

# Discovering and Understanding Events in Videos in Real Time

Alexander Martin

Johns Hopkins University

# The Instantaneous Email for X

Go put together a bunch of data and track trends so that we can have the best macro forecast on the street



Ok Jan. Will do



**First-Year Analyst**

# The Instantaneous Email to a Pod During an Event



## Analyst notes:

Trump spent 45 minutes talking about the 'Memphis Safe Task Force'

For 2 minutes trumped talked about the BBB and no tax on tips

He then spent another 45 min...

Trump said no tax on tips is going to be in the BBB, this will be good in the forecast



First-Year Analyst

# The Instantaneous Email about Unfolding Events



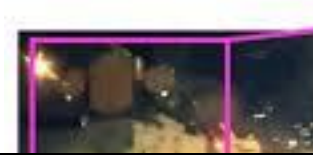


# The Instantaneous Email about Unfolding Events



NBC NEWS NOW

## How analysts are using TikTok to track Russian military movement

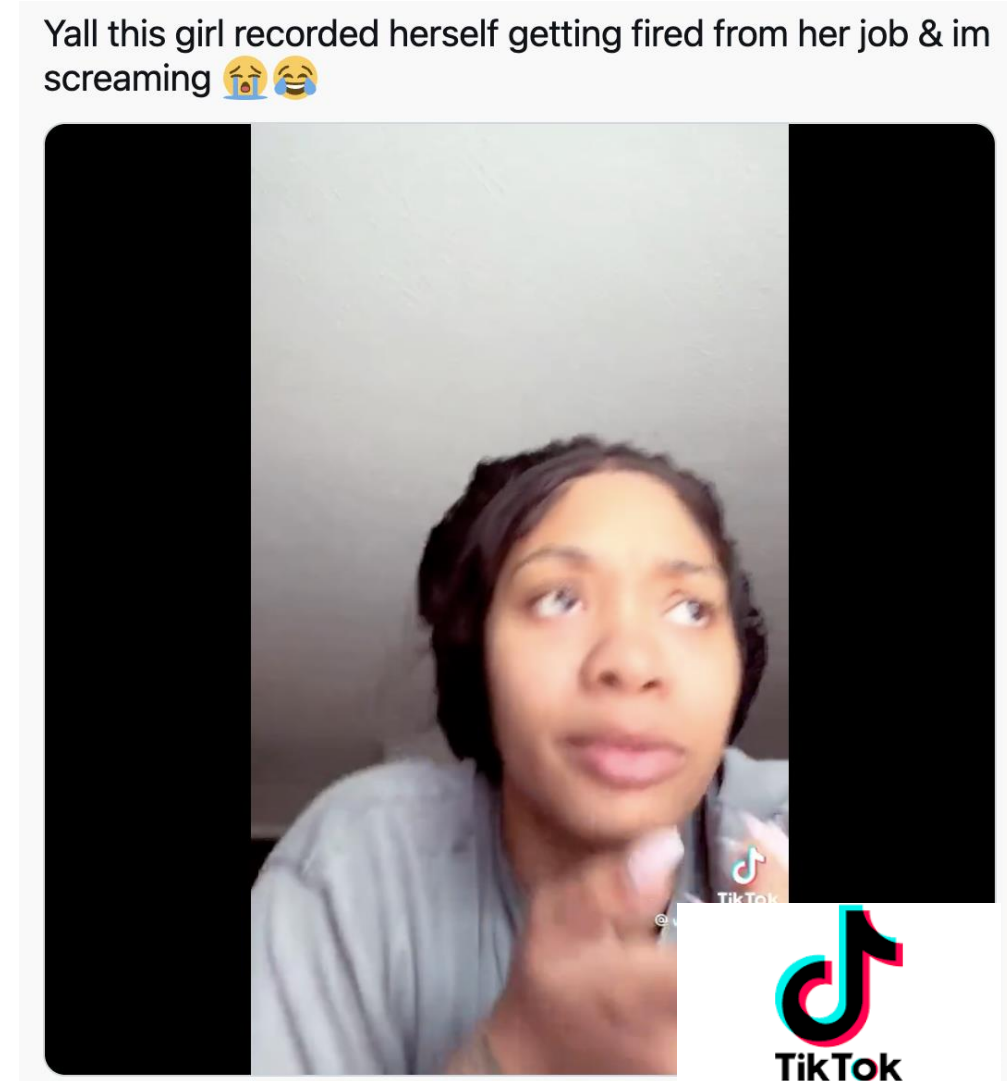
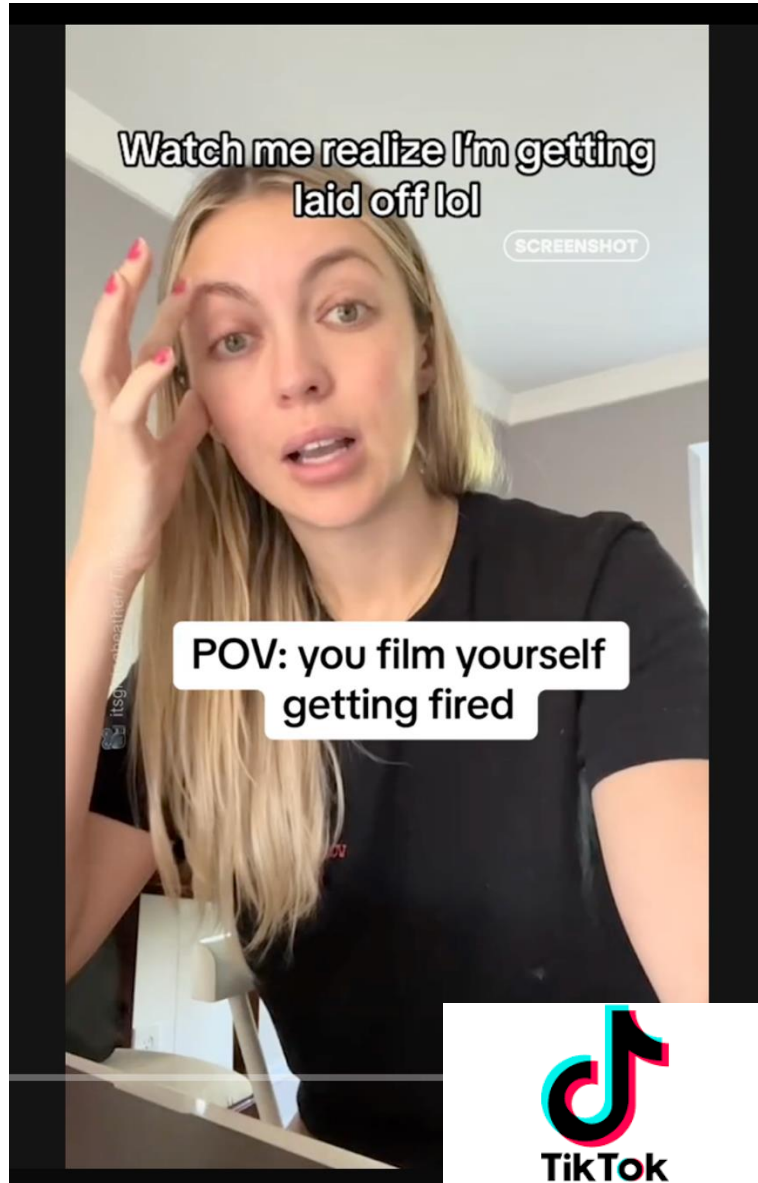


Looks like Russia is mobilizing to invade Ukraine, better put that in the forecast

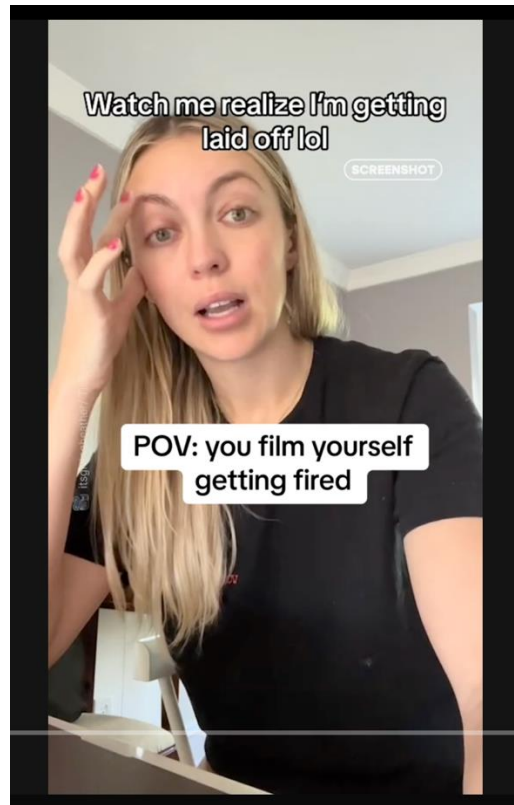


First-Year Analyst

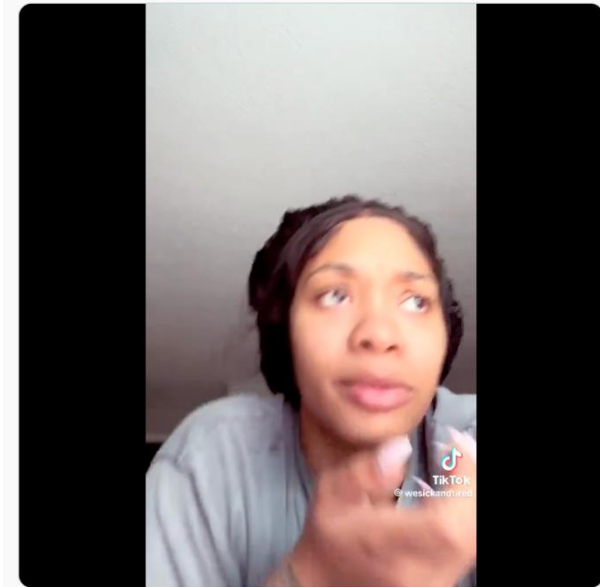
# The Instantaneous Email About New Trends in Media



# The Instantaneous Email About New Trends in Media



Yall this girl recorded herself getting fired from her job & im screaming 🤔😭



Seems like a lot of people are being fired this month, better put that in the forecast



**First-Year Analyst**



# The Instantaneous Forecast Adj. During an Event



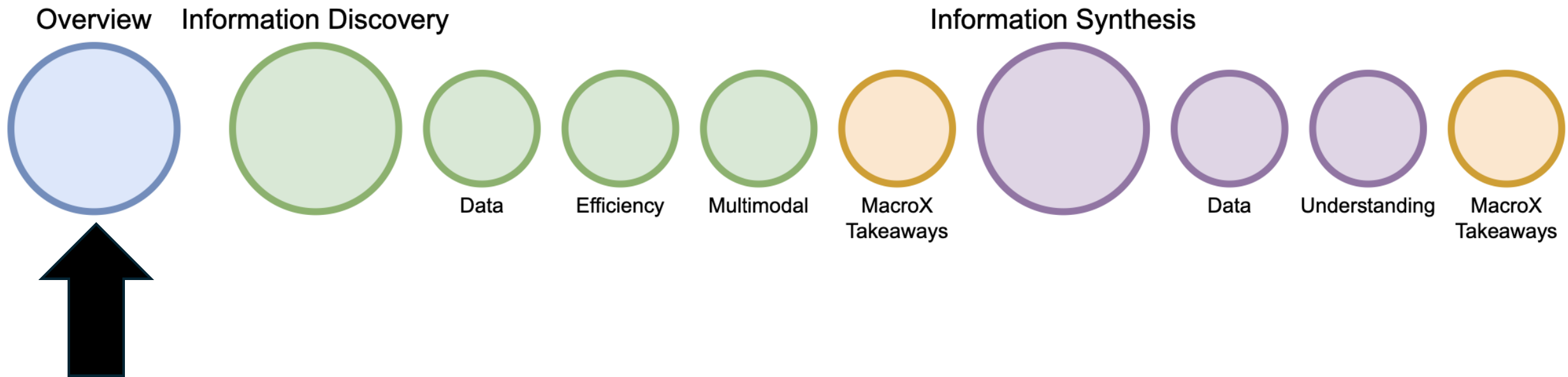
Donald Trump is shown in profile, holding a large black sign with the Presidential Seal and the title "Reciprocal Tariffs". The sign displays a table comparing the tariff rates charged by the U.S. to various countries with the rates those countries charge the U.S. The table lists 13 countries and their respective tariff percentages for both directions of trade.

Country	Tariff Charged to the U.S. (U.S. Tariff on imports from this country)	U.S. Tariff on Imports from this Country
China	67%	34%
European Union	38%	28%
Vietnam	98%	48%
Japan	64%	32%
Ukraine	46%	24%
India	52%	26%
South Korea	50%	25%
Thailand	22%	36%
Switzerland	81%	31%
Indonesia	84%	32%
Malaysia	47%	24%
Cambodia		

Looks like trump imposed reciprocal tariffs on our major trading partners, to the forecast!



First-Year Analyst



# ChatGPT, Prompted 9/15 (Abstains)

without searching, tell me about the liberation day tariffs

**I don't have a concept of**  
**something formally called**  
***“Liberation Day tariffs”* in**  
**economics, trade, or history as**  
**a standard term...**

Qwen3, Prompted 9/15 (Hallucinates)

without searching, tell me about the liberation day tariffs

**Liberation Day Tariffs refer to a set of economic measures and tariffs that were implemented in several countries, particularly in Europe and Asia, following their liberation from Axis occupation during World War II...**

# Qwen3, Prompted 9/15 (Hallucinates)

Liberation Day Tariffs refer to a set of economic measures and tariffs that were implemented in several countries during World War II...

Why can't frontier models do this task?



# Why does this happen? – Knowledge Cutoffs

- Knowledge cutoff – LLM is trained on the internet, but only up to a specific date
- LLM knowledge is static. As the world evolves, they don't
- Complete model failure (can't tell me things it hasn't seen)

# Why does this happen? – Knowledge Cutoffs

- Knowledge cutoff – LLM is trained on the internet, but only up to a specific date
- LLM knowledge is outdated
- Completion of training is a point in time

“why don’t we augment LLM knowledge by retrieving current information?”



retrieval augmented generation (RAG)

# How to write an article on demand?

An **article** is text containing high-level inference about a **topic**, either making inferences over observed content or probabilistic inferences about unobserved/future information.

Related to Article: forecast, analysis, debrief  
Related to Topic: information need, query, request

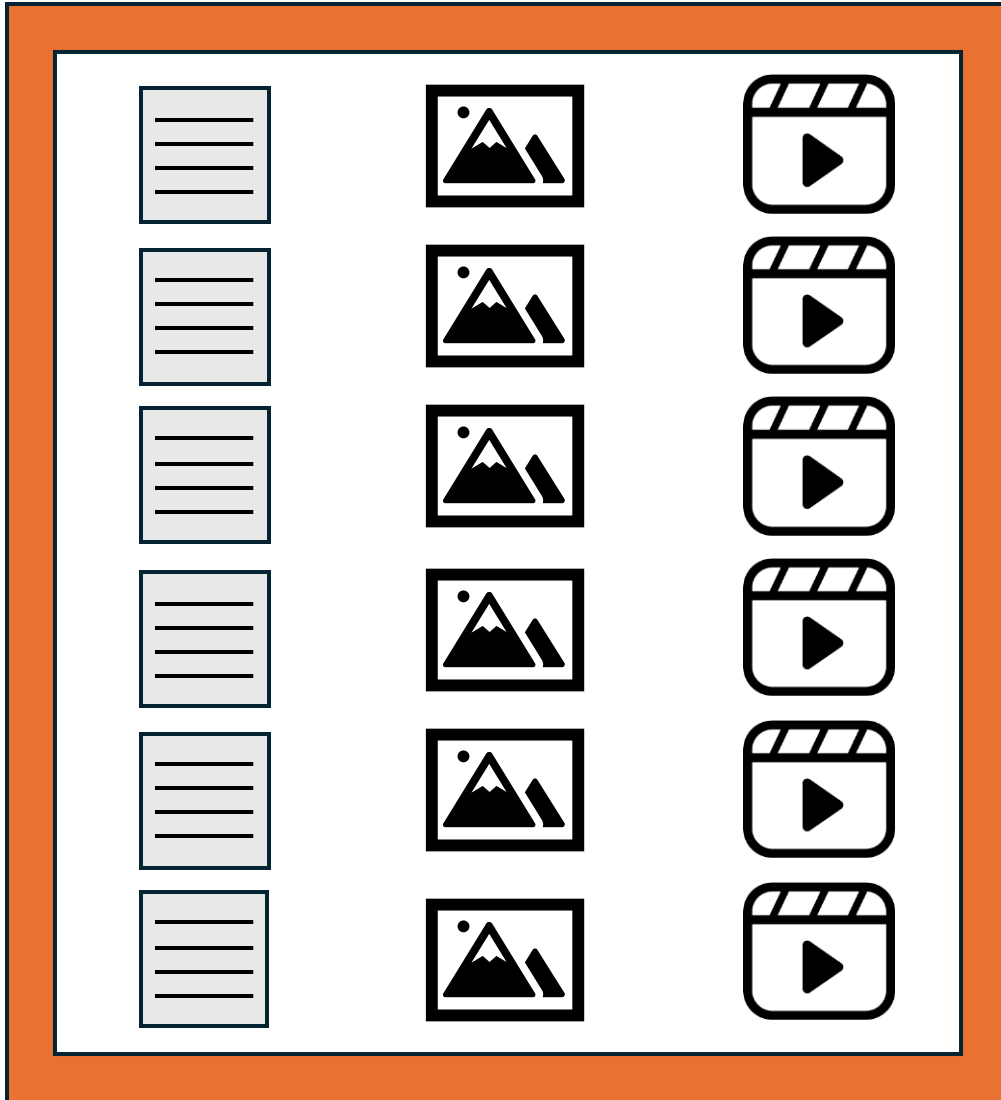
# How to write an article on demand?

**On Demand:** (1) instantaneous and (2) in response to a query/information need

**Live Data is**

1. Knowledge to put in your article
  - Trump just said no tax on tips in BBB
2. A query to the corpus
  - How does what Trump said relate to my other knowledge?

