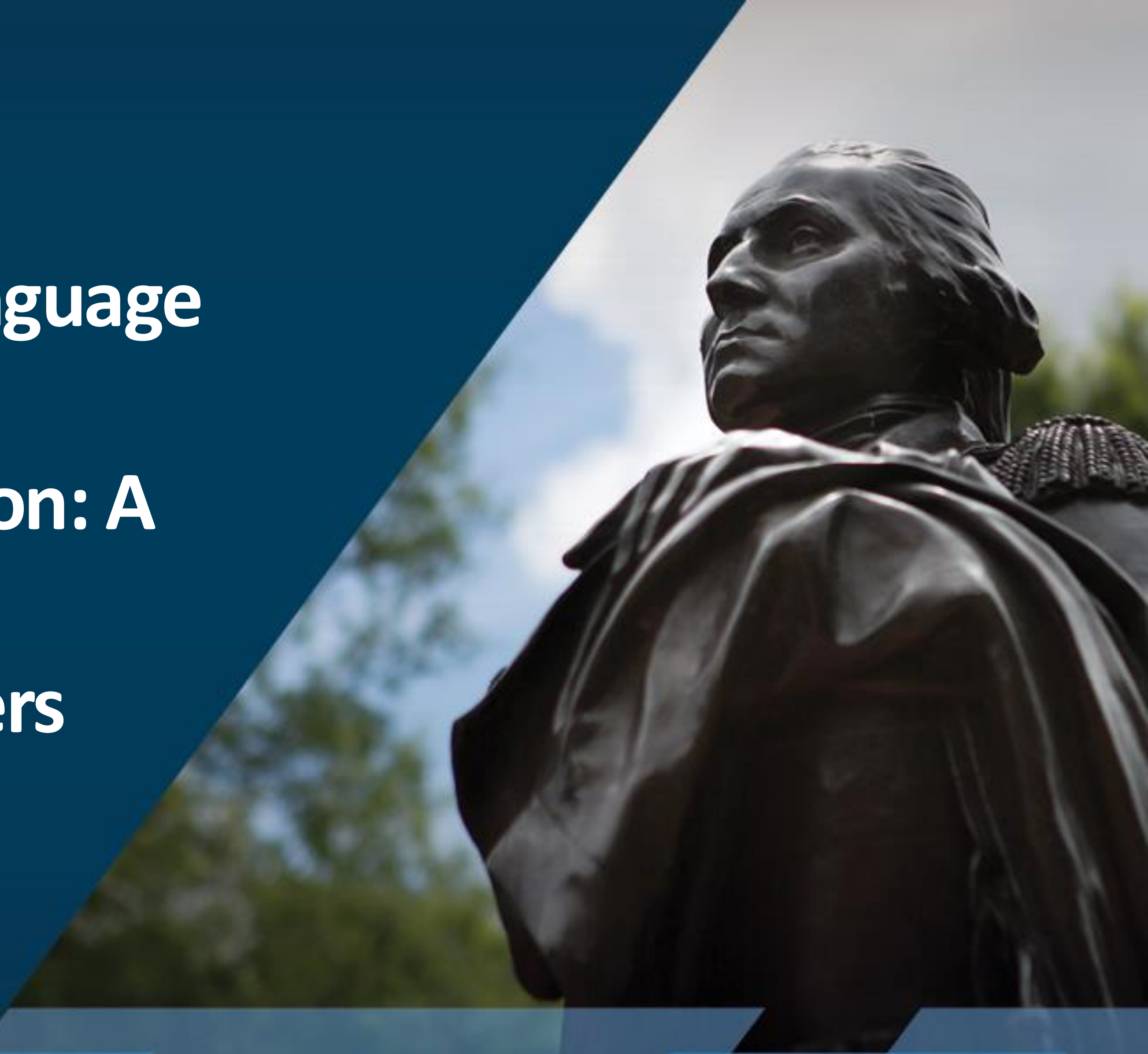
THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Leveraging Large Language Models for Logistics Information Extraction: A Case Study on Two International Disasters

