

Motivation

- A **unified framework** to analyze unstructured and structured data.
 - Structured** data (e.g., records, text, ...)
 - Unstructured** data (e.g., images, videos, ...)
- Compare our approach with the potential of LLMs for **holistic analysis**.
- This work is a prelude to understanding the capabilities and limitations of LLMs for each data type to move towards **holistic analysis**
 - Identify LLMs that are good for image/video analysis.
 - Identify LLMs that can perform well on structured data.

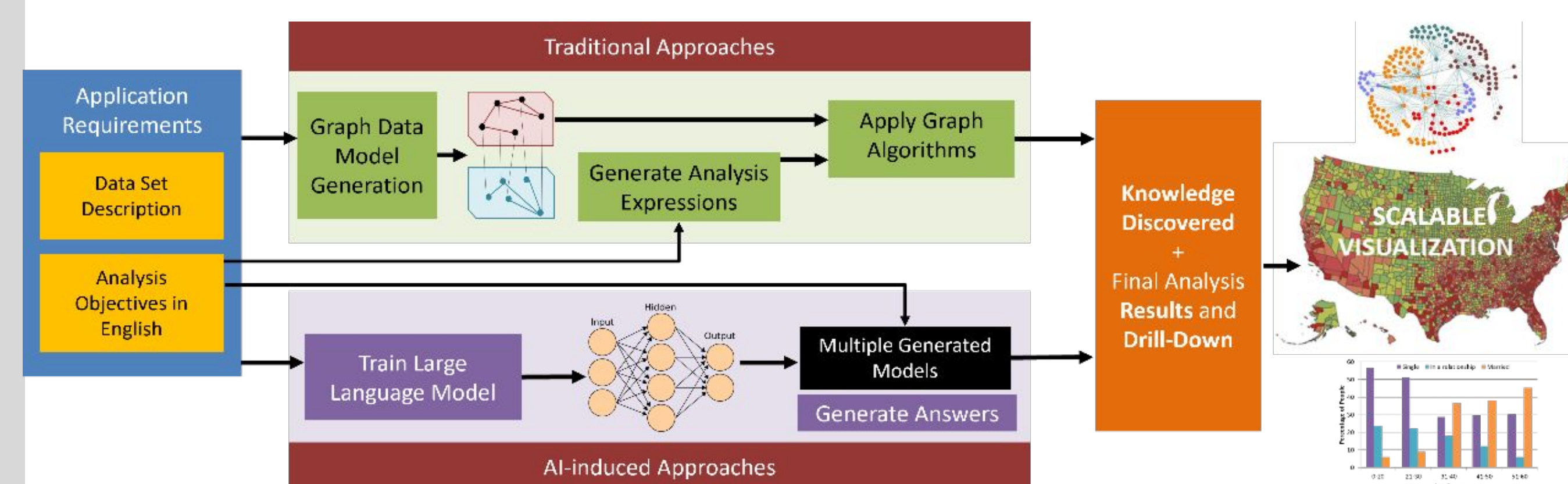
Objectives

- Compare open-source LLMs alternatives' for
 - Meaningful structured analysis/inference using traditional datasets: IMDB, US Census
- Analysis of images/video for identifying specific situations

Limitations of LLMs (without subscription)

- Input size is based on the **number of tokens** generated by the data
- GPU Memory** size determines the number of tokens that can be handled by the LLM
- Processing power** (CPU or GPU) determines the **time taken** for analysis
 - For **video analysis**, we were able to process **420 frames** in a single batch on one H100 GPU (80 GB memory) using the 4B-parameter Gemma model.
 - For **traditional data analysis**, Gemma 1B-parameter model processed **5000 rows** from the IMDB dataset in a single batch.

Lifecycle of Holistic Data Analysis (Two Approaches)



Structured Data Analysis (Human Expert vs LLMs)

IMDB Dataset Attributes: Tconst, TitleType, PrimaryTitle, Director, Language, Genre_label.

Number of rows in IMDb Dataset: ~95k

Summary of prompt given: *Here is the IMDB dataset schema, and imdb.csv. Answer the following questions.*

Q1: Count how many movies have ratings between 5 and 7.

Q2: Find the average rating of Action movies.

Q3: Find the top 5 actors with the most movies.