# How to write an article on demand?

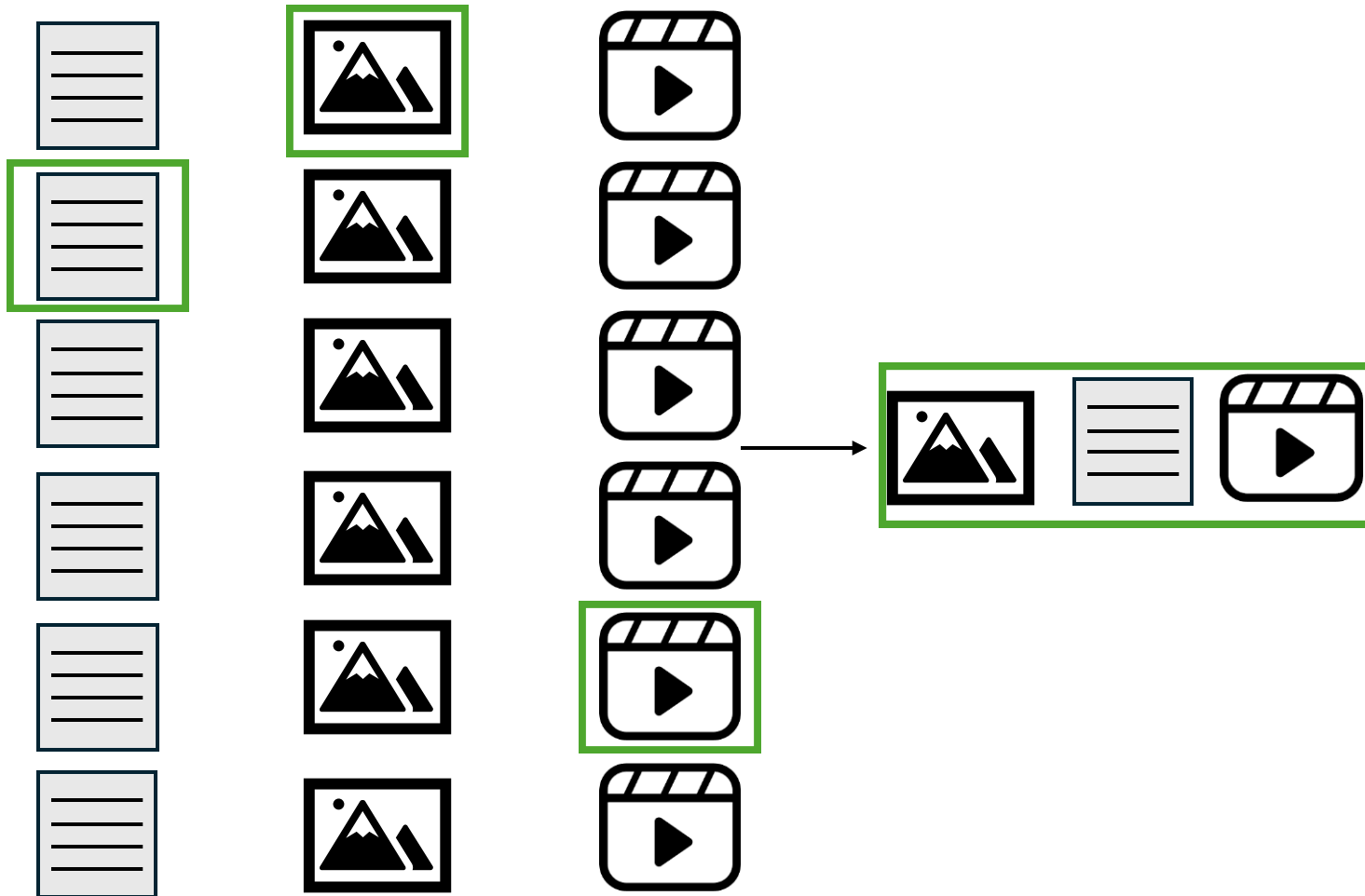


Internet Scale Data

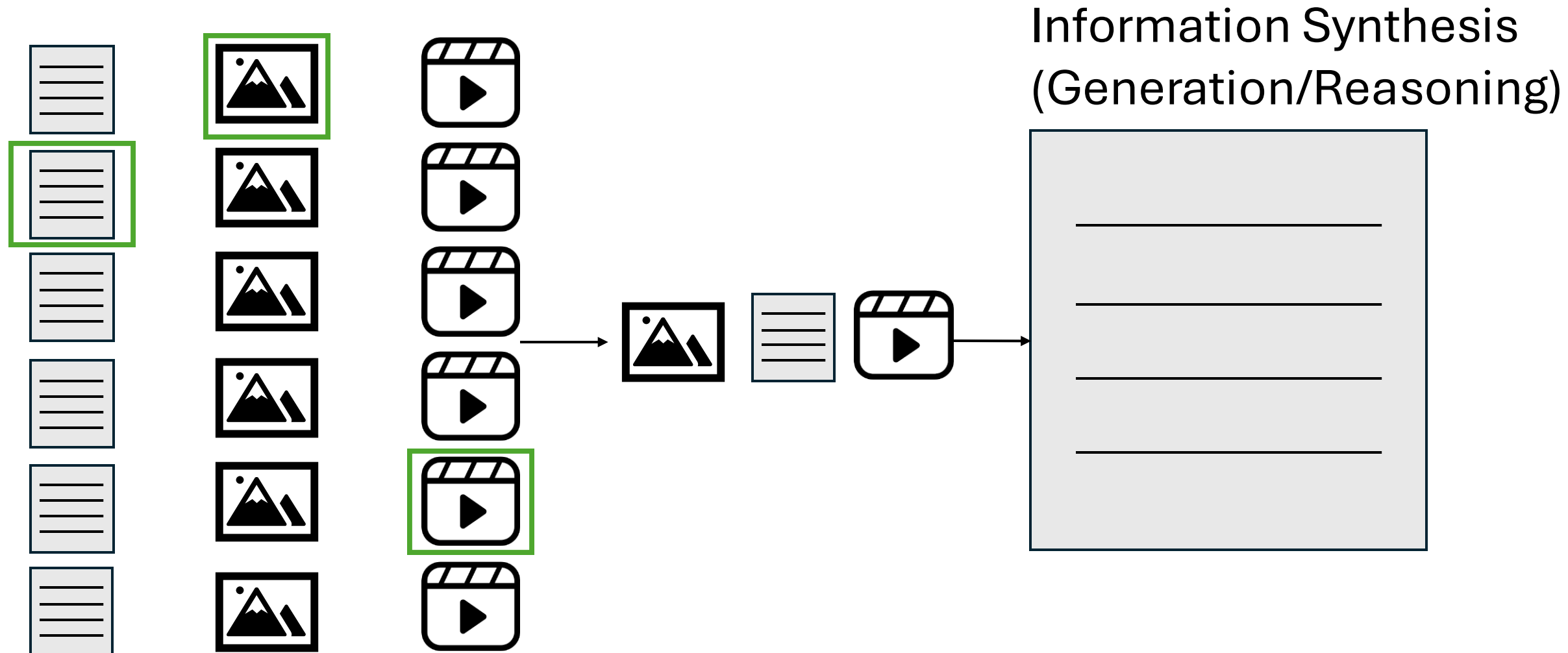


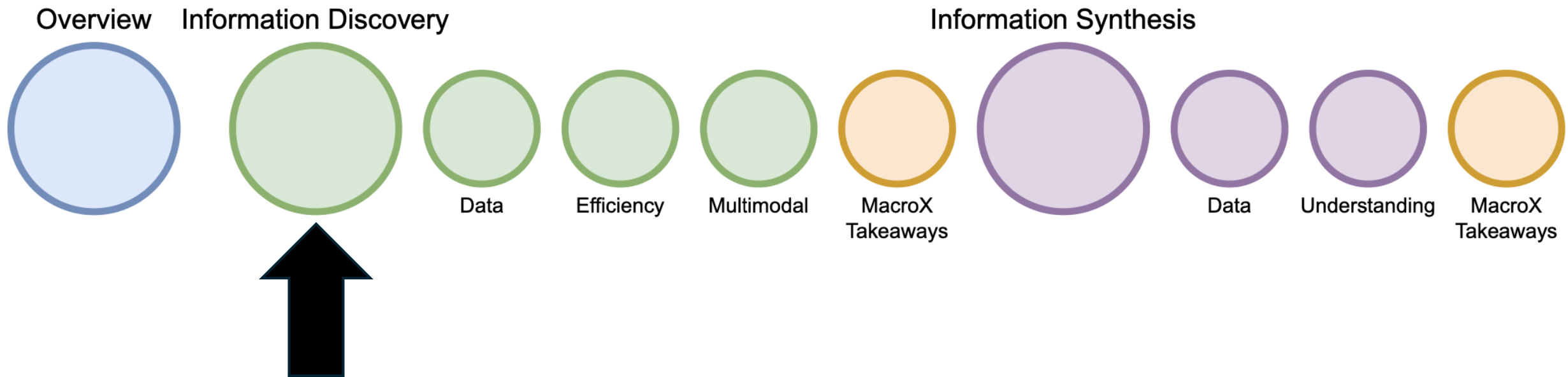
# How to write an article on demand?

Information Seeking  
(Retrieval)



# How to write an article on demand?

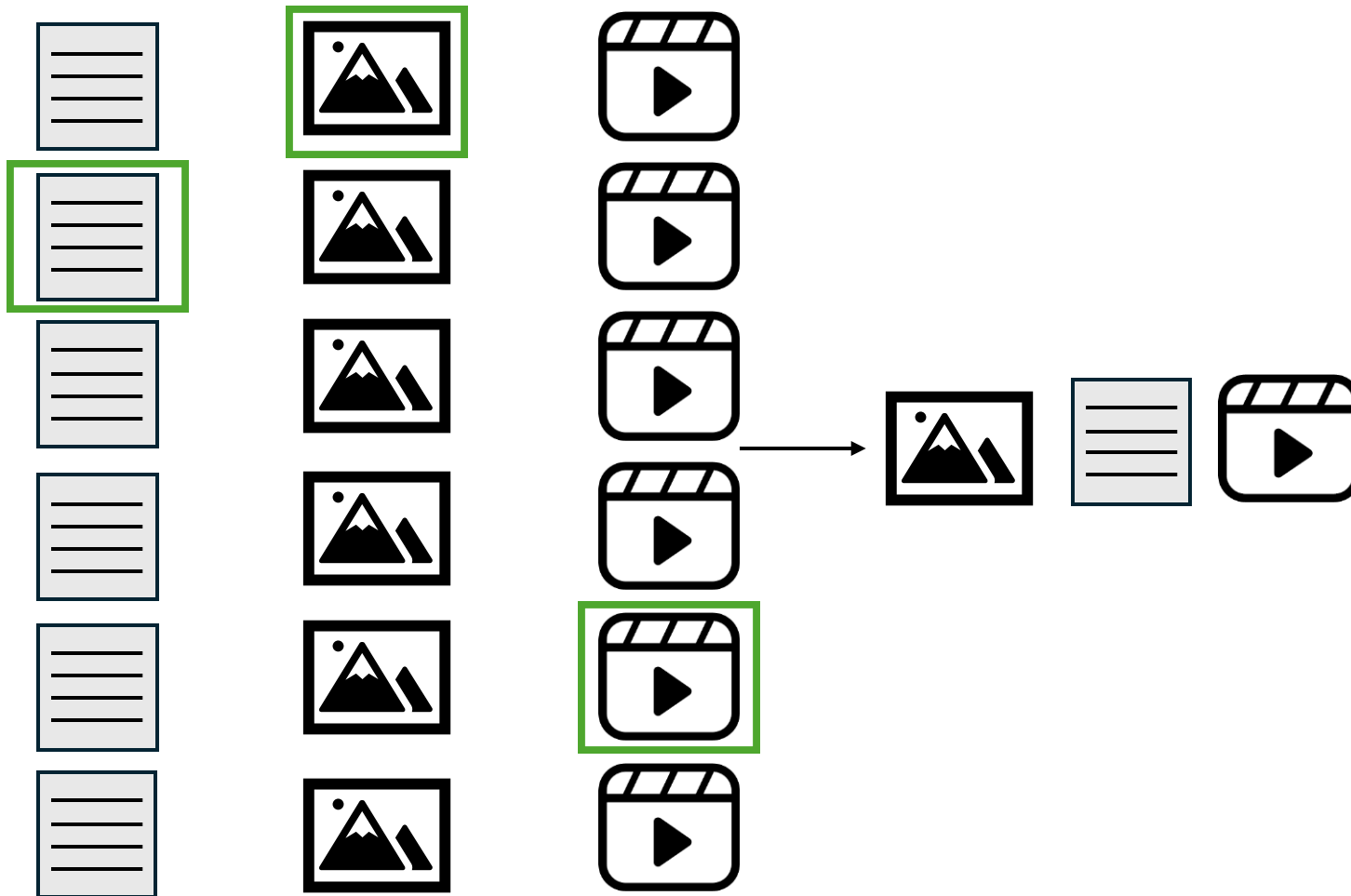




# Information Discovery

Information Retrieval

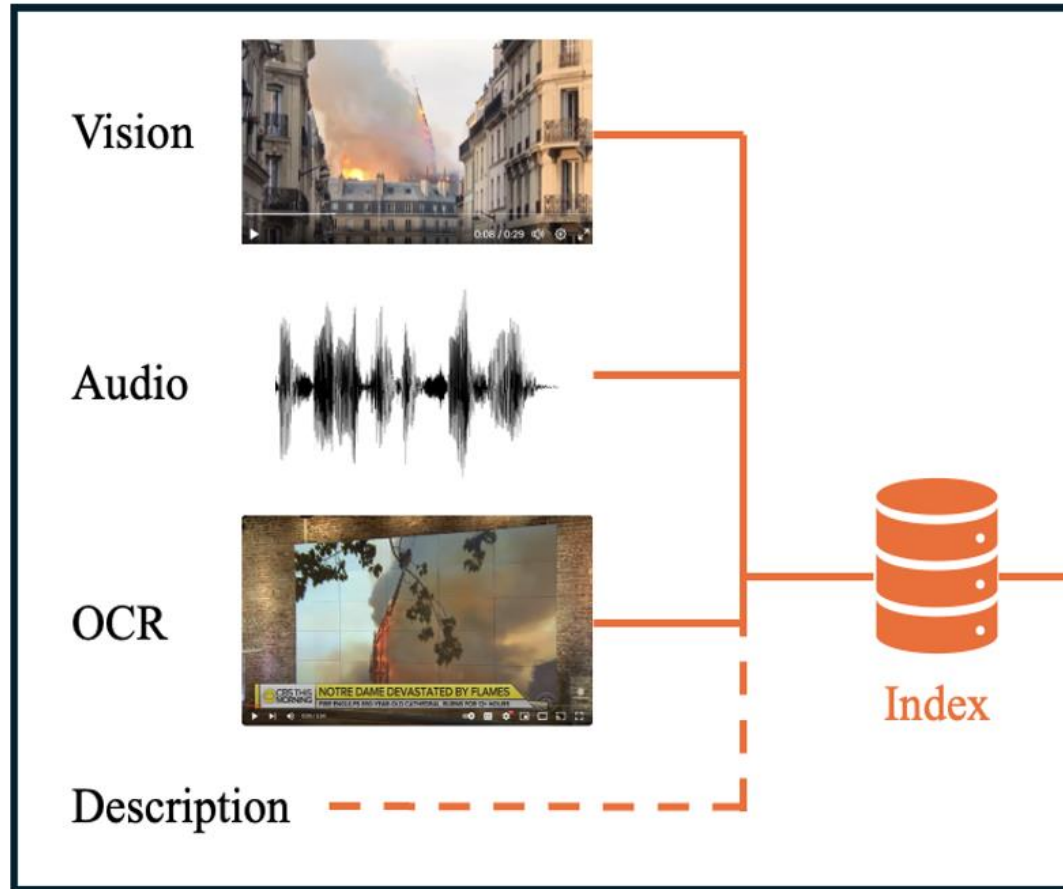
# How do I find information?



Information Seeking  
(Retrieval)

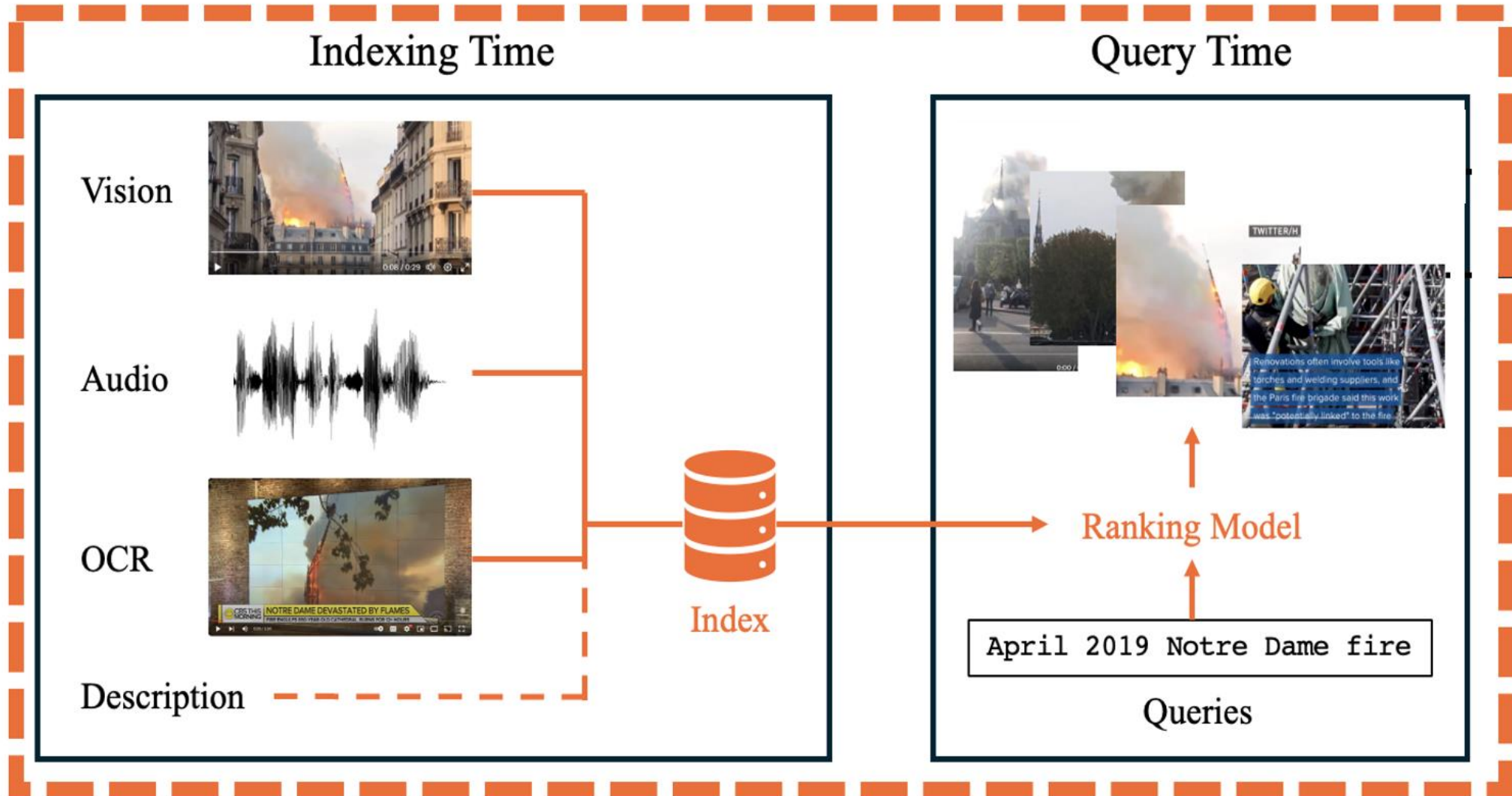
# Video Retrieval

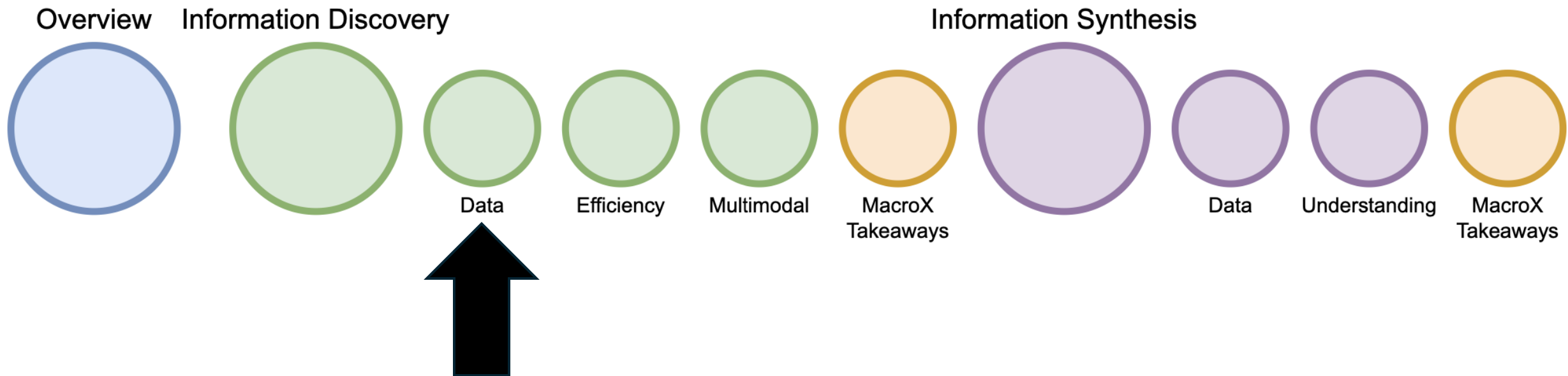
## Indexing Time





# Video Retrieval





# The Data for Real World Retrieval

MultiVENT 1.0, Sanders et al

MultiVENT 2.0, Kriz et al (incl Martin)