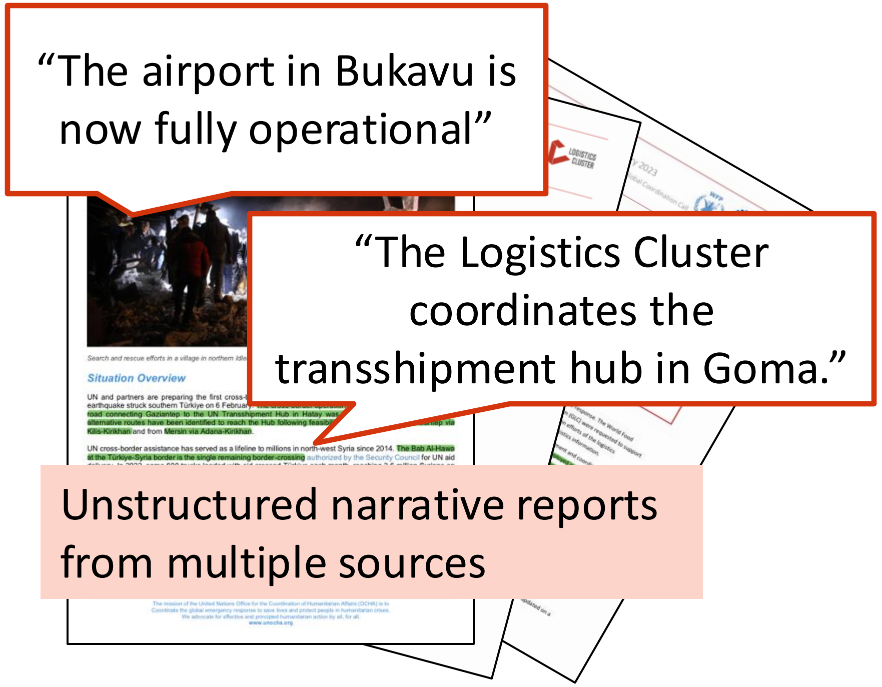


Zaid Kbah, PhD Candidate
Dr. Erica Gralla, Associate Professor

Background

- Situational awareness during disasters requires quick synthesizing of information from multiple sources

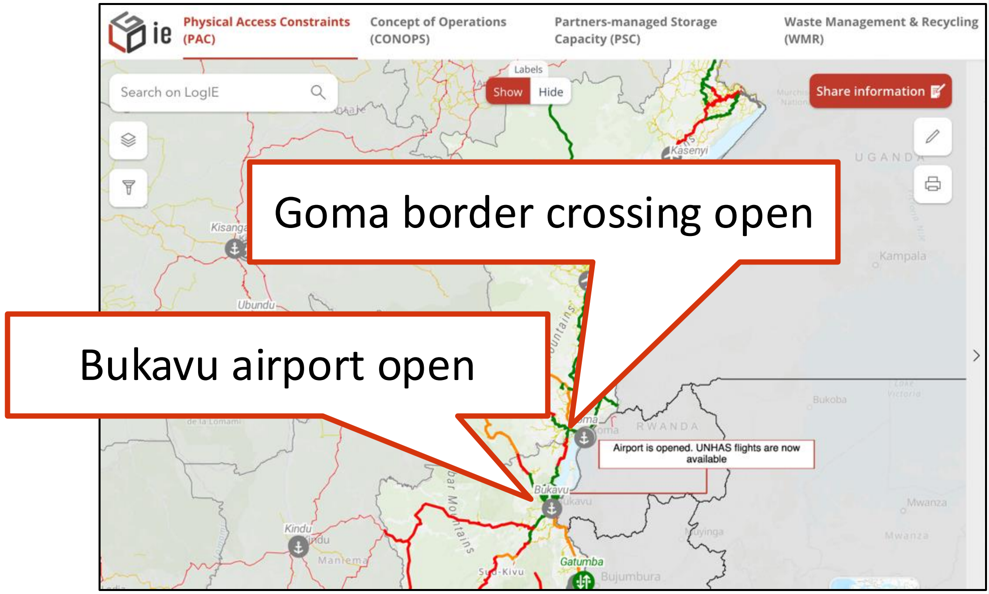
- Information management officers' triage, interpret, and convert text into structured, mission-usable products.
- Can a LLM accomplish the same task?



