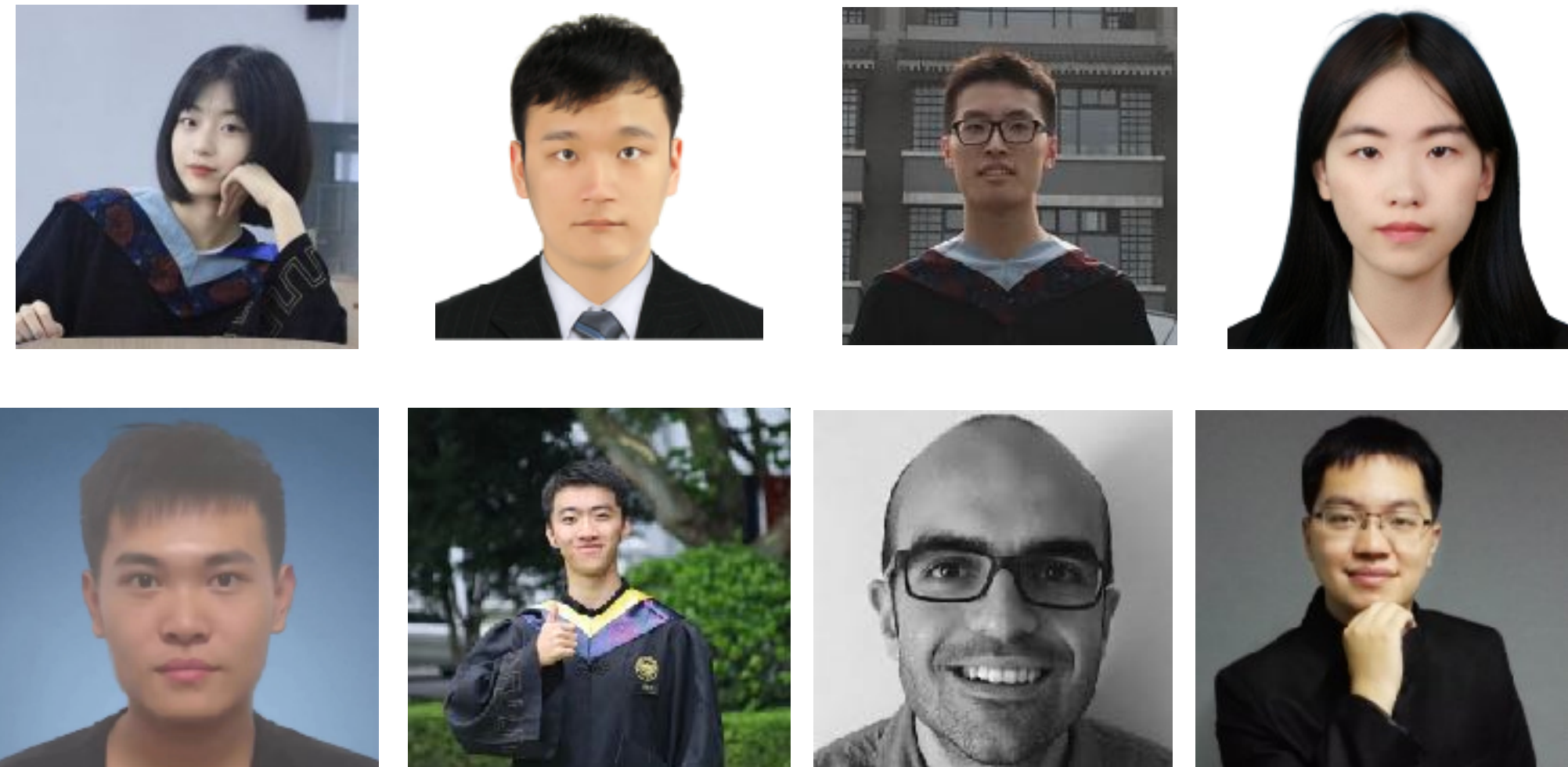
## Mitigate concerns when using LLMs

- *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.*