# MultiVENT

Emergencies



Politics



Social Events



Science & Business



1. Sanders, 1.0
2. Kriz, 2.0

# MultiVENT

Emergencies



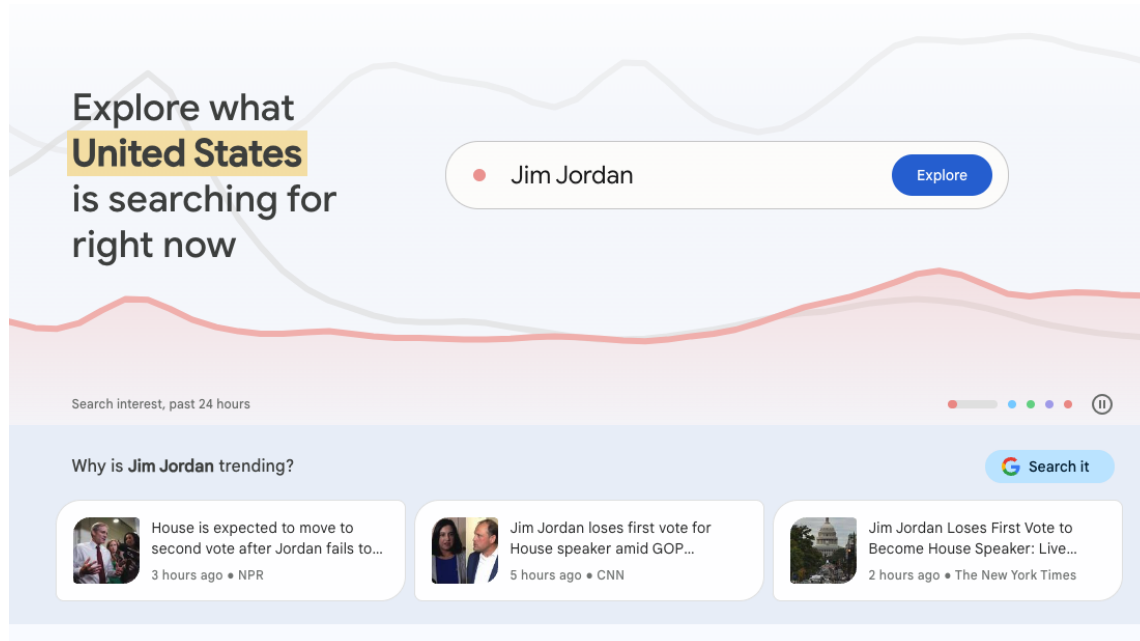
Politics



Social Events



Science & Business





# MultiVENT

## Emergencies



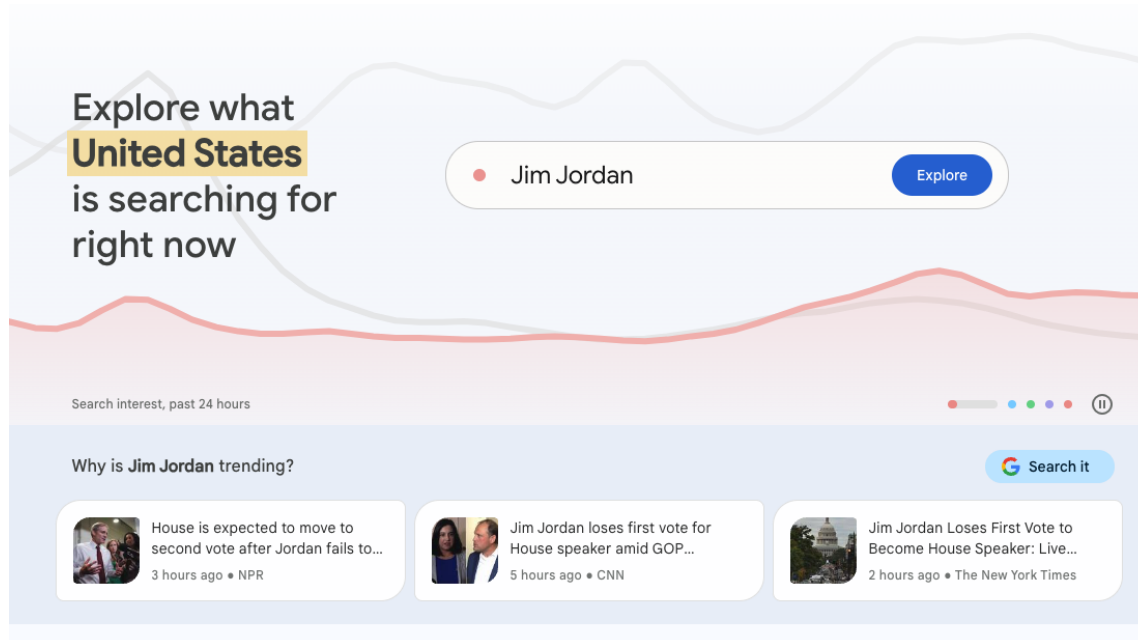
## Politics



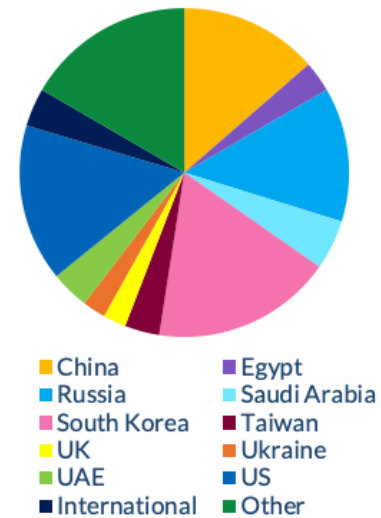
## Social Events



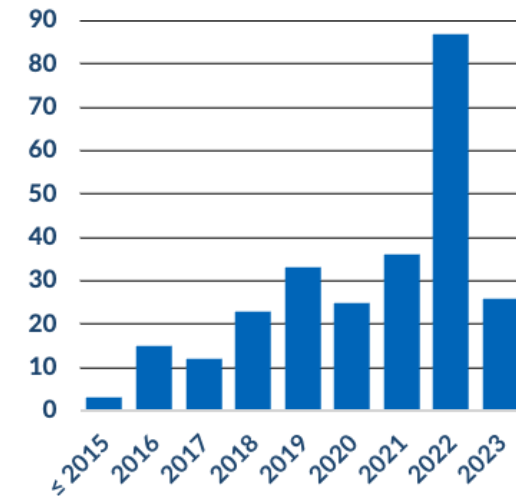
## Science & Business



## CURRENT EVENT LOCATIONS



## CURRENT EVENTS BY YEAR





# MultiVENT

## Example Topics:

- Elections (US Presidential/Senate, Korean, Ukrainian)
  - Disasters (Wildfires, earthquakes, typhoons)
- International conflicts (Border skirmishes, military mobilize)
  - Other msc (Trade wars, Brexit, Tariffs)

# MultiVENT – Video Types

Easy

Hard



Professional

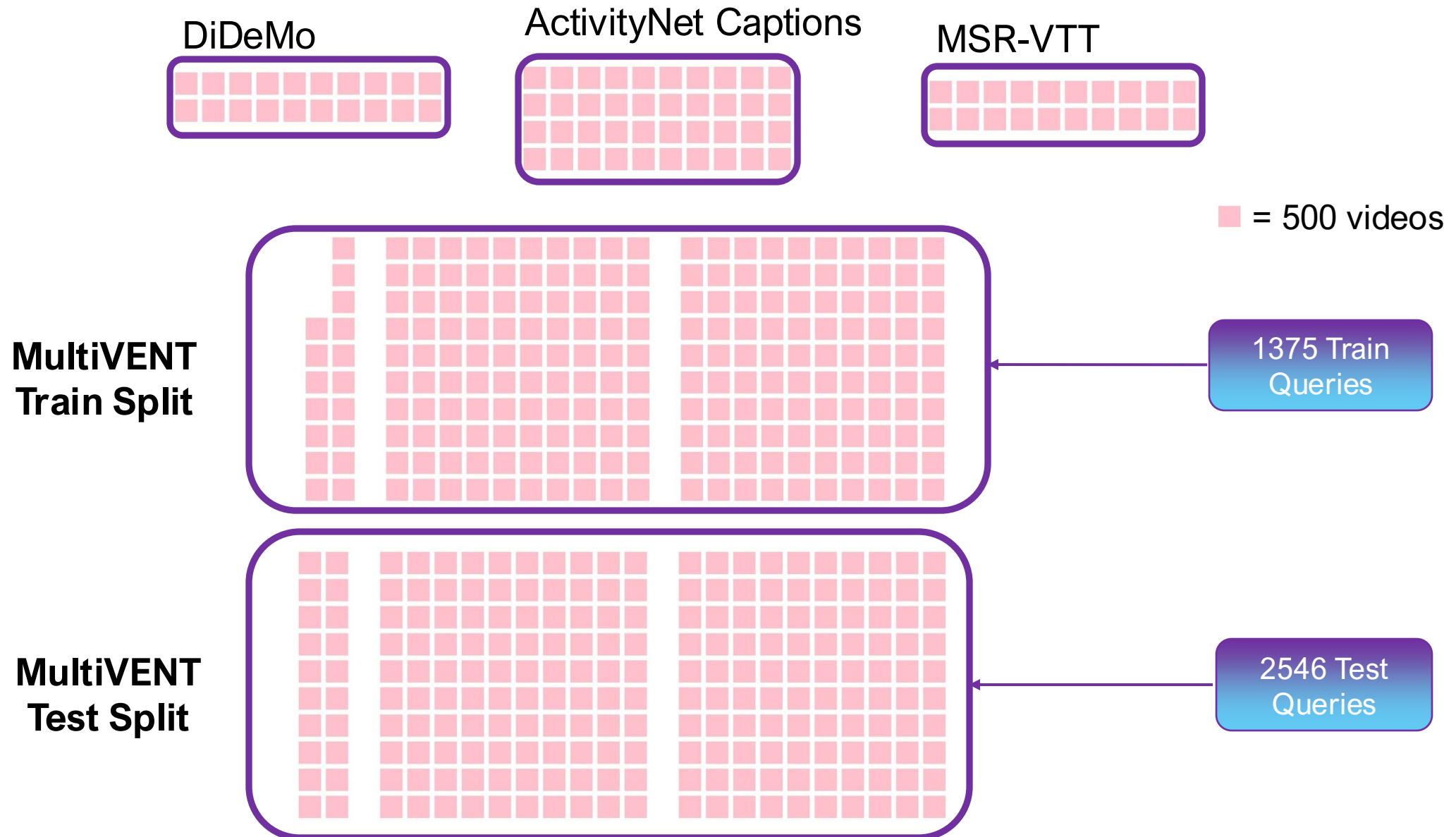


Edited

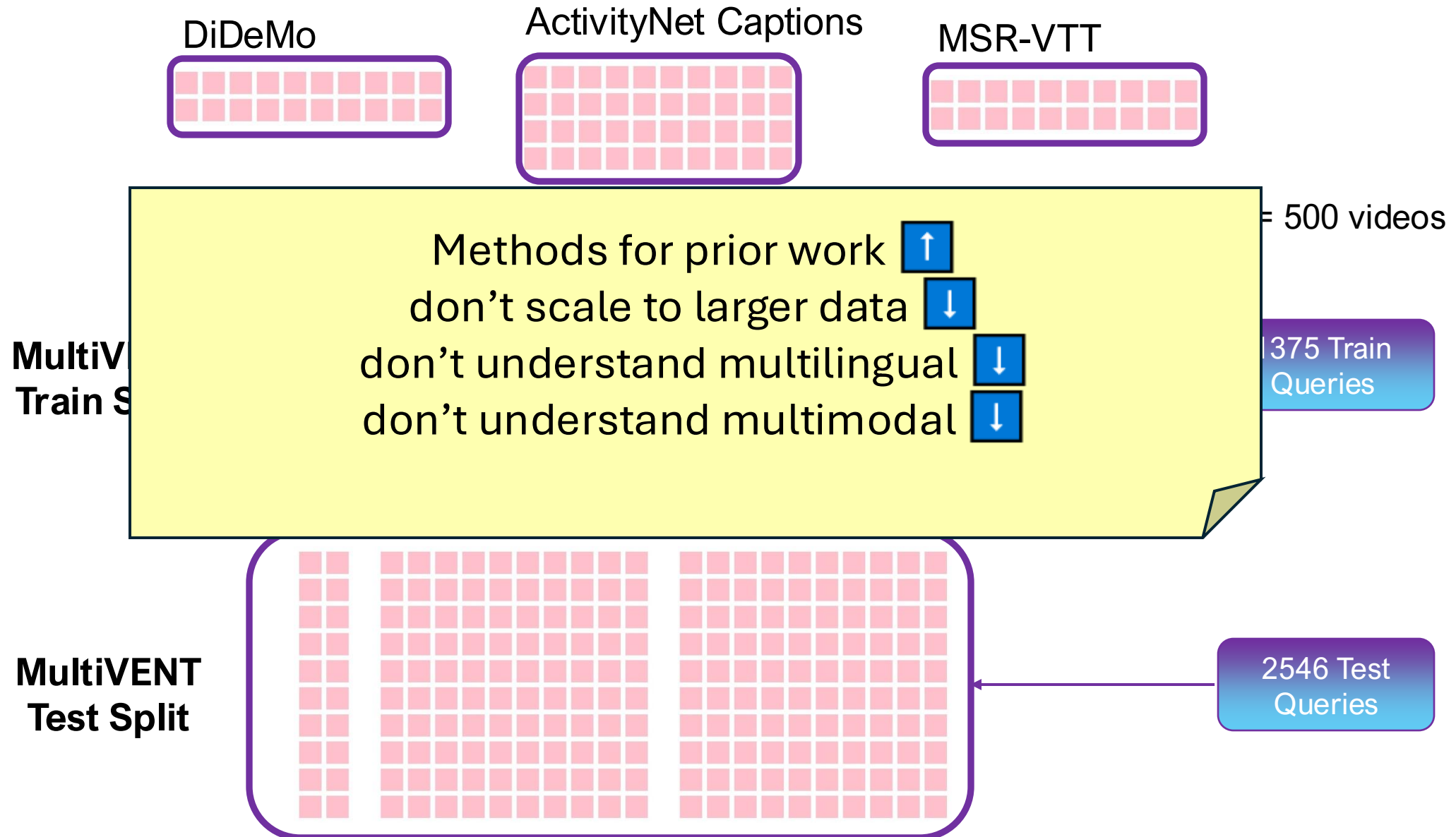


Raw Footage

# MultiVENT Size Comparison

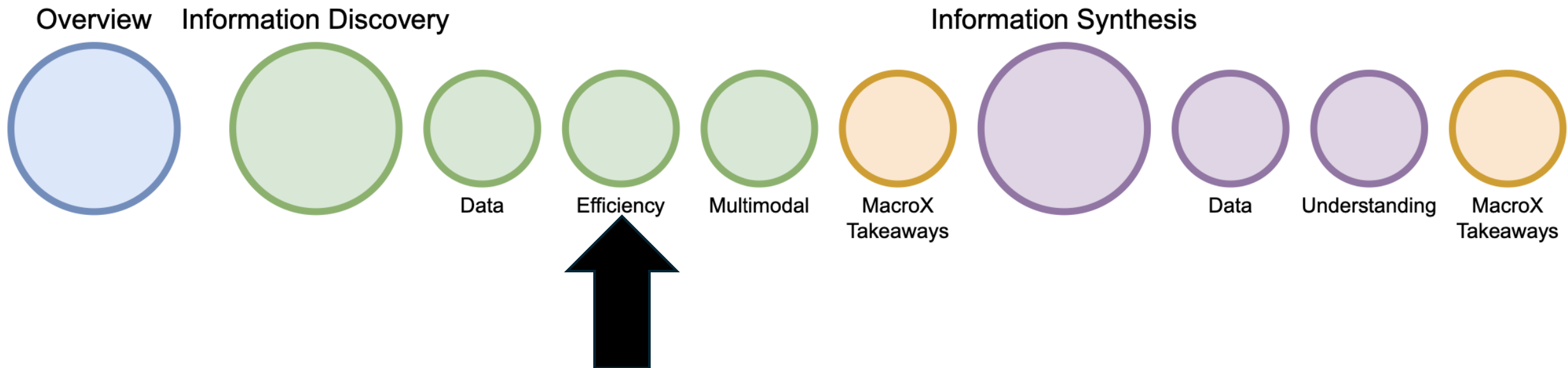


# MultiVENT Size Comparison



# Current Limitations

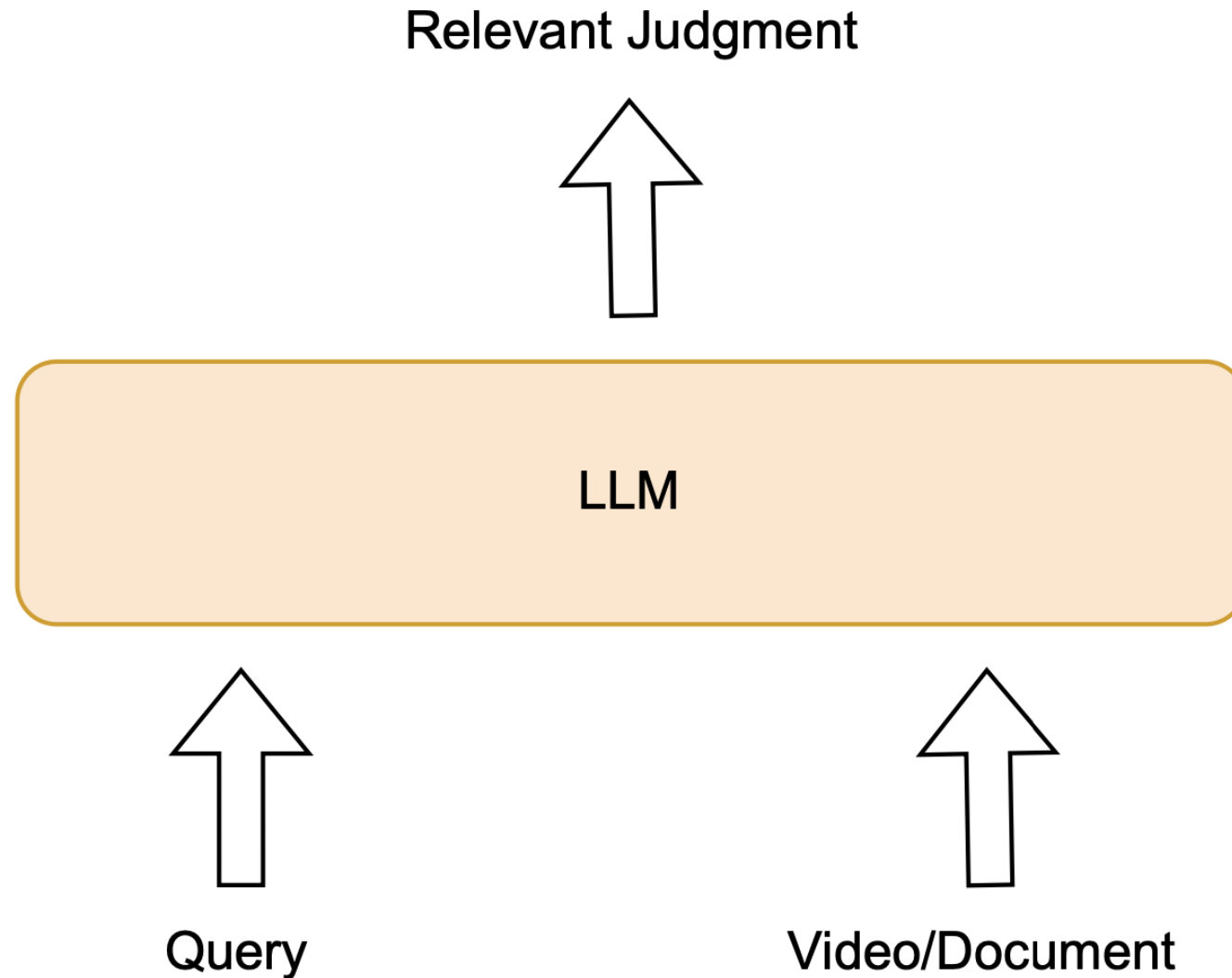
- Retrieval is not efficient
  - State of the art VideoLLMs are expensive to run on 109K+ videos
- Retrieval doesn't use multiple modalities
  - Most methods are vision only