“The airport in Bukavu is now fully operational”

“The Logistics Cluster coordinates the transshipment hub in Goma.”

Unstructured narrative reports from multiple sources



Goma border crossing open

Bukavu airport open

Airport is opened. UNHAS flights are now available

Physical Access Constraints (PAC)
Concept of Operations (CONOPS)
Partners-managed Storage Capacity (PSC)
Waste Management & Recycling (WMR)
Search on LogIE
Labels
Show
Hide
Share information
Kasenyi
UGANDA
Kampala
Kisangani
Ubundu
RWANDA
Bukavu
Kindu
Gatumba
Bujumbura
Mwanza

Common operating picture:
Maps of logistics infrastructure

Research Question

- **How effectively can ChatGPT extract relevant infrastructure and logistics information from text-based disaster reports?**
  - How do model configuration parameters, such as version and temperature, affect its performance?
  - What strategies can enhance the performance?

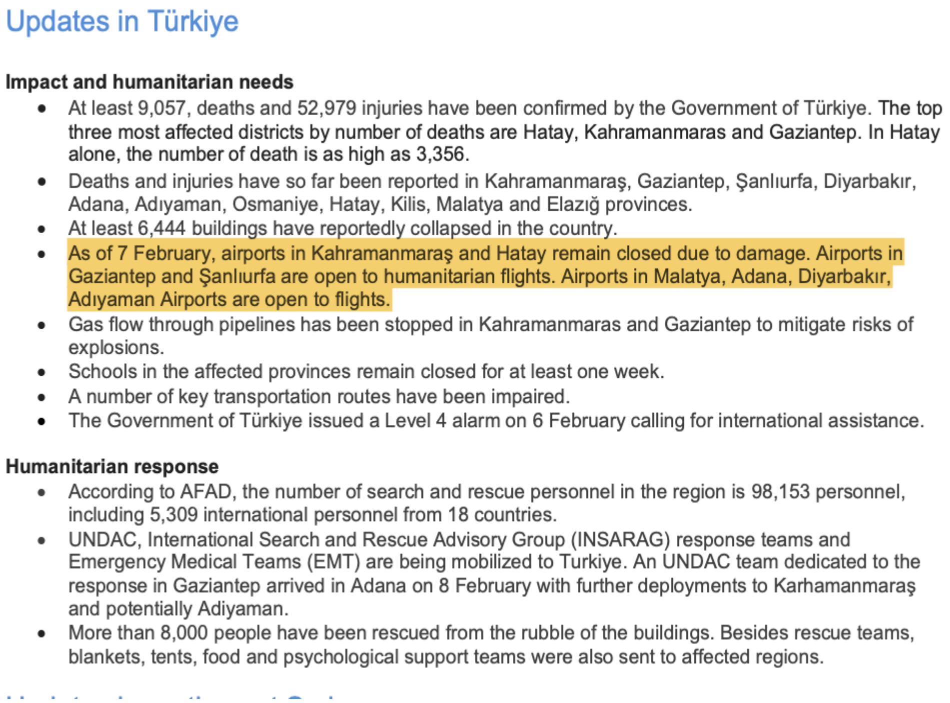
Literature Review

- Studies highlight LLM capabilities in various tasks:
  - Geolocation extraction – Yin et al. (2023)
  - Temporal relationship identification – Yuan et al. (2023)
  - Dialogue generation – Bai et al. (2023)
  - Annotation – Gilardi et al. (2023); Labruna et al. (2023)
  - Decision-making in wargame simulations – Lamparth et al. (2024), etc.
- Our task requires deep contextual understanding and domain-specific knowledge of logistics, and has not been explored

