

LightVA: Lightweight Visual Analytics With LLM Agent-Based Task Planning and Execution

Yuheng Zhao¹, Junjie Wang, Linbing Xiang, Xiaowen Zhang, Zifei Guo,
Cagatay Turkay³, Yu Zhang², Siming Chen¹



¹ Fudan University ² University of Oxford ³ University of
Warwick

Motivation



Visual Analytics System

Large exploration space

Fixed Space



Task planning

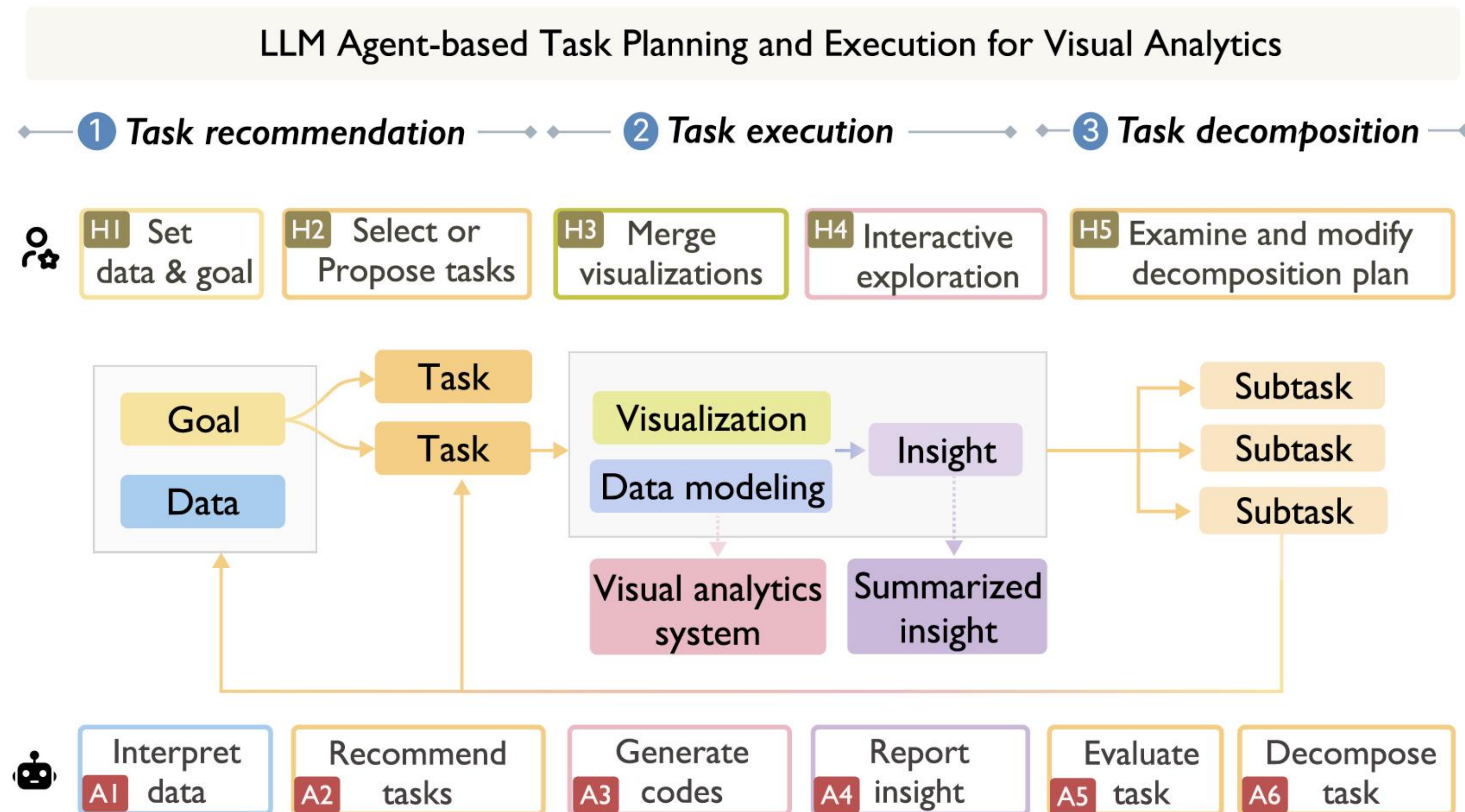


Analyst

Have to plan and decompose tasks

Some new tasks/ideas cannot be solved promptly

Framework



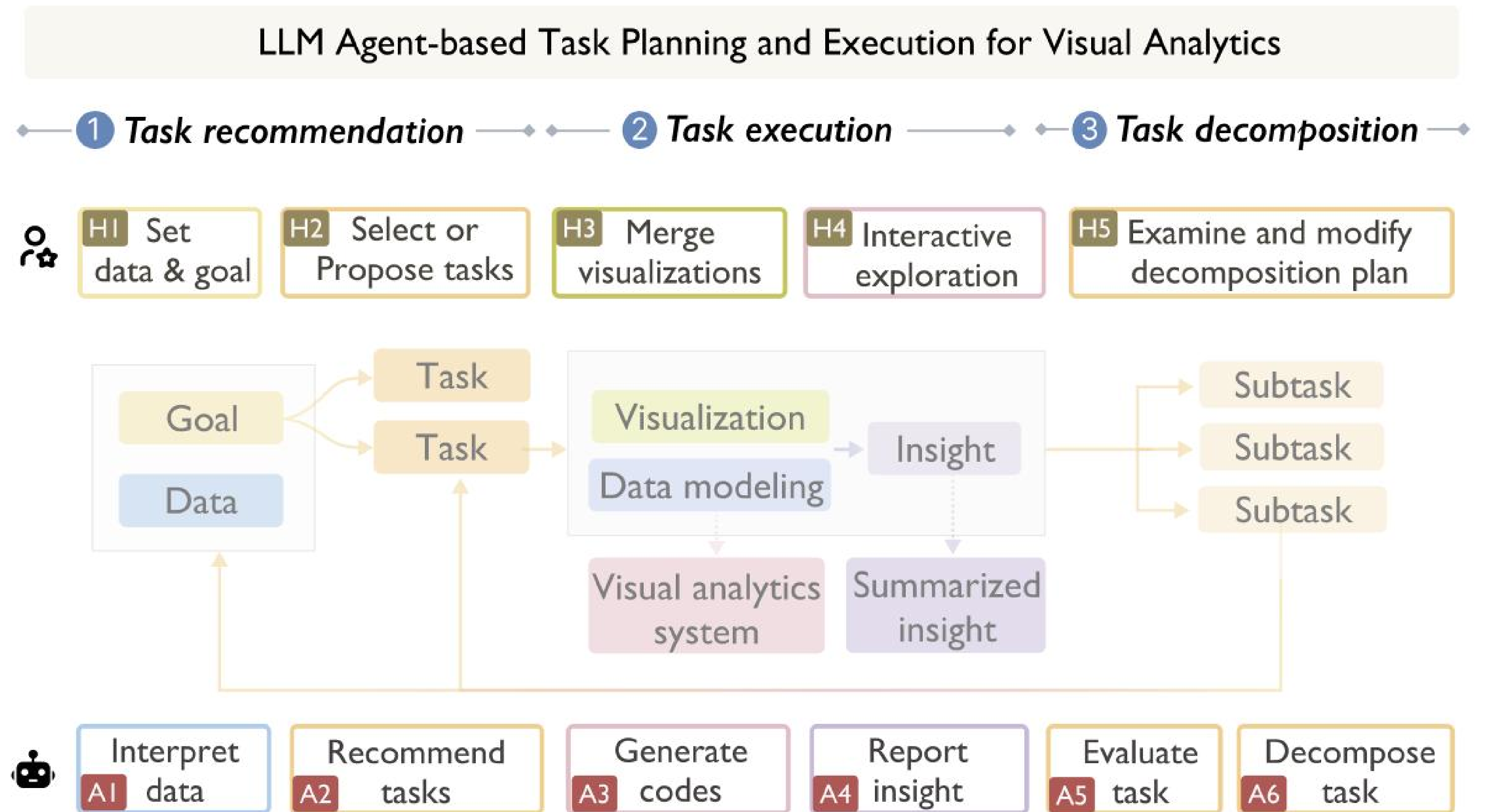
Task decomposition: The goal–task–subtask progressive structure that helps users navigate a large analytical space.

Malleable interface: generate visualization & insights based on user needs.