# Efficient Video Retrieval

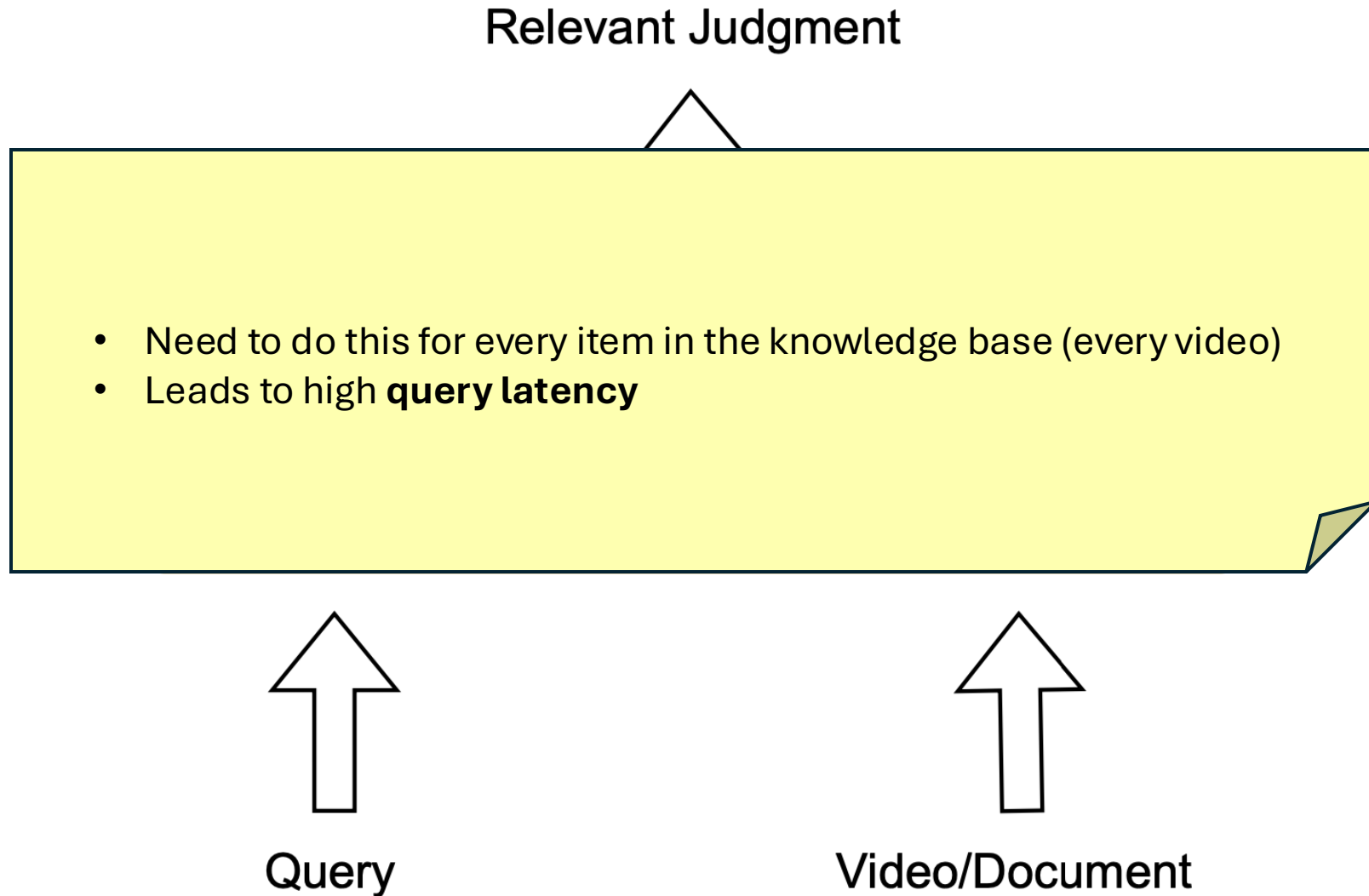
Video-ColBERT, Reddy\* & Martin\* et al

# Why I can't just use an LLM? (cross-encoder)

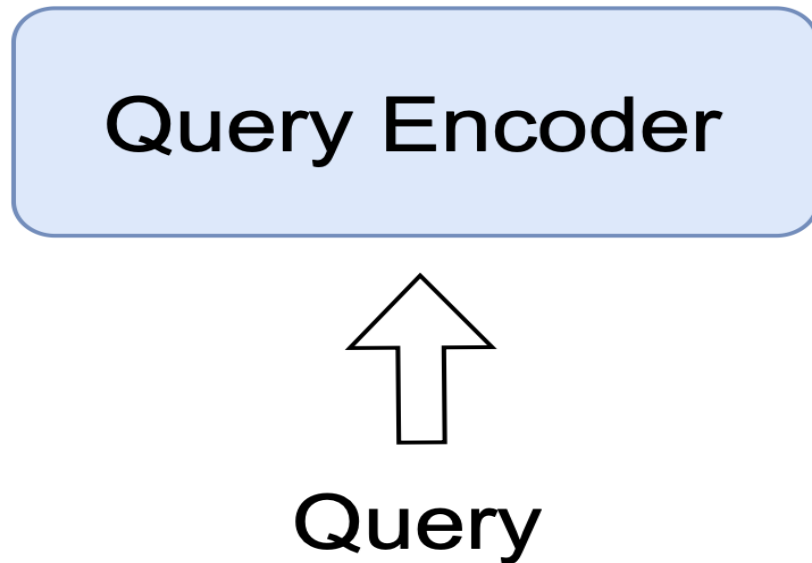




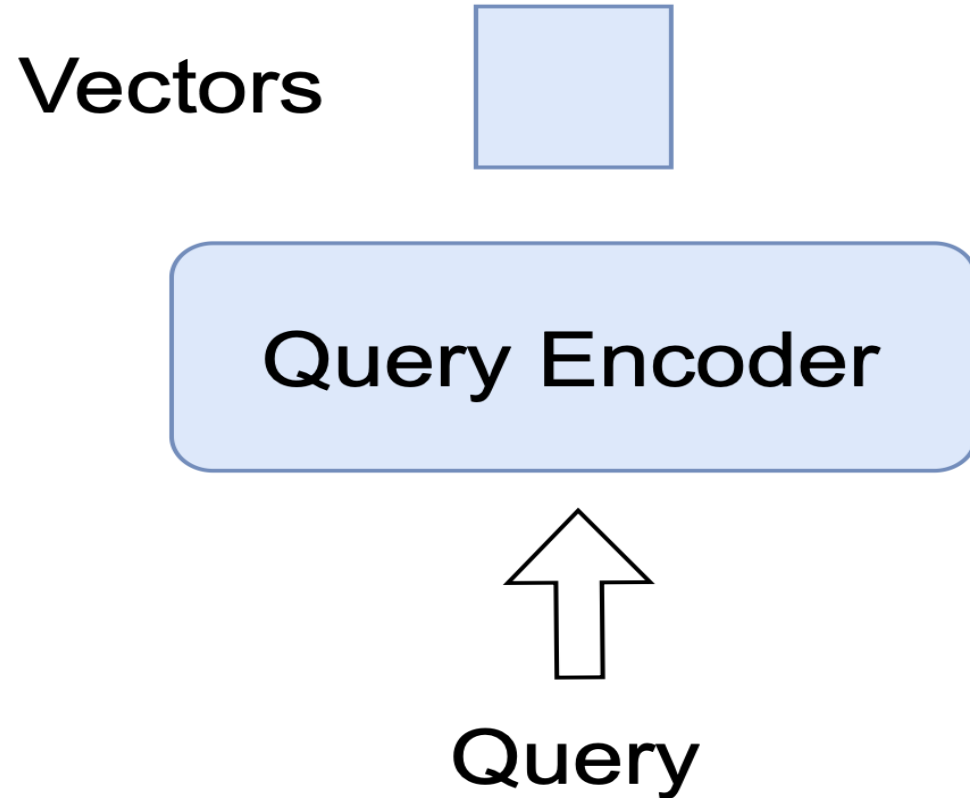
# Why I can't just use an LLM? (cross-encoder)



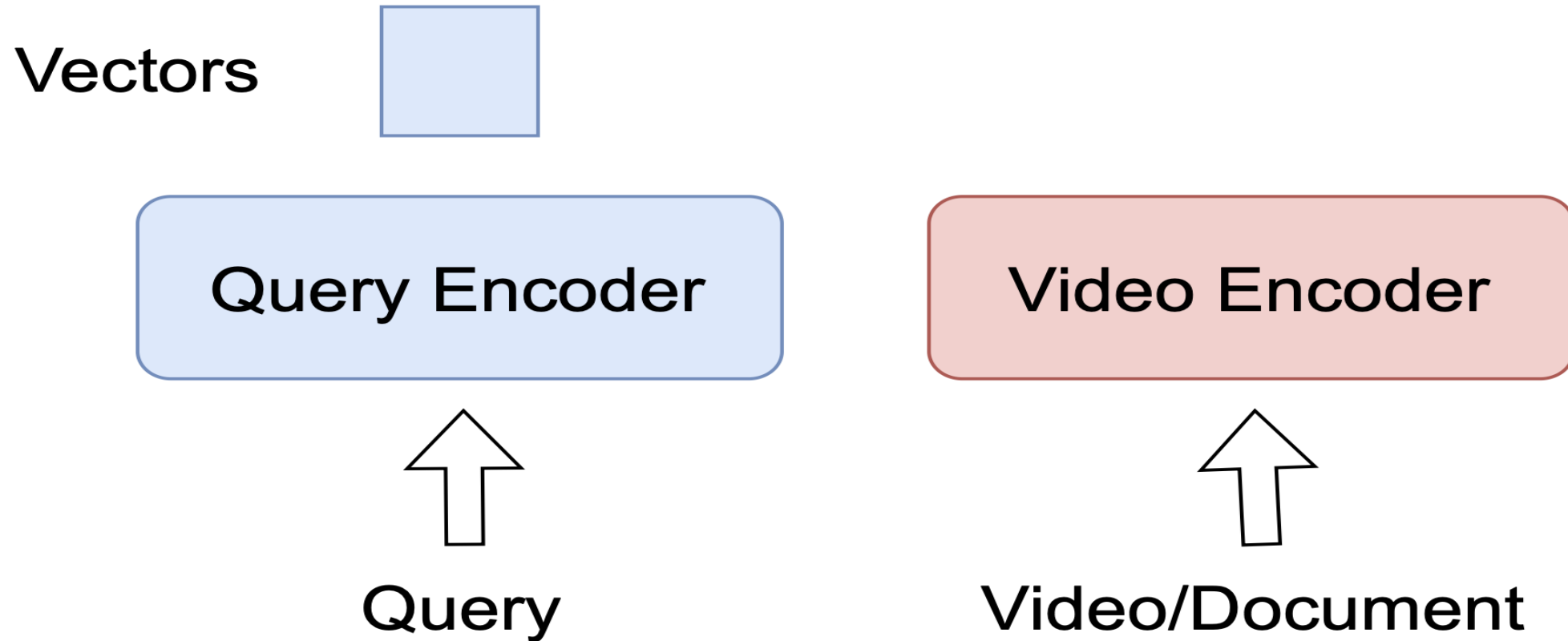
# Single Vector Methods



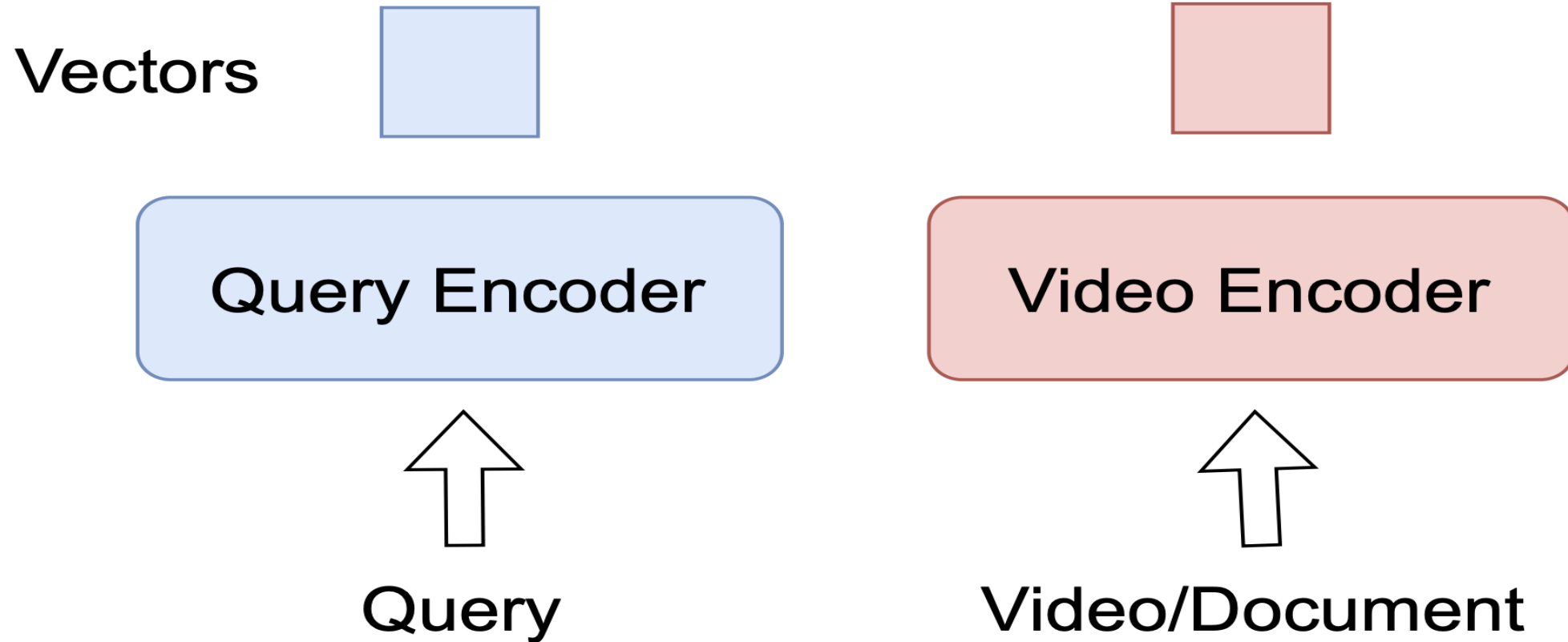
# Single Vector Methods



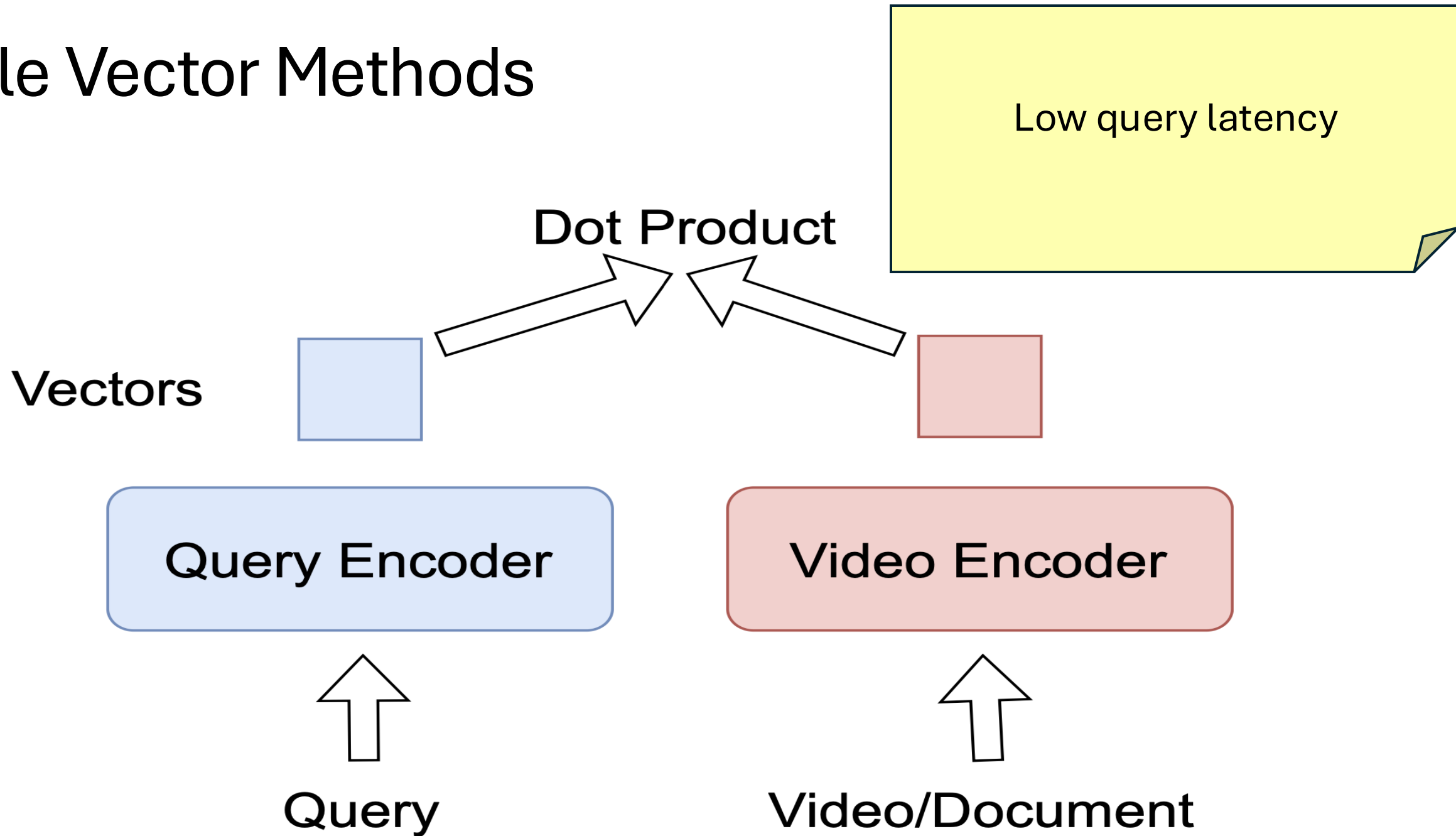
# Single Vector Methods



# Single Vector Methods

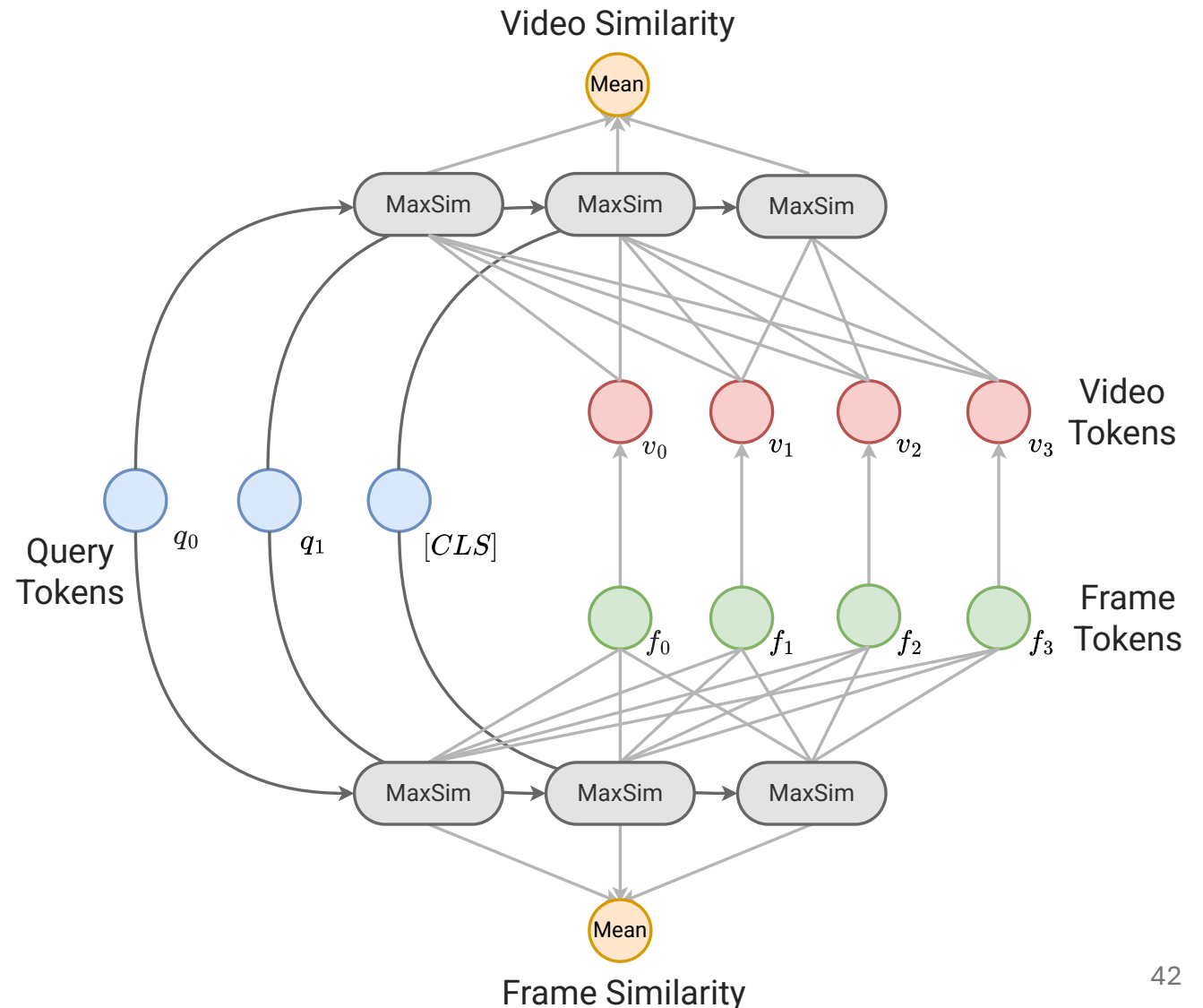


# Single Vector Methods

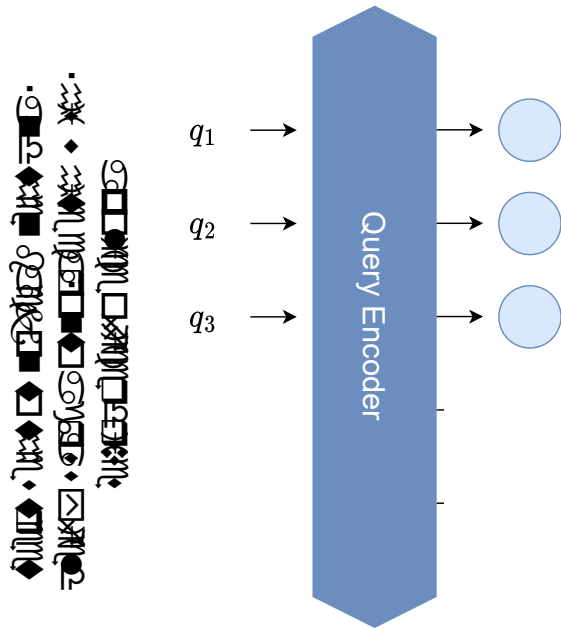


# Video-ColBERT – Efficient Video Retrieval

- Can we do video retrieval similar to SOTA text? (multi-vector sim)
- What will be the performance trade off compared to expensive methods?
- Will it scale to 100K+ videos?