Comparison of Human-written Query Output with LLM Generated Output

Query	MySQL Output	Gemini Output	ChatGPT Output
Q1	61,257 rows	61,257 rows (Same #rows as MySQL)	51,005 rows (Different from MySQL, Gemini)
Q2	5.79	5.79 (Same as MySQL)	5.96 (Different from MySQL, Gemini)
Q3	5 rows	5 rows (Different from MySQL)	5 rows (Different from MySQL, Gemini)

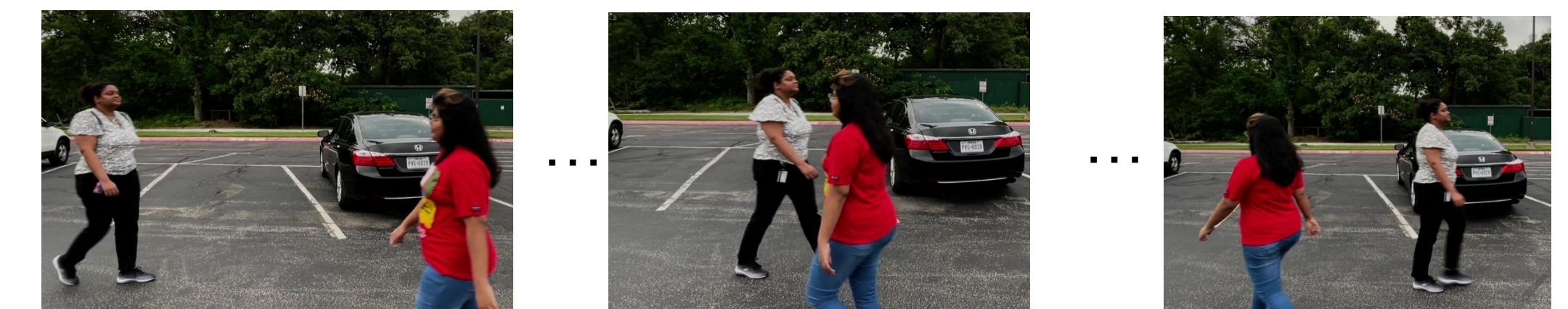
Query Q3 Outputs

MySQL		Gemini		ChatGPT	
Actor Name	Total Movies	Actor Name	Total Movies	Actor Name	Total Movies
Ove Sprogøe, Morten Grunwald, Poul Bundgaard	12	Amitabh Bachchan	125	Andy Lau	125
Vesa-Matti Loiri, Tapio Hämäläinen	10	Mithun Chakraborty	124	Amitabh Bachchan	124
Kaneta Kimotsuki	10	Simon Yam	119	Mithun Chakraborty	119
Kaneta Kimotsuki, Kazuya Tatekabe	9	Mohanlal	118	Simon Yam	118
Dermot Mulroney	9	Dermot Mulroney	114	Mohanlal	114

Observations: LLM outputs vary by query type (aggregation vs nested query, etc.)—accurate for simple queries but **inconsistent for complex ones** on the same dataset.

Video Situation Analysis

- Example Situation:** Two people crossing each other.
- An input video with 1600 frames



- LLM Used:** Gemma 4B, **System:** Stampede (H100 GPU, 80 GB memory)
- [Exp #1] **Sample Output with Prompt:** *State the action happening in this video ...*

# Frames	Manual Ground Truth (GT)	Model Output (Raw)
90	Two people coming closer to each other from opposite directions in a parking lot	Two people are walking towards a black car in a parking lot. (Similar to Manual GT)
240	Two people crossing each other in a parking lot	Two people are walking past a black car in a parking lot. (Different from Manual GT)
270	Two people crossing each other in a parking lot	\n * \n * \n * (Gibberish output)

[Exp #2] **Sample Output with Prompt:** *Choose the action labels from: [Two people exchanging an object, Two people coming closer to each other from opposite directions, ...]*

# Frames	Manual Ground Truth (GT)	Model Output (Manually Cleaned)
90	Two people coming closer to each other from opposite directions in a parking lot	Two people coming closer to each other from opposite directions (Same as Manual GT)
240	Two people crossing each other in a parking lot	Two people exchanging an object (Different from Manual GT)
270	Two people crossing each other in a parking lot	\n * \n * \n * (Gibberish output)

- Only 420 frames can be processed in a single batch on an H100 GPU

Observations: Gemma correctly selects the labels (as given in the prompt) for up to a certain number of frames. Gemma **outputs gibberish after 270 frames for both prompts**.

Conclusions (Public-domain LLMs)

- Our experiments raise several questions to be explored further
 - Do LLMs process **the whole dataset or a snapshot**?
 - How and why do LLMs generate **different outputs** (e.g., SQL queries) for similar prompts?
 - Why does an LLM give **gibberish/incorrect output**?
 - How to **improve LLM-generated** output (e.g., SQL, video analysis)?
- LLMs are **promising**, but still in their **infancy in processing large volumes and a variety of data** for holistic analysis
- Publication:** icABCD 2025