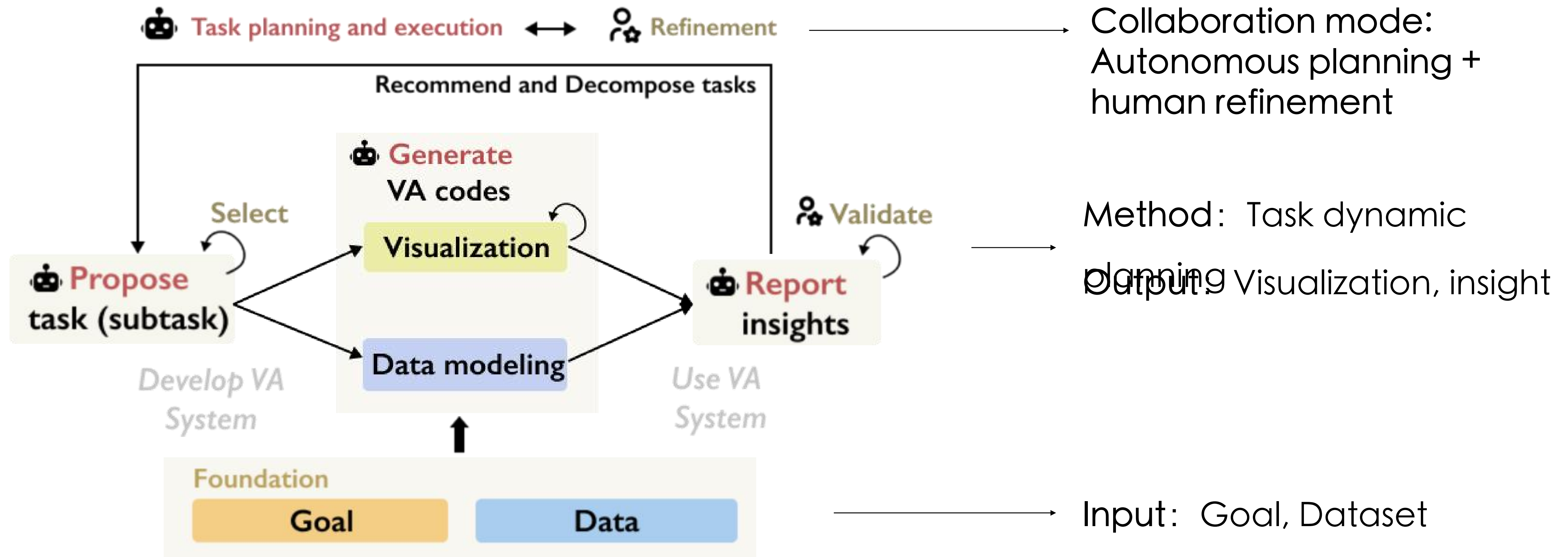
Framework

- Human roles:**
- Goal setting
 - Visual perception
 - Validation

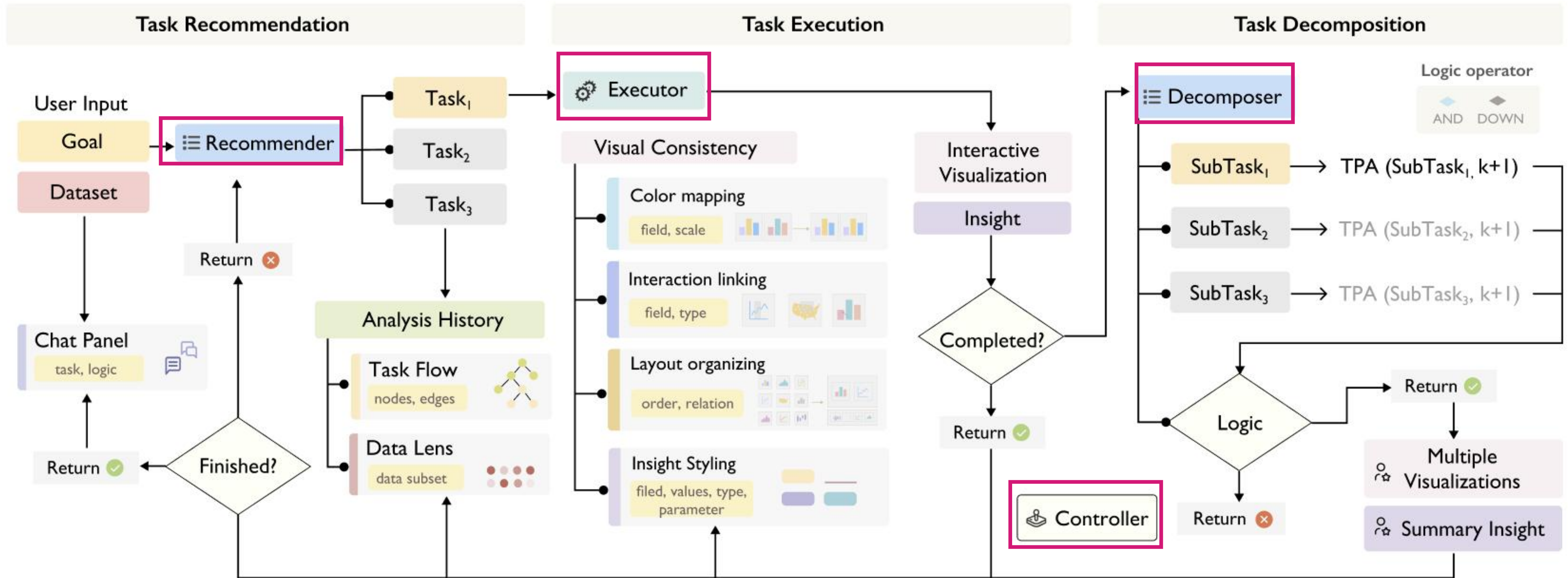
- Agent roles:**
- Task planning
 - Data modeling
 - Explanation



Framework



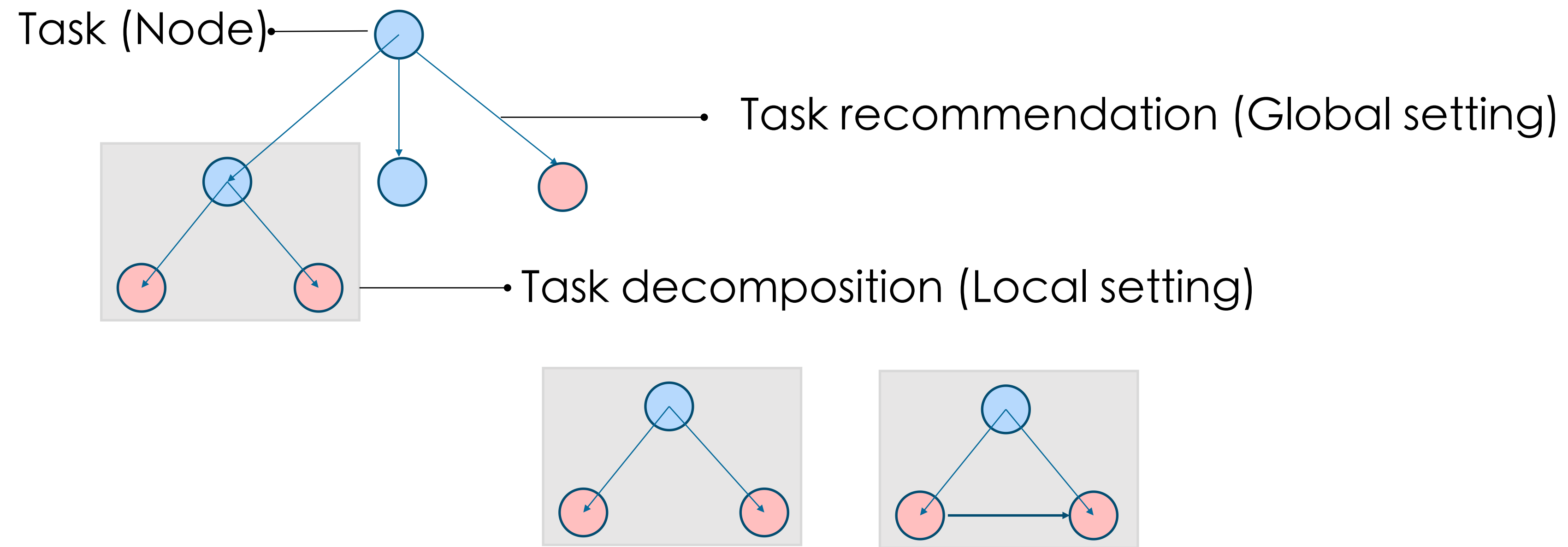
Workflow: Agent-Based Task Planning



Modeling

The definition of the task planning process

We define the task flow as a directed graph, where nodes represent tasks, edges represent the relationships between tasks, and tasks are executed with the logic.