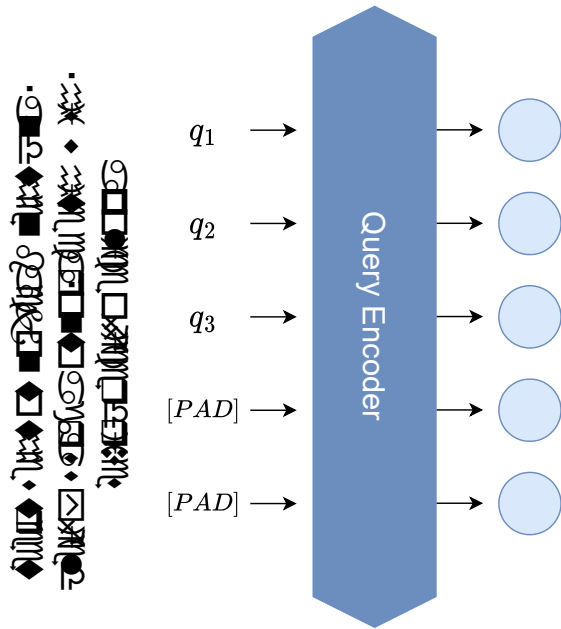


# Video-ColBERT – Efficient Video Retrieval

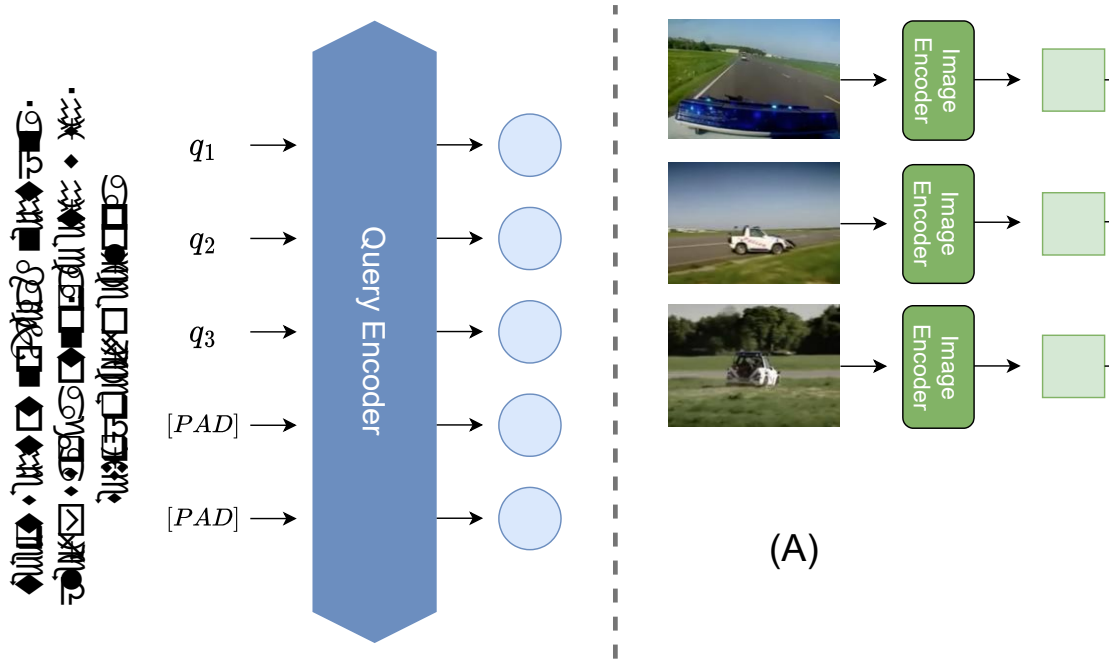




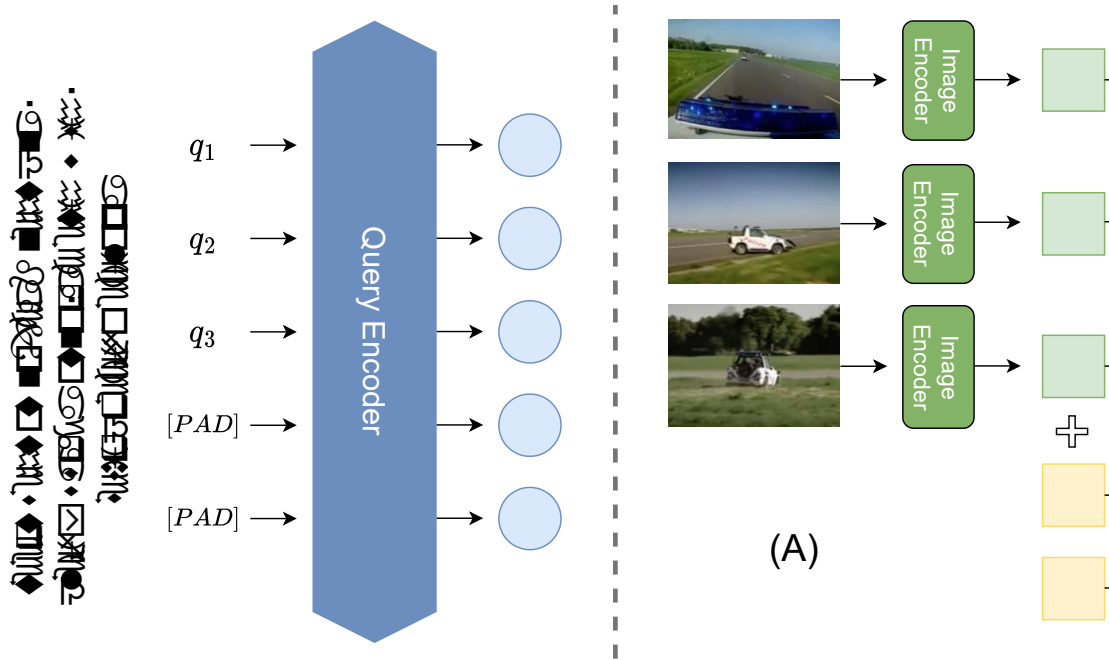
# Video-ColBERT – Efficient Video Retrieval



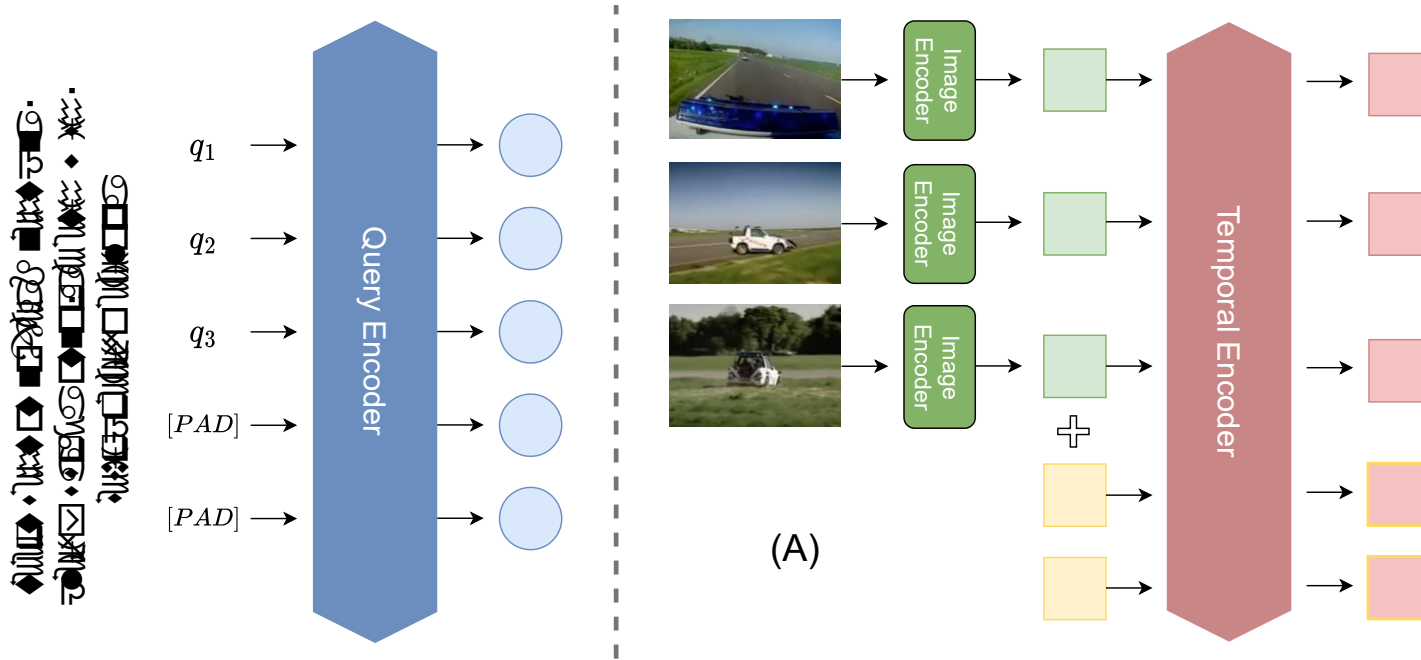
# Video-ColBERT – Efficient Video Retrieval



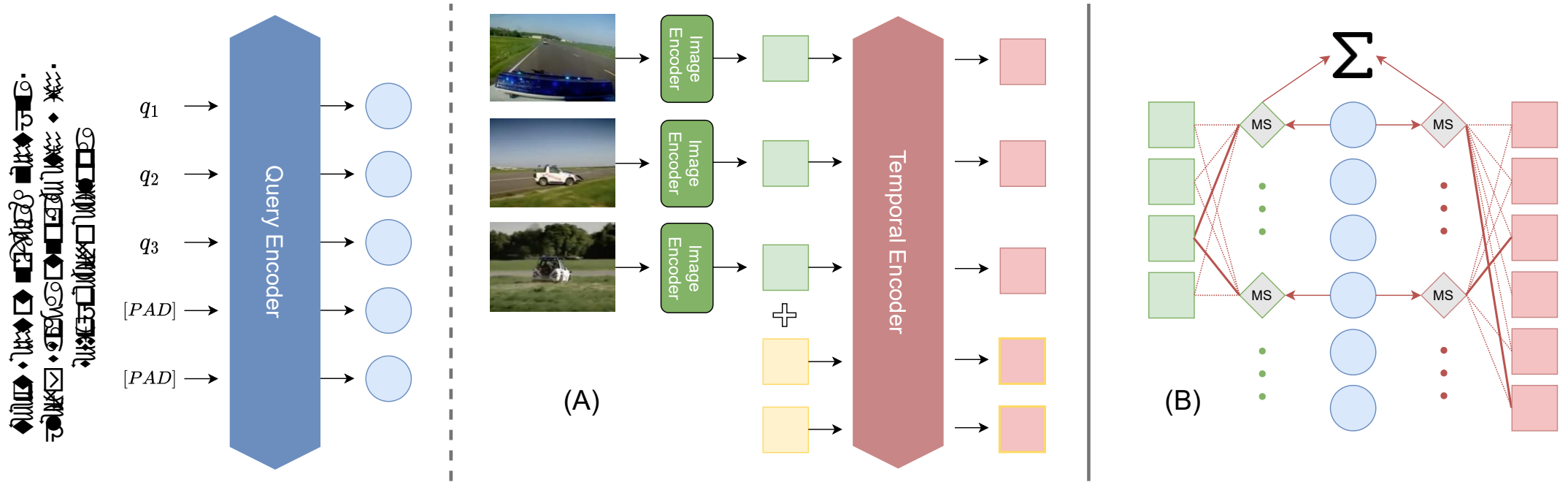
# Video-ColBERT – Efficient Video Retrieval

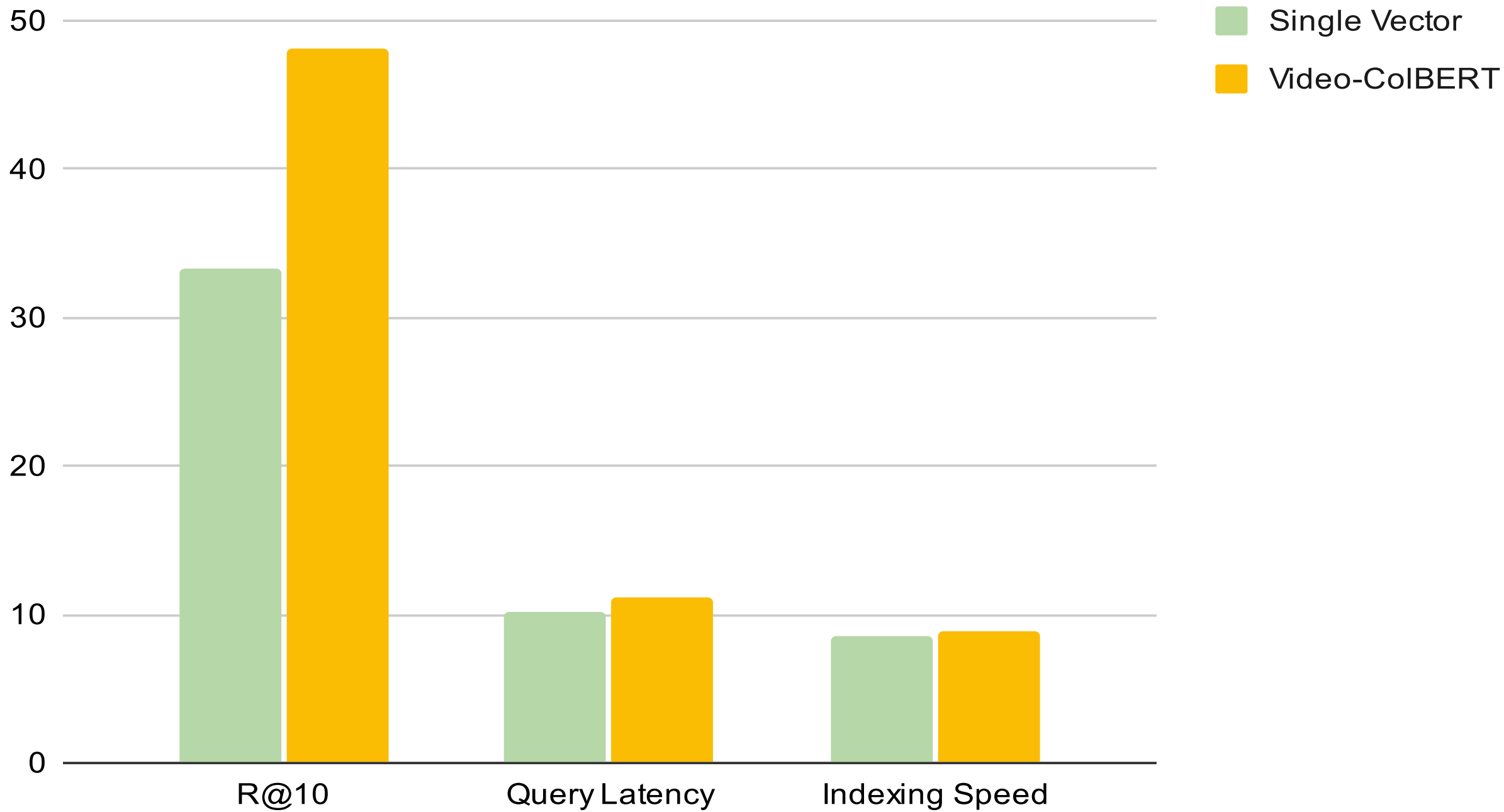


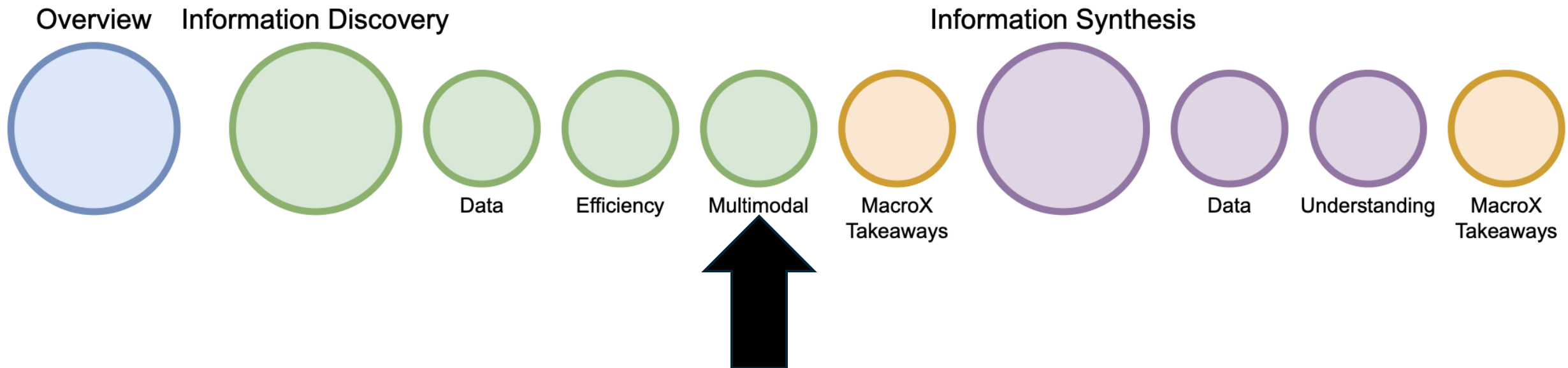
# Video-ColBERT – Efficient Video Retrieval