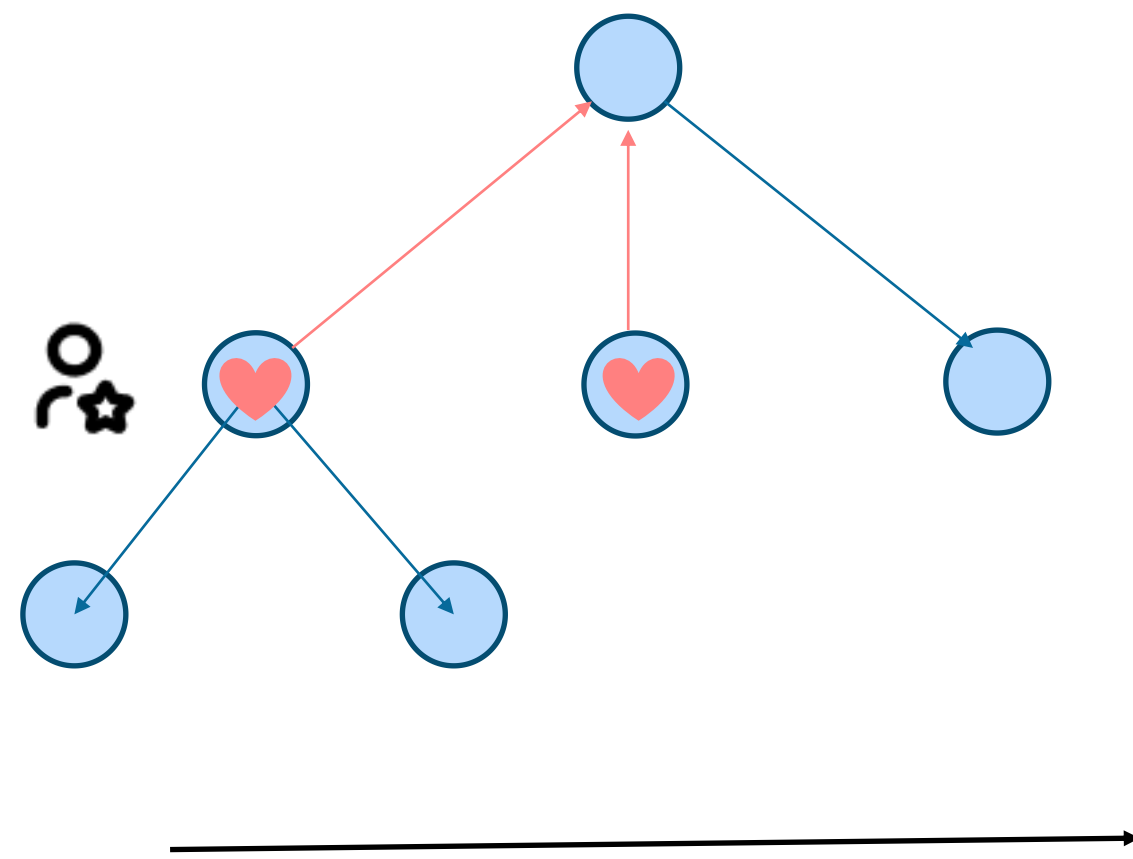
● Indicates a newly proposed task

Modeling

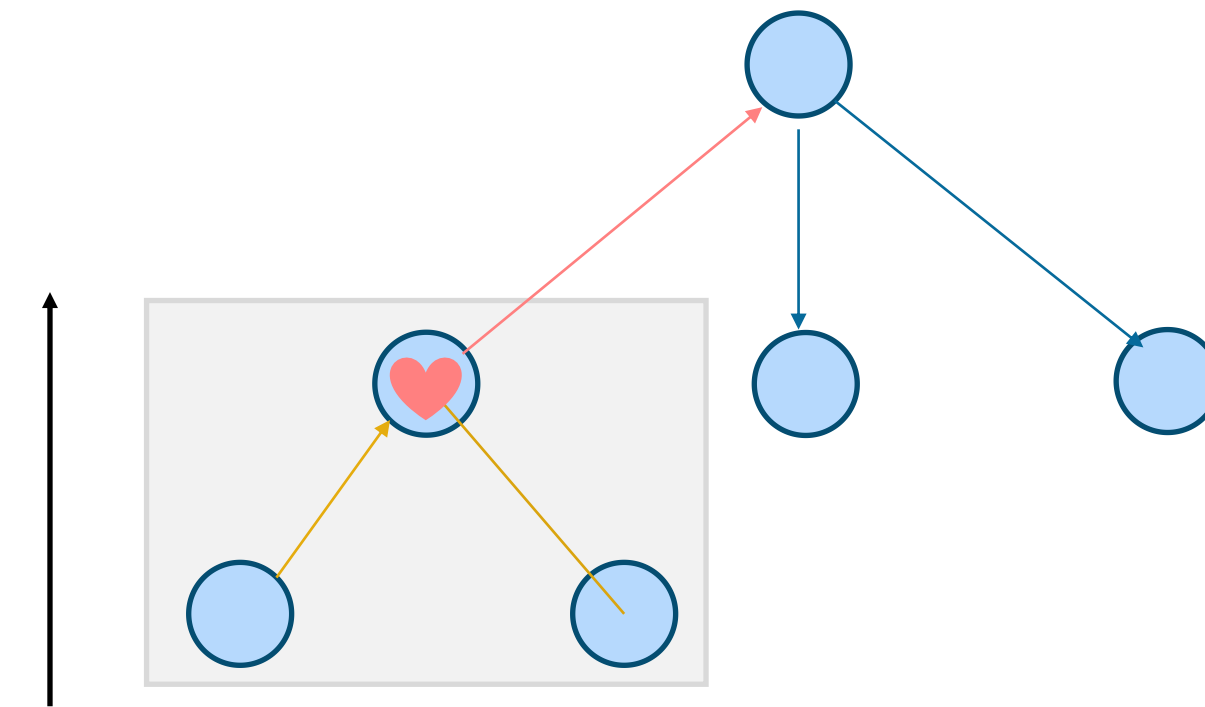
The definition of goal completion

We define *task progress* for visual analysis of the context, meaning:

- (1) The user first needs to confirm whether the task should be executed; if confirmed
- (2) The completion of this task depends on whether its subtasks are completed.



(1) When all tasks that the user expects to be completed are done, the goal is achieved.



(2) Subtasks are executed according to the decomposition plan; when all subtasks are completed, the main task is considered complete.

 The circle with a heart indicates a task the user expects to execute.

Modeling

Task recommendation

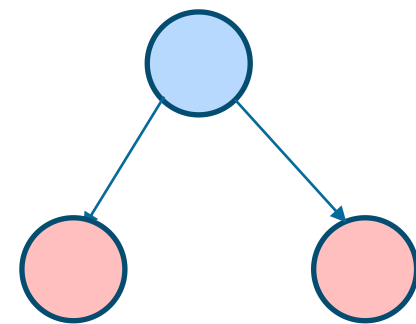
Goal: “Tell me about fuel efficiency”

Task 1: Identify the key factors affecting fuel efficiency, such as engine size, weight, and aerodynamics.

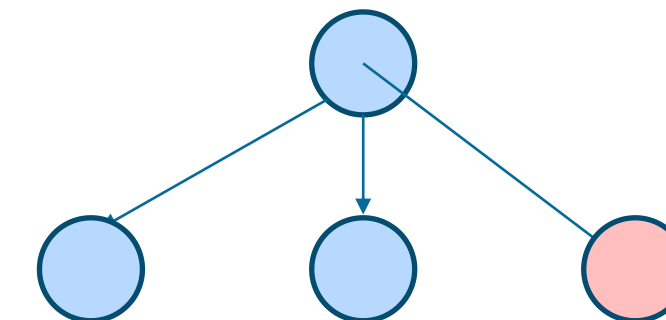
Task 1: Identify the key factors affecting fuel efficiency, such as engine size, weight, and aerodynamics.

Initial stage

- Initially determine the *data variables* related to the goal
- Recommend different *task types* according to the goal and data



 The circles with a red outline represent newly proposed tasks.



Based on historical analysis records

- Propose *new tasks* according to the current context and previous tasks
- Remind users of *unexplored tasks* that need to be revisited

Modeling

Task execution

Step1. Code generation

- Data transformation/modeling
- Vega-lite generation

Step2. Insight generation

- Compete Insight specification

Step3. Multi-view generation

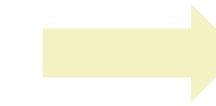
- Cognitive consistency constraint
 - Colors, layout [1]
- Interaction coordination constraints
 - Interaction Type (brush | filter | click)

```

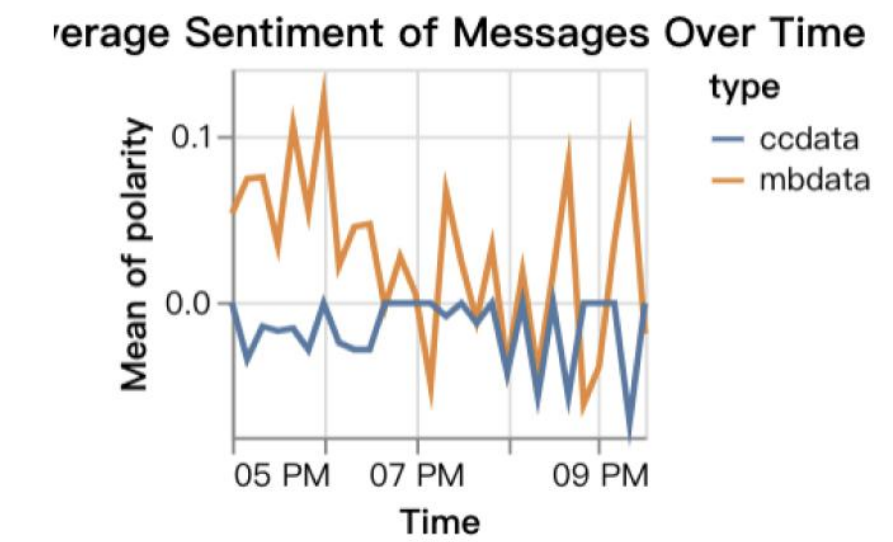
1 import altair as alt
2 import pandas as pd
3 def plot(data: pd.DataFrame):
4     # Data preprocessing
5     <codes>
6     # Chart generation
7     chart = alt.Chart().mark_bar().encode()
8     return chart
    
```

```

1 {
2   "text": "There is a correlation between A and B,
3     indicating xx generally exhibit lower B for xx in
4     C.",
5   "type": "'Correlation",
6   "parameters": "negative",
7   "data_variables": ['A', 'B'],
8   "data_values": {
9     'A': '2000 to 5000lbs',
10    'B': '10 to 35MPG',
11    'C': 'xx'
12  }
13 }
    
```



2 Analyze the sentiment of messages over time to identify any significant changes or trends.



Finding

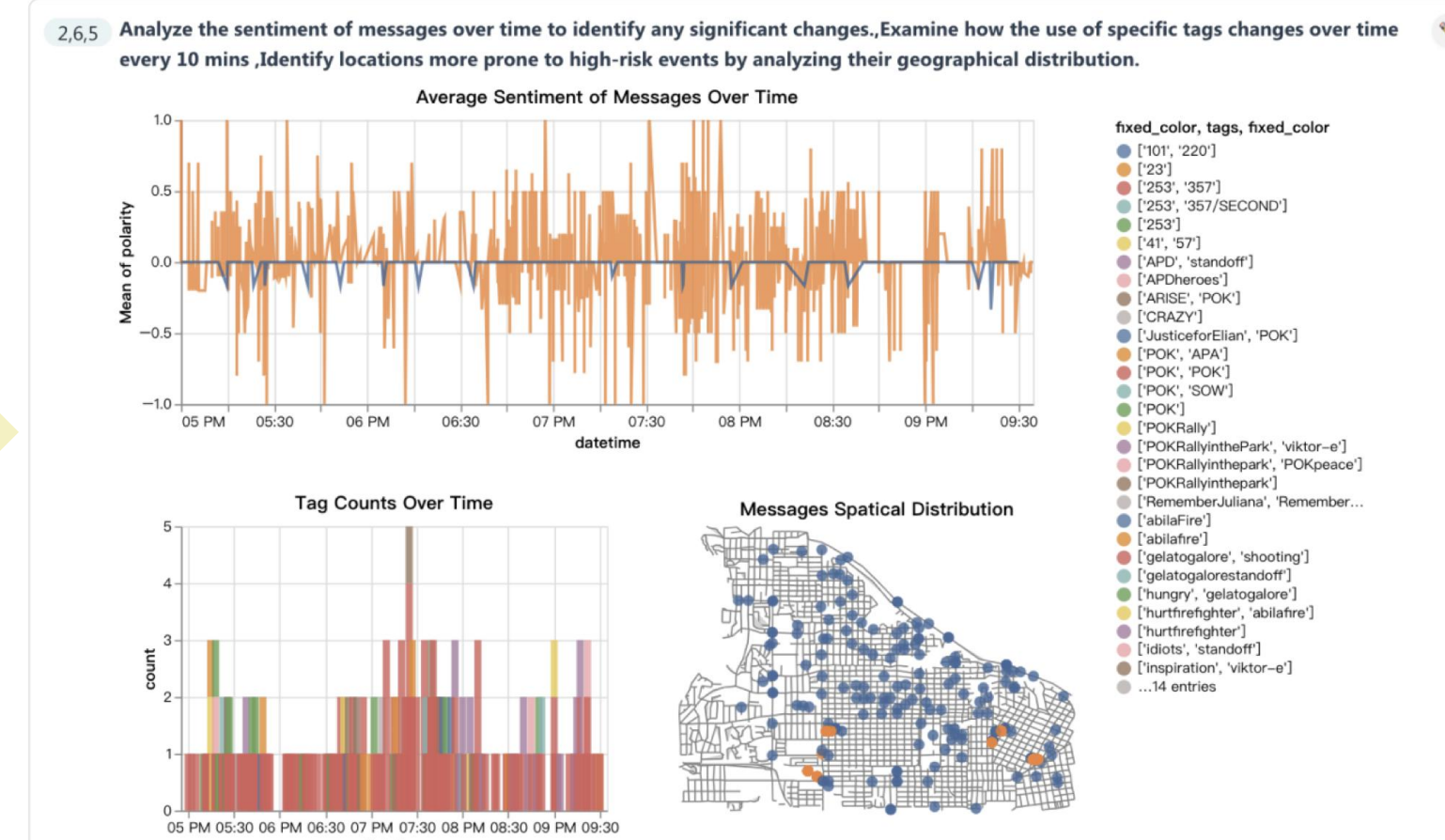
There is a significant drop **trend** in rolling average sentiment polarity under **mbdata**, indicating potential high-risk events. Specifically, these events were identified around **19:20:00**, with a minimum rolling mean polarity observed at approximately **-0.0975**.

Multi-linked Views

Cognitive consistency



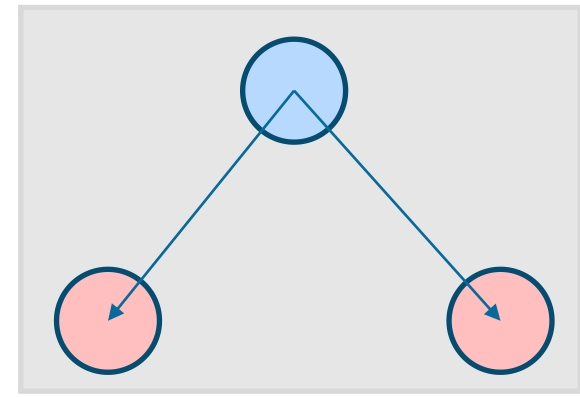
Interactive coordination



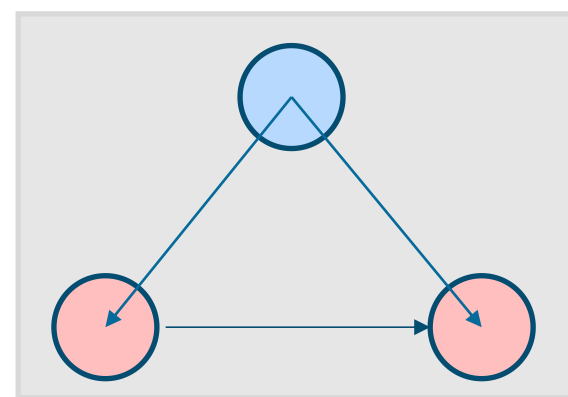
[1] Qu, Zening, and Jessica Hullman. "Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring." *IEEE transactions on visualization and computer graphics* 24, no. 1 (2017): 468-477.