# Video-ColBERT – Efficient Video Retrieval



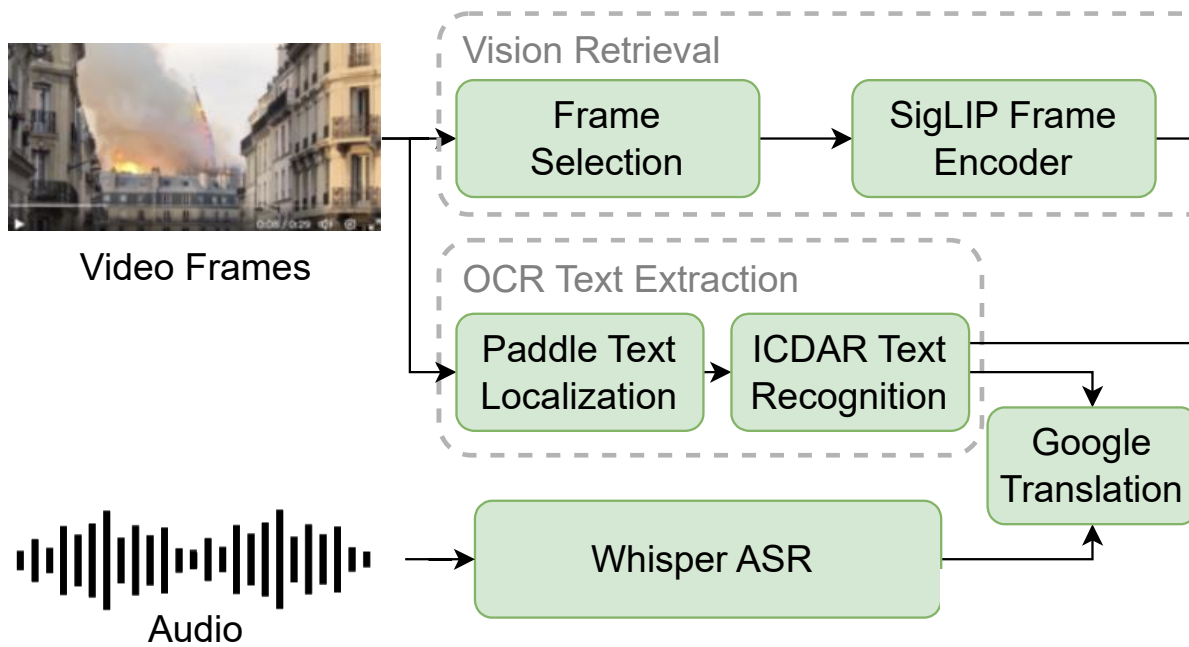




# Building Multimodal Indexes

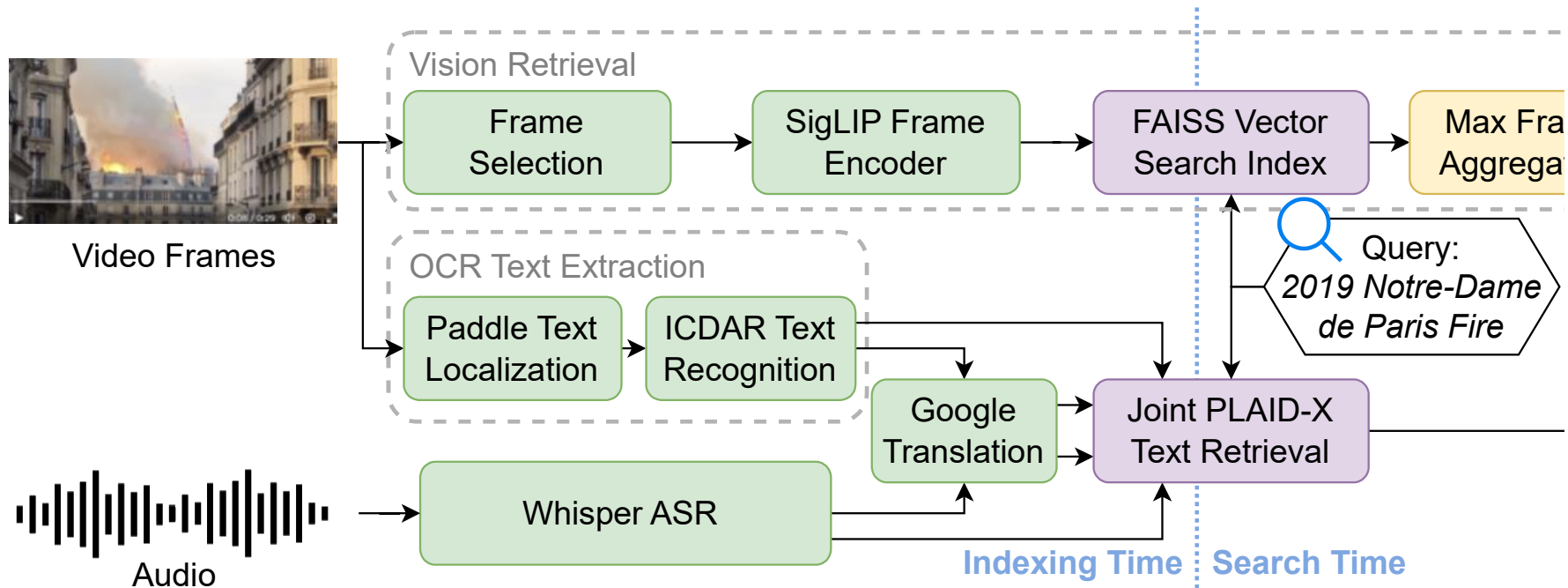
MMMORRF, Sameul et al (incl Martin)

# Multimodal Reciprocal Rank Fusion

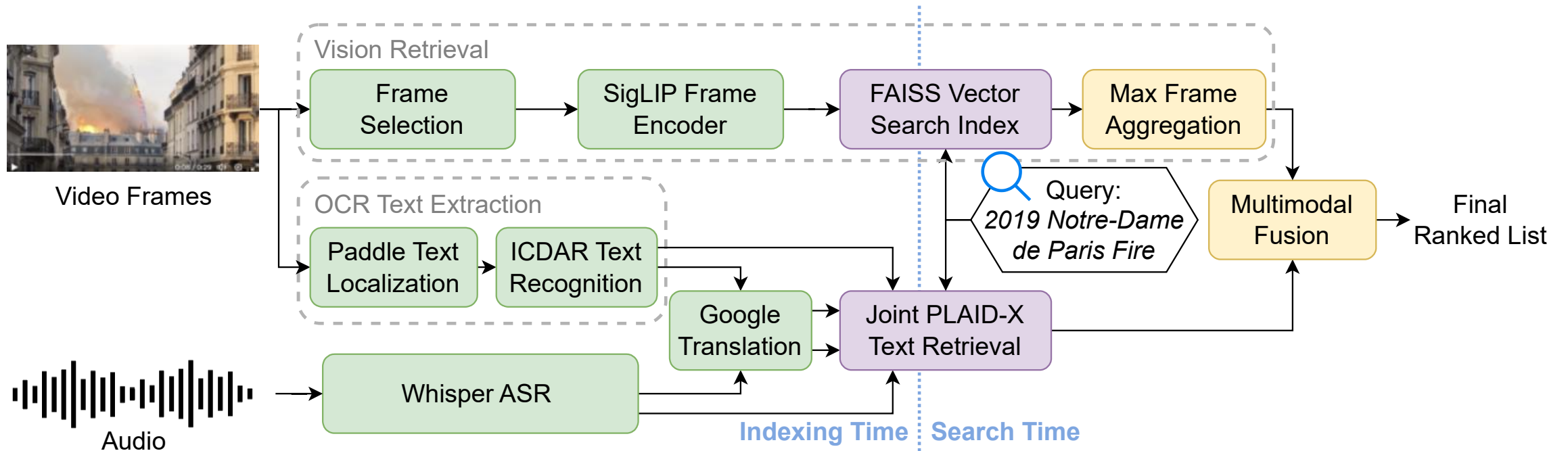




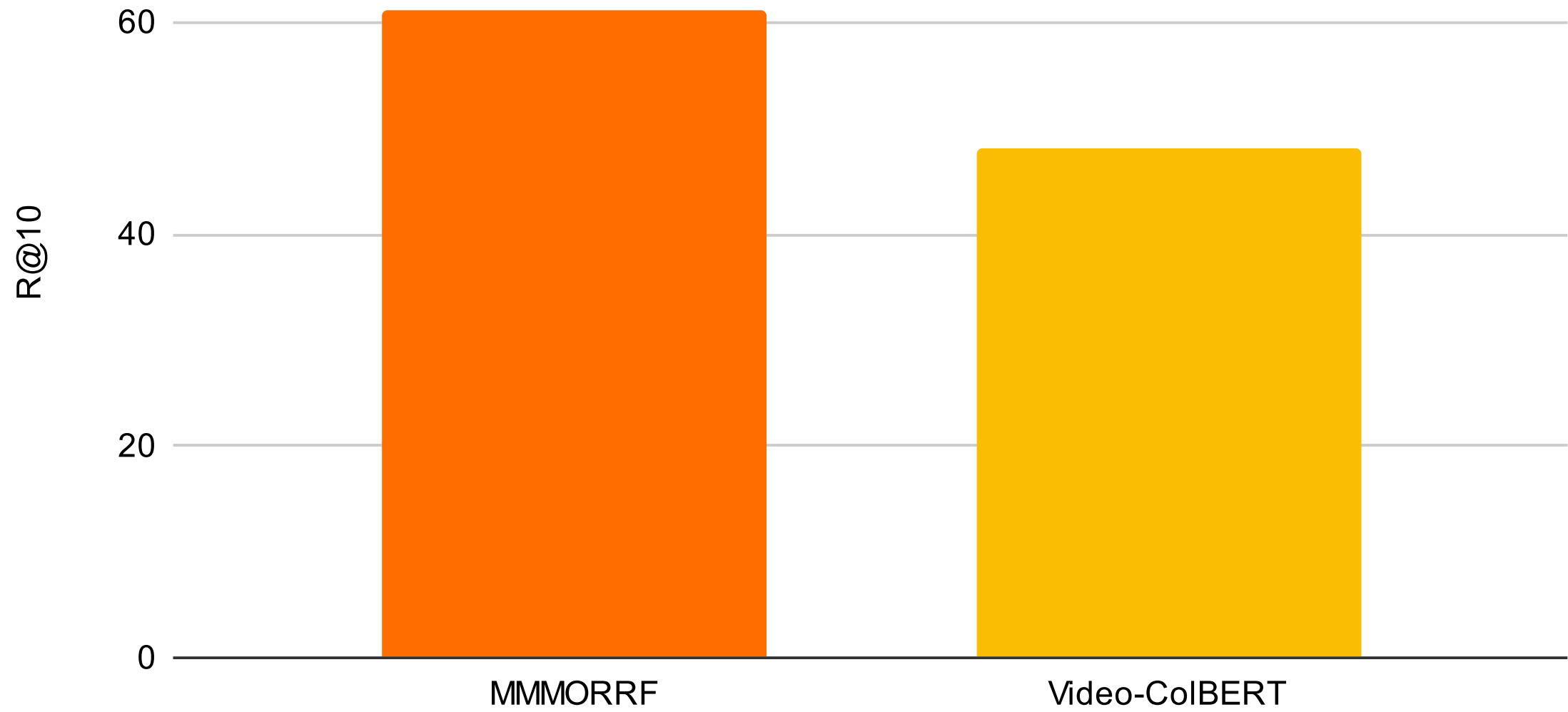
# Multimodal Reciprocal Rank Fusion

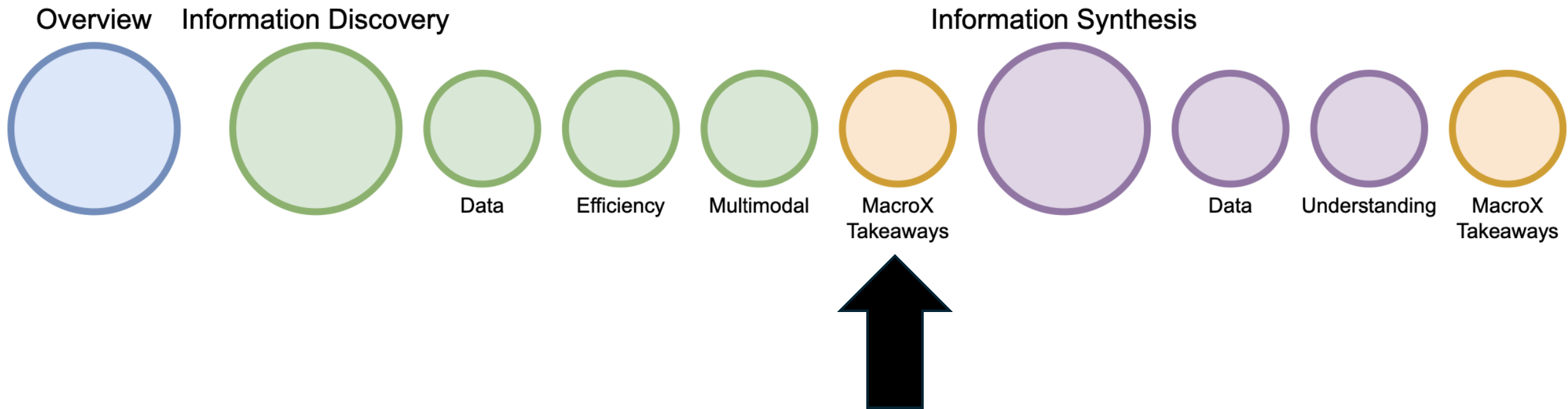


# Multimodal Reciprocal Rank Fusion



Model	Query Latency	Indexing Speed
MMMORRF	300	37000
Video-ColBERT	11.1	8.9





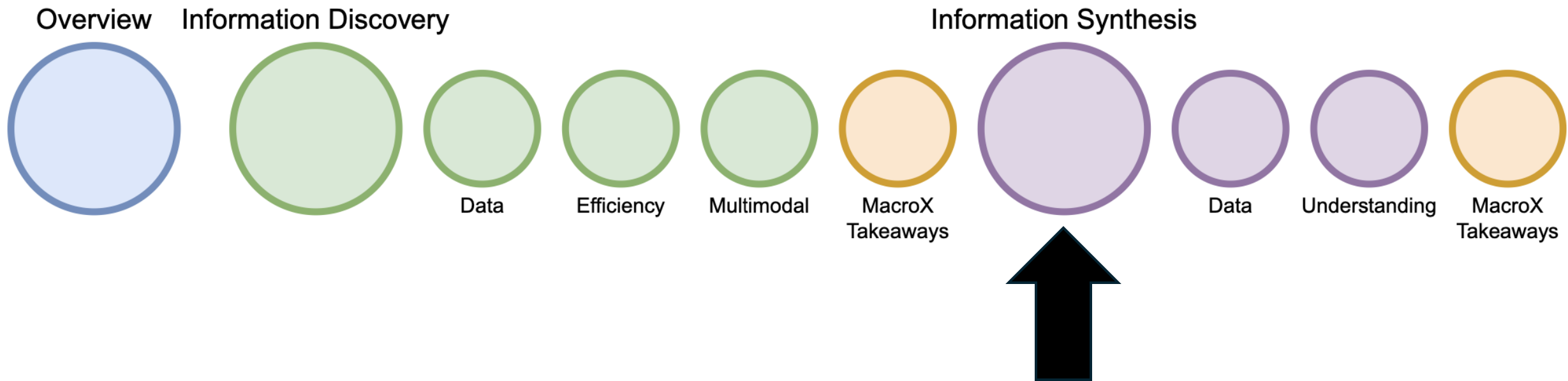
# MacroX Takeaway

# Scanning for unfolding events

- Don't know you're looking for something until after it happens
- Adapt lightweight encoders to filter the noise or cluster videos (Video-ColBERT)
- Use reasoning based rerankers to highlight relevant trends (Rank-1)

# Analysts finding information

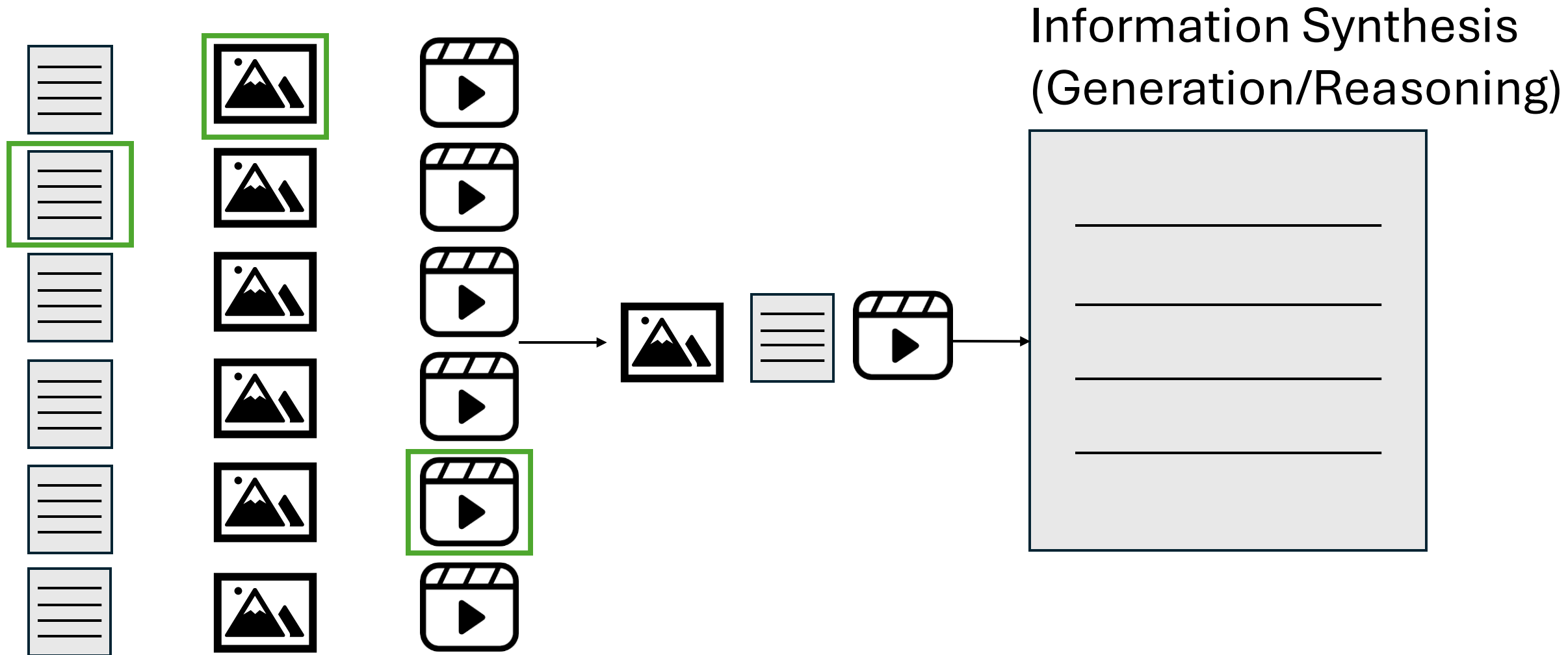
- Everyone can benefit from better search, but multimodal search is lacking
- Why leave out a huge information source (MMMORRF)