Modeling

Task Decomposition



and
(parallel)



down
(sequential)

Goal: “Tell me about fuel efficiency”

Task 1: Identify the key factors affecting fuel efficiency, such as engine size, weight, and aerodynamics.

Subtask1: Analyze the relationship between vehicle weight and mpg.

Subtask2: Analyze the correlation coefficient between vehicle weight and mpg.

Subtask3: Perform a linear regression analysis to quantify the relationship.

1. Task Completion Check

- Data coverage: Covers all relevant subsets
- Insight complexity: Need advanced analysis (regression, clustering, etc.)
- Visualization: Try better ways to present insights

2. If incomplete → Decomposition Plan

- And: Multiple agents run in parallel
- Down: Single agent runs sequentially, using previous results as context

Logic:= “(Subtask1 AND Subtask2) DOWN Subtask3”

Interface

TASK FLOW Decisions

Goal: My goal is to find high-risk events in this city. 50%

Visualization of task flow

Task 4: Temporal Analysis 0%

Task 11: 50%

Task 9: Temporal Analysis 0%

Task 14: Spatial-Temporal Analysis 0%

VISUALIZATION VIEW Grid Get Multi-view Export Multi-view Code

2 Analyze the sentiment of messages over time to identify any significant changes.

Average Sentiment of Messages Over Time

Finding

There is a significant drop trend in rolling average sentiment polarity under mbdata, indicating potential high-risk events. Specifically, these events were identified around 19:20:00, with a minimum rolling mean polarity observed at approximately -0.0975.

3 Conduct a frequency analysis of keywords within the messages.

Top 20 Most Frequent Keywords in Me:

Finding

The 'fire', 'police', '#abilapost', '#pok', '#kronostar' are frequently mentioned keywords in this evening, which might suggest some high-risk events in Abila.

4 Examine how the use of specific tags changes over time.

Tag Counts Over Time

Finding

The use of tags has evolved over time, with different tags being prominent in different years. In 2018, tag 'emergency' was the most frequently used, while in 2019 it was 'dispatch'. The year 2020 saw an increase in the use of tag 'call center'.

11 Examine how the use of specific tags changes over time every 10 mins

Tag Counts Every 10 Minutes

Finding

The analysis of keyword within messages reveals specific, recurrent topics of interest among the communication

5 Identify locations more prone to high-risk events by analyzing their geographical distribution.

Messages Geographical Distribution

Finding

Among geotagged entries, there are 174 instances classified as ccd data. Comparatively, mbdata events, totaling

15 Apply clustering analysis to identify spatial patterns of high-risk events.

Clustering of High-Risk Events by Location and P

Finding

Clustering analysis reveals spatial patterns in high-risk event locations, grouping areas based on geographic

CHAT VIEW

keywords within the messages.

Examine how the use of specific tags changes over time.

Identify locations more prone to high-risk events by analyzing their geographical distribution.

trend view added

Score: The score of the result for this task is 8/10, I can further explore it.

Explanation: High-risk events identified but further analysis could include more detailed data segmentation and advanced statistical methods for deeper insights.

Task Decomposition

- Apply data segmentation to analyze sentiment fluctuations before and
- Utilize statistical methods to identify patterns in sentiment data
- Implement advanced statistical techniques for trend analysis and

Run Deny

Task Describe your questions of the dataset

Direct manipulation

Natural Language

Users steer the agent's reasoning via visualization, direct manipulation, and natural language.

Interface

Spatial analysis -> Cluster analysis + hot spot detection analysis

Insight gallery

ii. VISUALIZATION VIEW Grid Get Multi-view Export Multi-view Code

2 Analyze the sentiment of messages over time to identify any significant changes.

Finding
There is a significant drop **trend** in rolling average sentiment **polarity** under **mbdata**, indicating potential high-risk events. Specifically, these events were identified around **19:20:00**, with a minimum rolling mean **polarity** observed at approximately **-0.0975**.

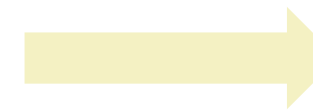
3 Conduct a frequency analysis of keywords within the messages.

Finding
The **'fire'**, **'police'**, **#abilapost**, **#pok**, **#kronostar** are frequently mentioned keywords in this evening, which might suggest some high-risk events in Abila.

4 Examine how the use of specific tags changes over time.

Finding
The use of tags has evolved over time, with different tags being prominent in different years. In 2018, tag 'emergency' was the most frequently used, while in 2019 it was 'dispatch'. The year 2020 saw an increase in the use of tag 'call center'.

Depth



5 Identify locations more prone to high-risk events by analyzing their geographical distribution.

Finding
Among geotagged entries, there are **174** instances classified as **'ccdata'**.
Comparatively, **'mbdata'** events, totaling

15 Apply clustering analysis to identify spatial patterns of high-risk events.

Finding
Clustering analysis reveals **spatial** patterns in high-risk event locations, grouping areas based on **geographic**

16 Utilize hotspot detection methods to pinpoint areas with the highest concentration of risk events.

Finding
Areas around the coordinates **(36.06, 24.86)** and **(36.05, 24.90)** have

Multi-link view

Multi-linked Views

11 Examine how the use of specific tags changes over time every 10 mins

Finding
The analysis of keyword within messages reveals specific, recurrent topics of interest among the communication

5 Identify locations more prone to high-risk events by analyzing their geographical distribution.

Finding
Among geotagged entries, there are **174** instances classified as **'ccdata'**. Comparatively, **'mbdata'** events, totaling

15 Apply clustering analysis to identify spatial patterns of high-risk events.

Finding
Clustering analysis reveals **spatial** patterns in high-risk event locations, grouping areas based on **geographic**

Breadth



2,6,5 Analyze the sentiment of messages over time to identify any significant changes. Examine how the use of specific tags changes over time every 10 mins. Identify locations more prone to high-risk events by analyzing their geographical distribution.

fixed_color, tags, fixed_color

- ['101', '220']
- ['23']
- ['253', '357']
- ['253', '357/SECOND']
- ['253']
- ['41', '57']
- ['APD', 'standoff']
- ['APDheroes']
- ['ARISE', 'POK']
- ['CRAZY']
- ['JusticeforElian', 'POK']
- ['POK', 'APA']
- ['POK', 'POK']
- ['POK', 'SOW']
- ['POK']
- ['POKRally']
- ['POKRallyinthePark', 'viktor-e']
- ['POKRallyinthePark', 'POKpeace']
- ['POKRallyinthePark']
- ['RememberJuliana', 'Remember...']
- ['abilaFire']
- ['abilafire']
- ['gelatogalore', 'shooting']
- ['gelatogalorestandoff']
- ['hungry', 'gelatogalore']
- ['hurfirefighter', 'abilafire']
- ['hurfirefighter']
- ['idiots', 'standoff']
- ['inspiration', 'viktor-e']
- ...14 entries

A Case Study Demo

Usage Scenario · VAST Challenge 2021 · Mini-Challenge 3



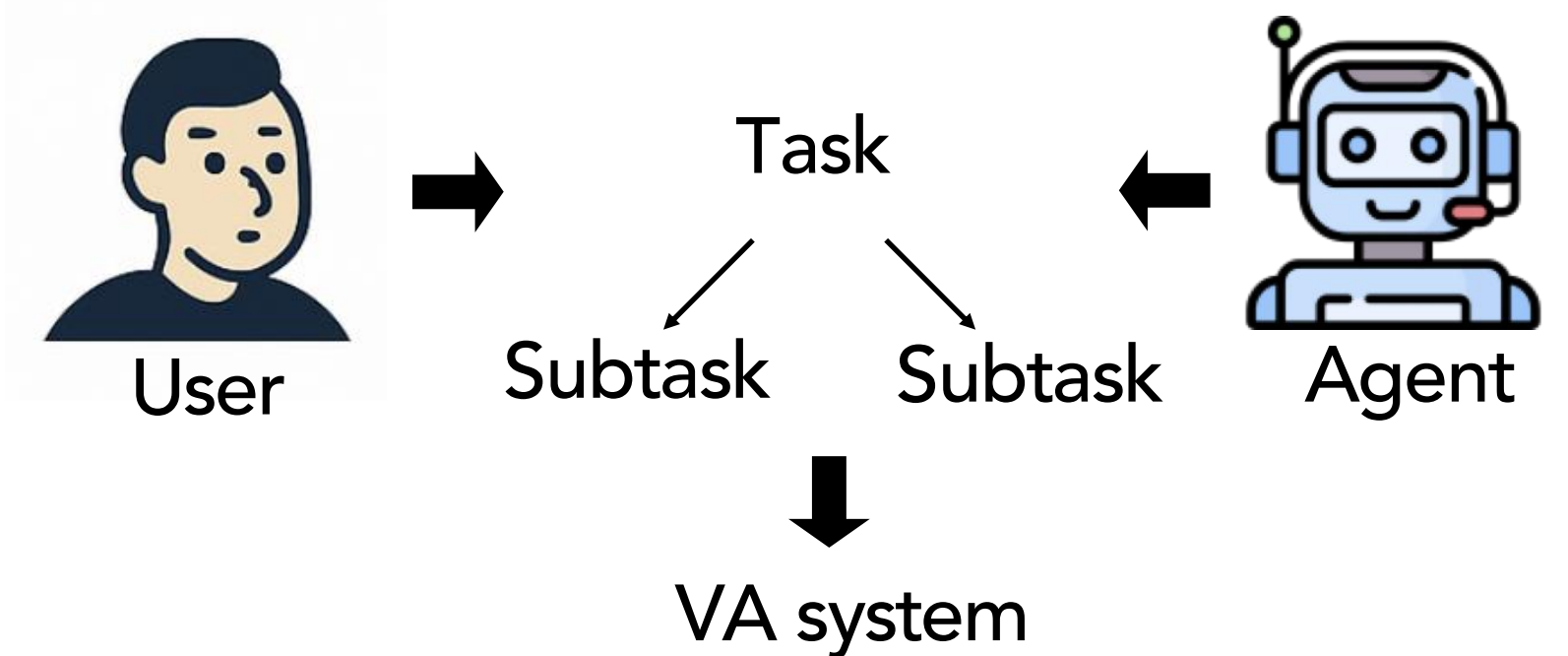
- **Background**: On January 23, 2014, multiple events unfolded in Abila.
- **Data**: Social media messages with sentiment scores and locations

To evaluate our framework

Discussion

Findings from the expert studies

- LightVA shortens the process of system development and analysis that originally took about 30 days to just 2-3 hours, expanding the depth and breadth of exploration.
- Compared with end-to-end agent-based data analysis, progressive task co-planning help enhance user engagement.



Open questions

Towards mixed-initiative collaboration, not only tool, but also partners:

- Visualization (perception & interaction) can serve as medium a that benefits customization, validation, and **explanation** of planning process.
- Building the task graph from different domain VA as knowledge base, more **adaptive** task planning
- Rethinking **evaluation** — beyond task accuracy toward collaborative performance (insight discovery ratio)

Thank you!

yuhengzhao@fudan.edu.cn
simingchen@fudan.edu.cn

More details in our paper

Project Website: <https://zyh1222.github.io/LightVA/>