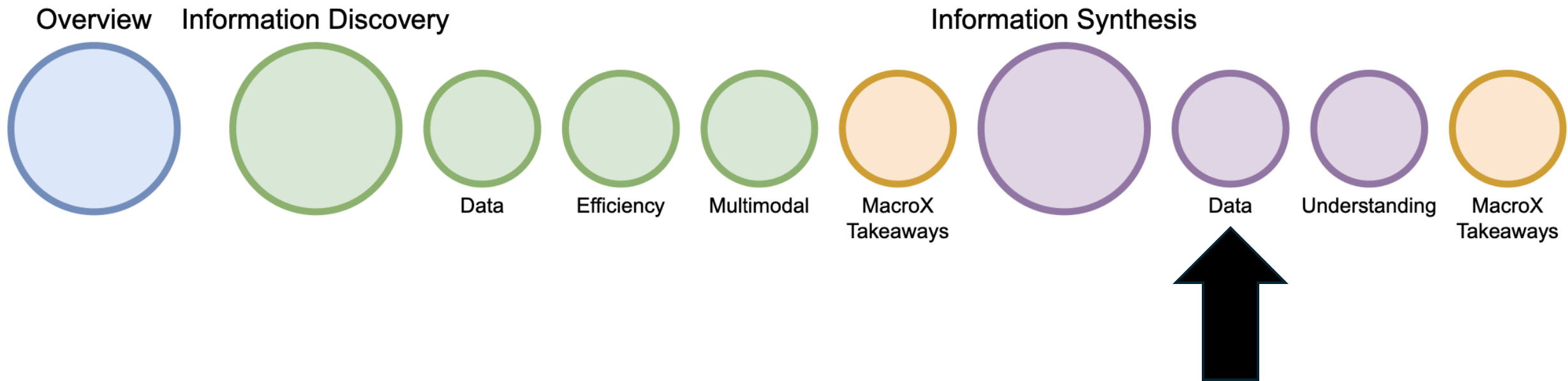
# Information Synthesis

Article Generation, Nowcasting, Reasoning

# What do I do with the information I found?



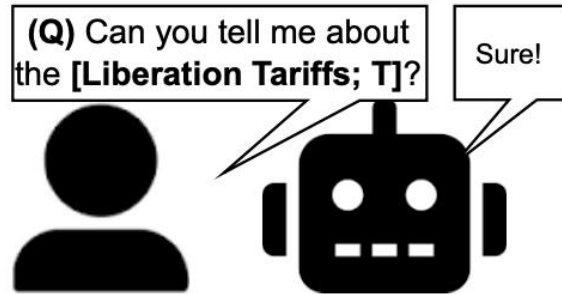




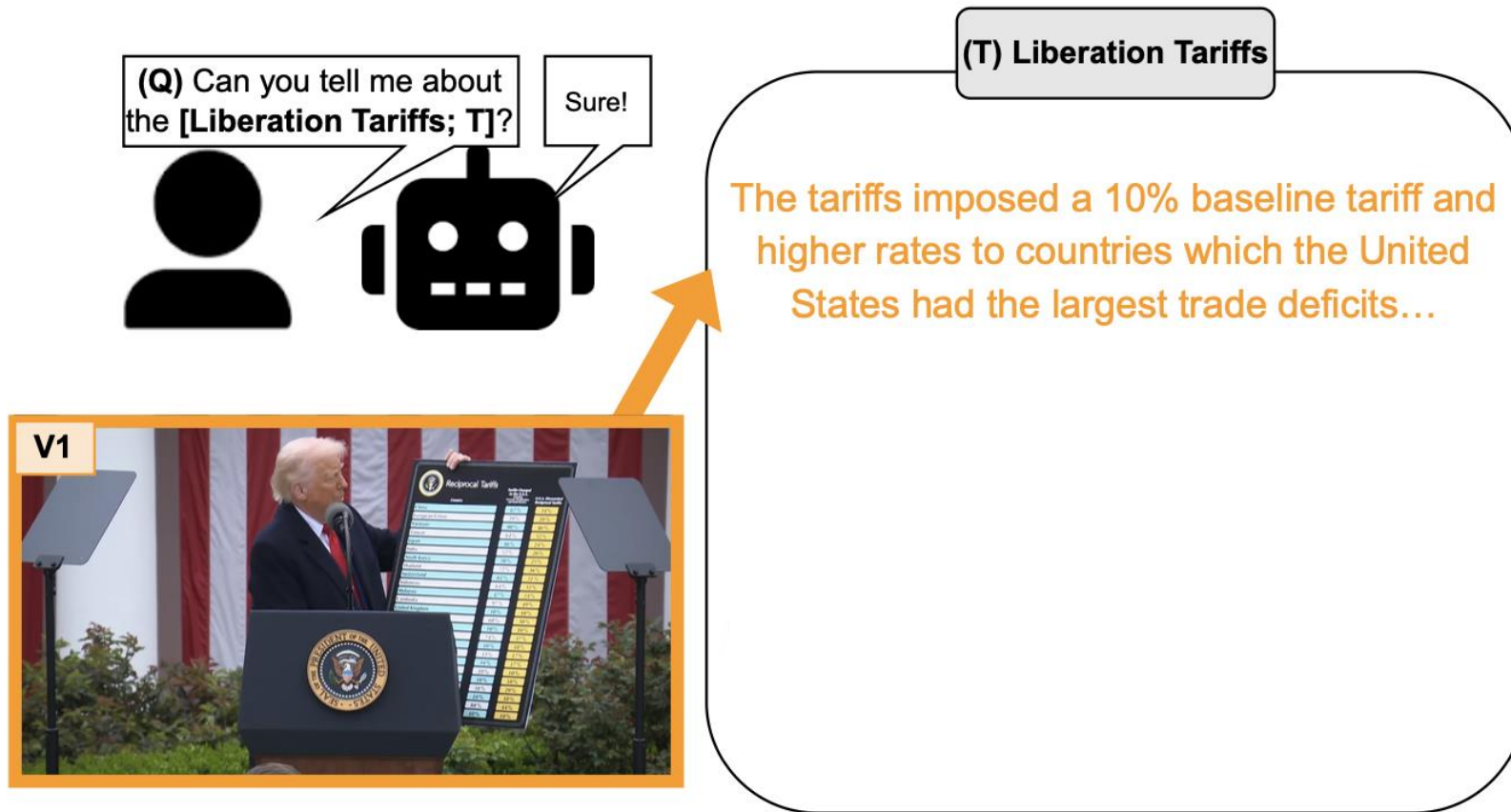
# Data for Multimodal Reasoning

WikiVideo, Martin et al

# WikiVideo: Article Generation from Multiple Videos



# WikiVideo: Article Generation from Multiple Videos



# WikiVideo: Article Generation from Multiple Videos

(Q) Can you tell me about the [Liberation Tariffs; T]?

Sure!

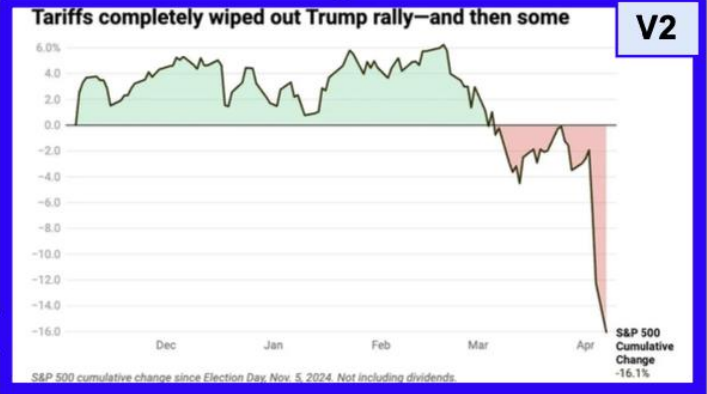


## (T) Liberation Tariffs

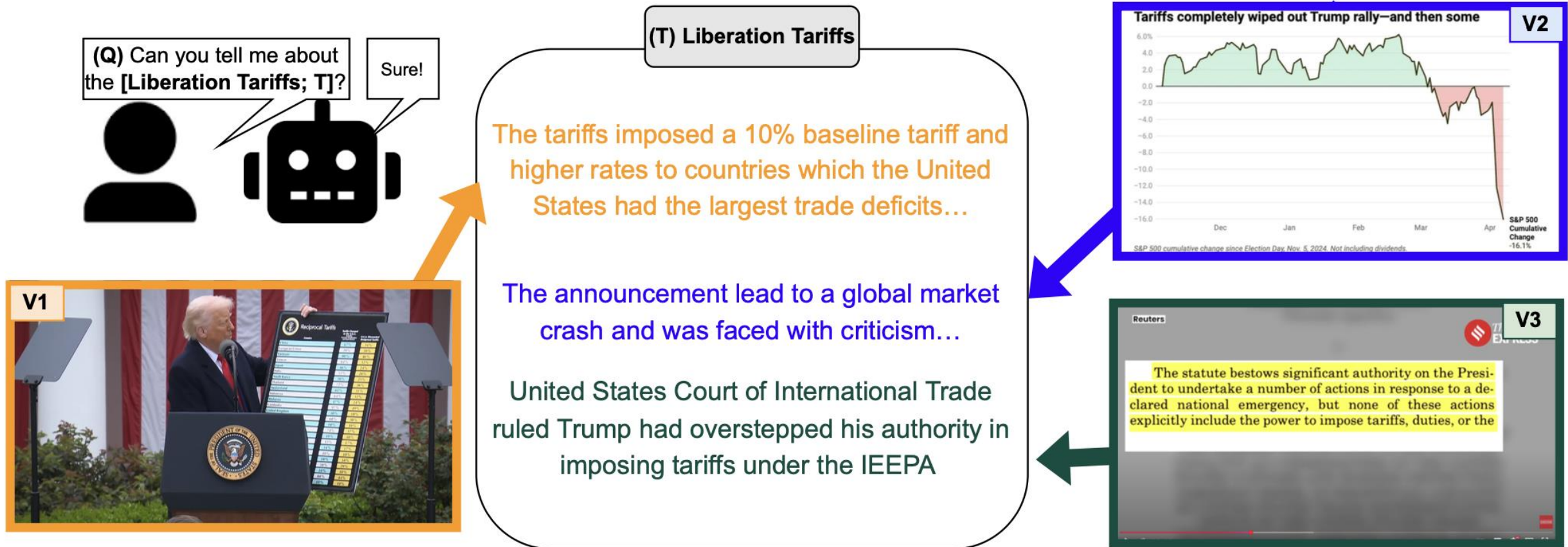
The tariffs imposed a 10% baseline tariff and higher rates to countries which the United States had the largest trade deficits...

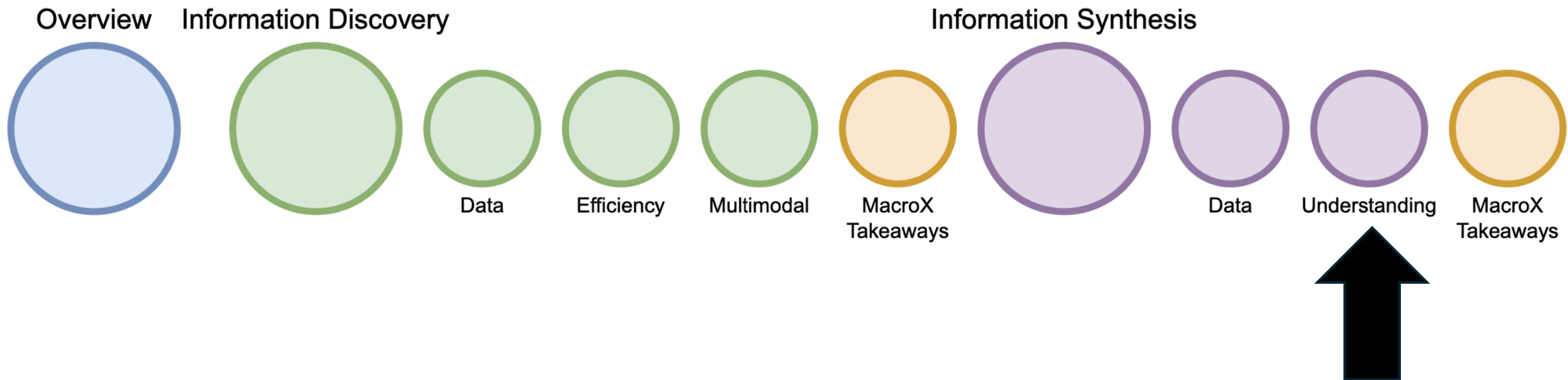
The announcement lead to a global market crash and was faced with criticism...

V1



# WikiVideo: Article Generation from Multiple Videos





# Improving Understanding

WikiVideo, Martin et al



# WikiVideo: VideoLLM Performance



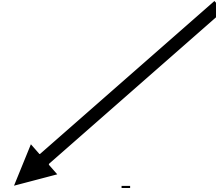
## QwenVL2.5 Output

The video shows a man in a dark suit and red tie standing at a podium with an American flag in the background. He appears to be speaking into a microphone, likely delivering a speech or address. The setting is outdoors, as indicated by the greenery visible behind him. At one point, he gestures with his hands while holding some papers, suggesting he is referencing notes or emphasizing points during his speech. The date "April 2, 2025" is displayed in the top left corner of the screen, indicating when the event is taking place. The logo "N" in the top right corner suggests that the video may be from a news source.

# WikiVideo: VideoLLM Performance



Recognizes  
key entity



## QwenVL2.5 Output

The video shows a man in a dark suit and red tie standing at a podium with an American flag in the background. He appears to be speaking into a microphone, likely delivering a speech or address. The setting is outdoors, as indicated by the greenery visible behind him. At one point, he gestures with his hands while holding some papers, suggesting he is referencing notes or emphasizing points during his speech. The date "April 2, 2025" is displayed in the top left corner of the screen, indicating when the event is taking place. The logo "N" in the top right corner suggests that the video may be from a news source.



# WikiVideo: VideoLLM Performance



Recognizes  
key event

## QwenVL2.5 Output

The video shows a man in a dark suit and red tie standing at a podium with an American flag in the background. He appears to be speaking into a microphone, likely delivering a speech or address. The setting is outdoors, as indicated by the greenery visible behind him. At one point, he gestures with his hands while holding some papers, suggesting he is referencing notes or emphasizing points during his speech. The date "April 2, 2025" is displayed in the top left corner of the screen, indicating when the event is taking place. The logo "N" in the top right corner suggests that the video may be from a news source.

# WikiVideo: VideoLLM Performance



## QwenVL2.5 Output

The video shows a man in a dark suit and red tie standing at a podium with an American flag in the background. He appears to be speaking into a microphone, likely delivering a speech or address. The setting is outdoors, as indicated by the greenery visible behind him. At one point, he gestures with his hands while holding some papers, suggesting he is referencing notes or emphasizing points during his speech. The date "April 2, 2025" is displayed in the top left corner of the screen, indicating when the event is taking place. The logo "N" in the top right corner suggests that the video may be from a news source.

**Recognizes  
key object**

# WikiVideo: VideoLLM Performance

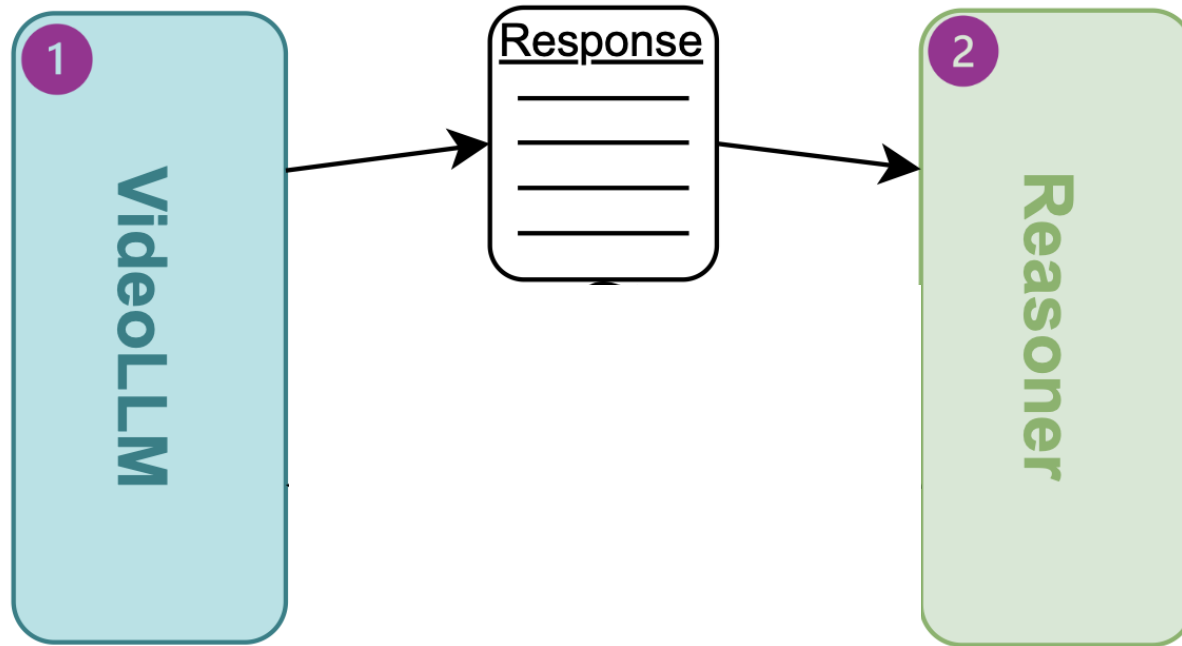


## QwenVL2.5 Output

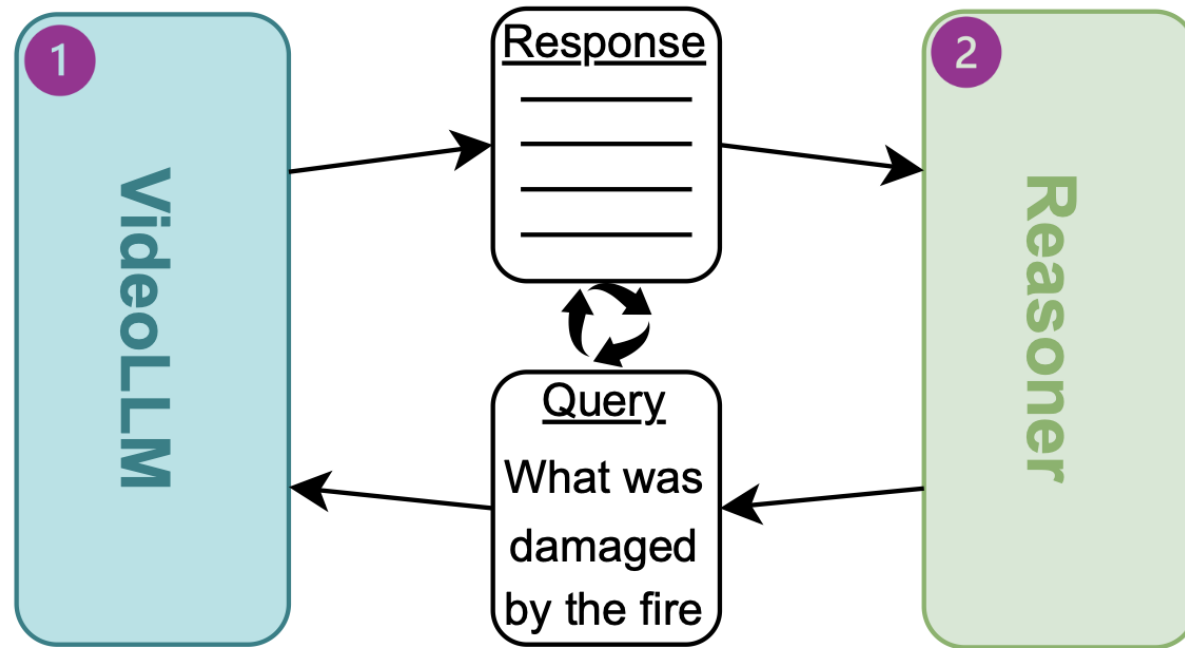
The video shows a man in a dark suit and red tie standing at a podium with an American flag in the background. He appears to be speaking into a microphone, likely delivering a speech or address. The setting is outdoors, as indicated by the greenery visible behind him. At one point, he gestures with his hands while holding some papers, suggesting he is referencing notes or emphasizing points during his speech. The date "April 2, 2025" is displayed in the top left corner of the screen, indicating when the event is taking place. The logo "N" in the top right corner suggests that the video may be from a news source.

**FAILS TO DRAW “HIGHER-LEVEL” INFERENCES**

# WikiVideo: Collaborative Article Generation

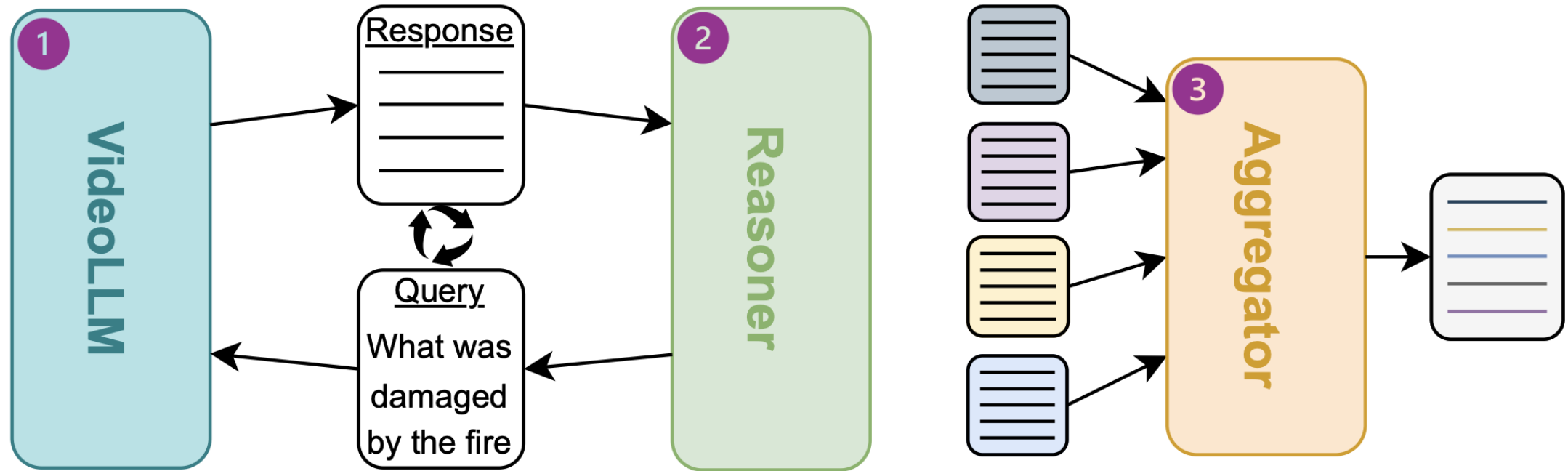


# WikiVideo: Collaborative Article Generation



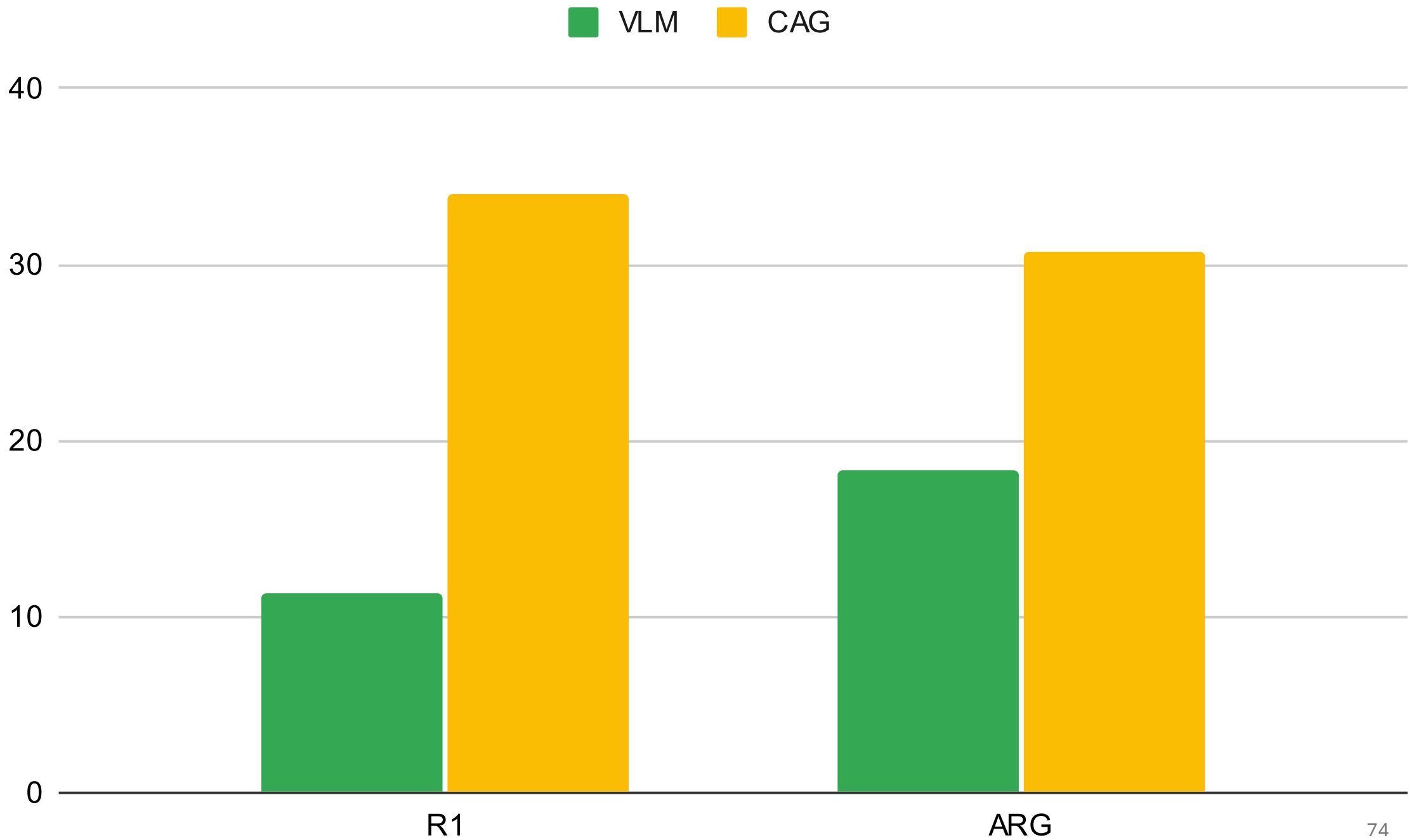
**A** **<think>** What new query can I think of to help me find more information about the Notre-Dame fire?... The video shows... Considering all this, I think a good follow-up would focus on the extent of the damage **</think>**

# WikiVideo: Collaborative Article Generation



**A** **<think>** What new query can I think of to help me find more information about the Notre-Dame fire?... The video shows... Considering all this, I think a good follow-up would focus on the extent of the damage **</think>**

**B** **<think>** This video captures information about the aftermath. No new query needed. **</think>**



# WikiVideo: VideoLLM Performance



## QwenVL2.5 Output

The video shows a man in a dark suit and red tie standing at a podium with an American flag in the background. He appears to be speaking into a microphone, likely delivering a speech or address. The setting is outdoors, as indicated by the greenery visible behind him. At one point, he gestures with his hands while holding some papers, suggesting he is referencing notes or emphasizing points during his speech. The date "April 2, 2025" is displayed in the top left corner of the screen, indicating when the event is taking place. The logo "N" in the top right corner suggests that the video may be from a news source.

**FAILS TO DRAW “HIGHER-LEVEL” INFERENCES**



# WikiVideo: CAG-2 Output

On April 2, 2025, President Donald Trump delivered a speech in Washington, D.C., announcing new tariffs on imports from several countries, including China, the European Union, Vietnam, Taiwan, Japan, India, South Korea, Thailand, Switzerland, Indonesia, and Malaysia. The tariffs, ranging from 20% to 36%, were designed to address what the administration perceived as unfair trade practices and to encourage domestic manufacturing. Trump emphasized that the U.S. would not impose a full 100% reciprocal tariff, stating his intention to be \"kind to the world\" while protecting American industries. The announcement, made after the stock market closing bell, had already contributed to market volatility, with significant losses reported in the stock market, including a \$3 trillion drop in value, referred to as a \"stock market armageddon.\" Financial analysts, such as those from JPMorgan Chase & Co., raised the recession risk forecast to 60%, up from 40% prior to the announcement. The legal and economic implications of these tariffs were widely discussed, with experts debating the constitutionality of the President's actions under the International Emergency Economic Powers Act (IEEPA) and the potential long-term impacts on global trade dynamics and consumer prices.

# WikiVideo: CAG-2 Output

On April 2, 2025, **President Donald Trump** delivered a speech in Washington, D.C., announcing new tariffs on imports from several countries, including China, the European Union, Vietnam, Taiwan, Japan, India, South Korea, Thailand, Switzerland, Indonesia, and Malaysia. The tariffs, ranging from 20% to 36%, were designed to address what the administration perceived as unfair trade practices and to encourage domestic manufacturing. Trump emphasized that the U.S. would not impose a full 100% reciprocal tariff, stating his intention to be \"kind to the world\" while protecting American industries. The announcement, made after the stock market closing bell, had already contributed to market volatility, with significant losses reported in the stock market, including a \$3 trillion drop in value, referred to as a \"stock market armageddon.\" Financial analysts, such as those from JPMorgan Chase & Co., raised the recession risk forecast to 60%, up from 40% prior to the announcement. The legal and economic implications of these tariffs were widely discussed, with experts debating the constitutionality of the President's actions under the International Emergency Economic Powers Act (IEEPA) and the potential long-term impacts on global trade dynamics and consumer prices.

# WikiVideo: CAG-2 Output

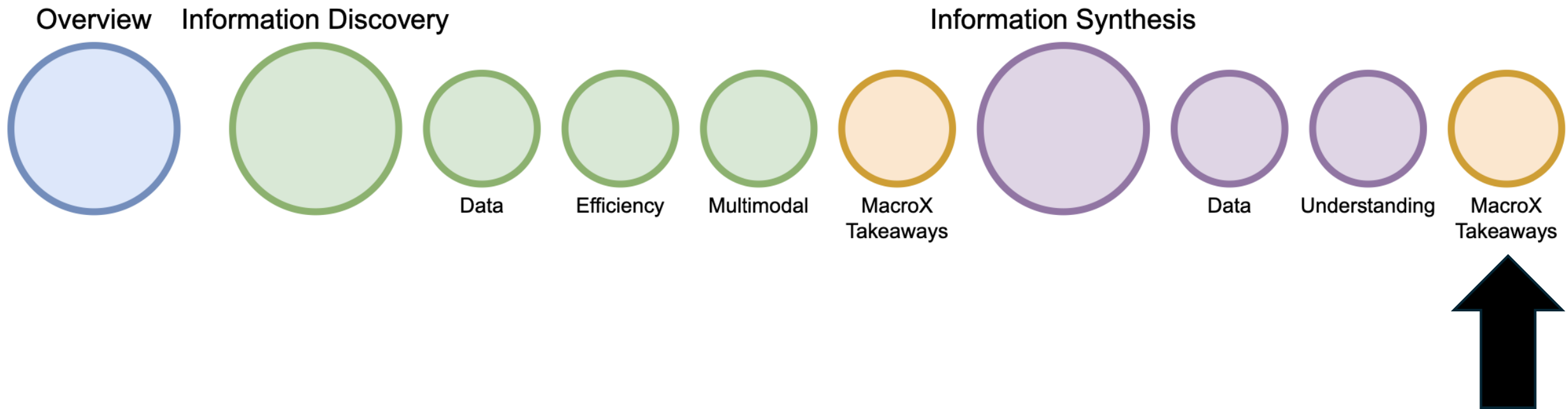
On April 2, 2025, **President Donald Trump** delivered a speech in Washington, D.C., announcing **new tariffs on imports from several countries, including China, the European Union, Vietnam, Taiwan, Japan, India, South Korea, Thailand, Switzerland, Indonesia, and Malaysia.** The tariffs, ranging from 20% to 36%, were designed to address what the administration perceived as unfair trade practices and to encourage domestic manufacturing. Trump emphasized that the U.S. would not impose a full 100% reciprocal tariff, stating his intention to be \"kind to the world\" while protecting American industries. The announcement, made after the stock market closing bell, had already contributed to market volatility, with significant losses reported in the stock market, including a \$3 trillion drop in value, referred to as a \"stock market armageddon.\" Financial analysts, such as those from JPMorgan Chase & Co., raised the recession risk forecast to 60%, up from 40% prior to the announcement. The legal and economic implications of these tariffs were widely discussed, with experts debating the constitutionality of the President's actions under the International Emergency Economic Powers Act (IEEPA) and the potential long-term impacts on global trade dynamics and consumer prices.

# WikiVideo: CAG-2 Output

On April 2, 2025, **President Donald Trump** delivered a speech in Washington, D.C., announcing **new tariffs on imports from several countries, including China, the European Union, Vietnam, Taiwan, Japan, India, South Korea, Thailand, Switzerland, Indonesia, and Malaysia.** The tariffs, **ranging from 20% to 36%,** were designed to address what the administration perceived as unfair trade practices and to encourage domestic manufacturing. Trump emphasized that the U.S. would not impose a full 100% reciprocal tariff, stating his intention to be \"kind to the world\" while protecting American industries. The announcement, made after the stock market closing bell, had already contributed to market volatility, with significant losses reported in the stock market, including a \$3 trillion drop in value, referred to as a \"stock market armageddon.\" Financial analysts, such as those from JPMorgan Chase & Co., raised the recession risk forecast to 60%, up from 40% prior to the announcement. The legal and economic implications of these tariffs were widely discussed, with experts debating the constitutionality of the President's actions under the International Emergency Economic Powers Act (IEEPA) and the potential long-term impacts on global trade dynamics and consumer prices.

# WikiVideo: CAG-2 Output

On April 2, 2025, **President Donald Trump** delivered a speech in Washington, D.C., announcing **new tariffs on imports from several countries, including China, the European Union, Vietnam, Taiwan, Japan, India, South Korea, Thailand, Switzerland, Indonesia, and Malaysia.** The tariffs, **ranging from 20% to 36%,** were designed to address what the administration perceived as unfair trade practices and to encourage domestic manufacturing. Trump emphasized that the U.S. would not impose a full 100% reciprocal tariff, stating his intention to be \"kind to the world\" while protecting American industries. The announcement, made after the stock market closing bell, had already contributed to market volatility, with significant losses reported in the stock market, including a **\$3 trillion drop in value,** referred to as a \"stock market armageddon.\" **Financial analysts, such as those from JPMorgan Chase & Co., raised the recession risk forecast to 60%, up from 40% prior to the announcement.** The legal and economic implications of these tariffs were widely discussed, with experts debating the constitutionality of the President's actions under the International Emergency Economic Powers Act (IEEPA) and the potential long-term impacts on global trade dynamics and consumer prices.



# MacroX Takeaway

# Analysts writing reports (Assisted writing)

**Our Reasoning Stack**  
CAG  
Video-ColBERT/MMMORRF



**Analyst**

Great job on this report analyst!  
Enjoy your weekend off.

## **Best Macro Forecast**

---

---

---

---



# Feed cleaner & specific signal



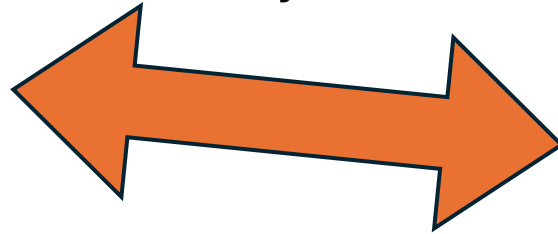


# Generalist Forecast Understanding

- Varying query needs
- Expand to on demand reasoning for any forecast, information need
- Examples: Analyst projecting recession, farmer worrying about immigration policy and hiring workers, etc.

# Explainable Forecasting

announcing new tariffs on imports from several countries, including China, the European Union, Vietnam, Taiwan, Japan, India, South Korea, Thailand, Switzerland, Indonesia, and Malaysia, ranging from 10-36% **[Source]**

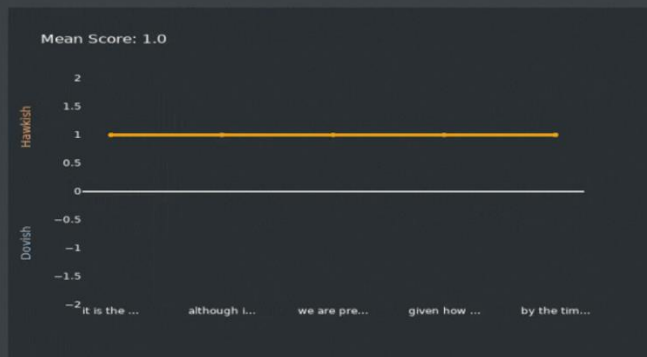


# Improving this tool I found on your website

## Live hawkish/dovish score and summary for Chairman Powell's speech



### LIVE SCORE



It is the Fed's job to bring inflation down to our 2% goal, and we will do so. We have tightened

### LIVE SUMMARY

#### Opening Remarks

Mostly Hawkish: +0.5

Speaker: Chair Jerome H. Powell

Date: 10/19/2023

#### Future Policy Stance:

- Future policy aims to sustainably reduce inflation to the target of 2 percent, with recognition of ongoing inflation concerns and the need for caution.
- Policy decisions will be data-driven, considering economic growth, labor market conditions, and the

# Improving and Expanding the Tool

Live Data



Information Need (Query)



**Live Forecasting Tool**



Augmented Knowledge

- More informed
- Exactly relevant to query
- Situated in up-to-date world knowledge



A little bit of advertising

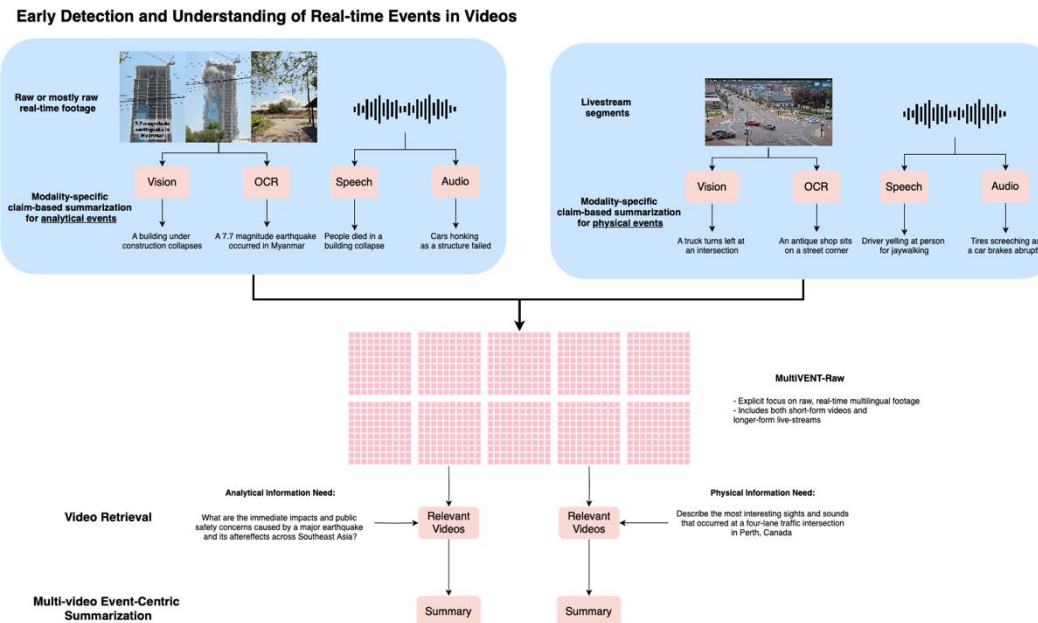
# About Me

- JHU PhD
- “multimodal understanding and reasoning”
- A tag line for macroX: “scalable domain expertise through augmenting reasoning with internet scale information”
- My research: modality agnostic approaches to efficient information seeking and synthesis



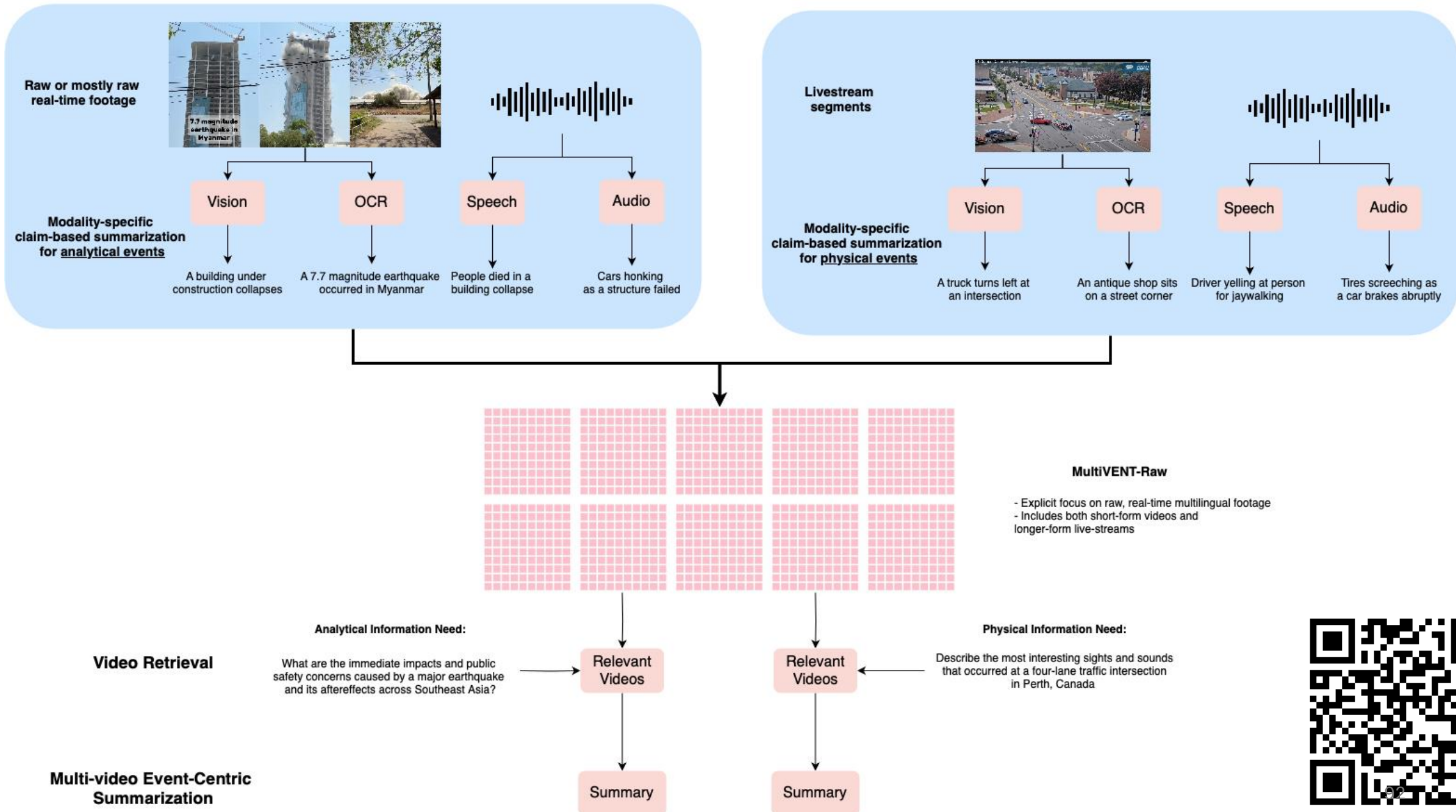
# SCALE26: Event Understanding and Summarization from Real-time Videos

- Summer Workshop hosted by my research center
- Theme 1: finding raw, real-time, multilingual videos
  - (think social media based event detection and retrieval)
- Theme 2: multi-video article generation (wikivideo)





# Early Detection and Understanding of Real-time Events in Videos



Video-ColBERT



WikiVideo



MultiVENT



[nlp.jhu.edu/multivent/](http://nlp.jhu.edu/multivent/)

ACL Workshop on  
Multimodal RAG



[nlp.jhu.edu/magmar/](http://nlp.jhu.edu/magmar/)