Task: Extract structured logistics information from a narrative report

Text Narrative Report

Structured Data on the
Status of Logistics Infrastructure (Statements)



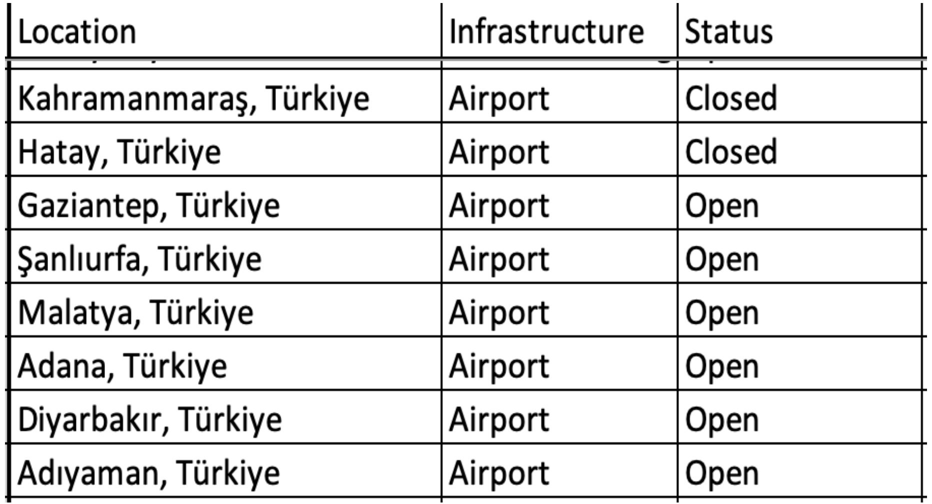
Updates in Türkiye

Impact and humanitarian needs

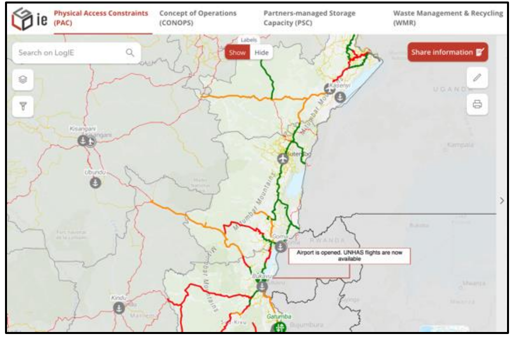
- At least 9,057, deaths and 52,979 injuries have been confirmed by the Government of Türkiye. The top three most affected districts by number of deaths are Hatay, Kahramanmaras and Gaziantep. In Hatay alone, the number of death is as high as 3,356.
- Deaths and injuries have so far been reported in Kahramanmaraş, Gaziantep, Şanlıurfa, Diyarbakır, Adana, Adıyaman, Osmaniye, Hatay, Kilis, Malatya and Elazığ provinces.
- At least 6,444 buildings have reportedly collapsed in the country.
- As of 7 February, airports in Kahramanmaraş and Hatay remain closed due to damage. Airports in Gaziantep and Şanlıurfa are open to humanitarian flights. Airports in Malatya, Adana, Diyarbakır, Adıyaman Airports are open to flights.
- Gas flow through pipelines has been stopped in Kahramanmaras and Gaziantep to mitigate risks of explosions.
- Schools in the affected provinces remain closed for at least one week.
- A number of key transportation routes have been impaired.
- The Government of Türkiye issued a Level 4 alarm on 6 February calling for international assistance.

Humanitarian response

- According to AFAD, the number of search and rescue personnel in the region is 98,153 personnel, including 5,309 international personnel from 18 countries.
- UNDAC, International Search and Rescue Advisory Group (INSARAG) response teams and Emergency Medical Teams (EMT) are being mobilized to Turkiye. An UNDAC team dedicated to the response in Gaziantep arrived in Adana on 8 February with further deployments to Karhamanmaraş and potentially Adiyaman.
- More than 8,000 people have been rescued from the rubble of the buildings. Besides rescue teams, blankets, tents, food and psychological support teams were also sent to affected regions.



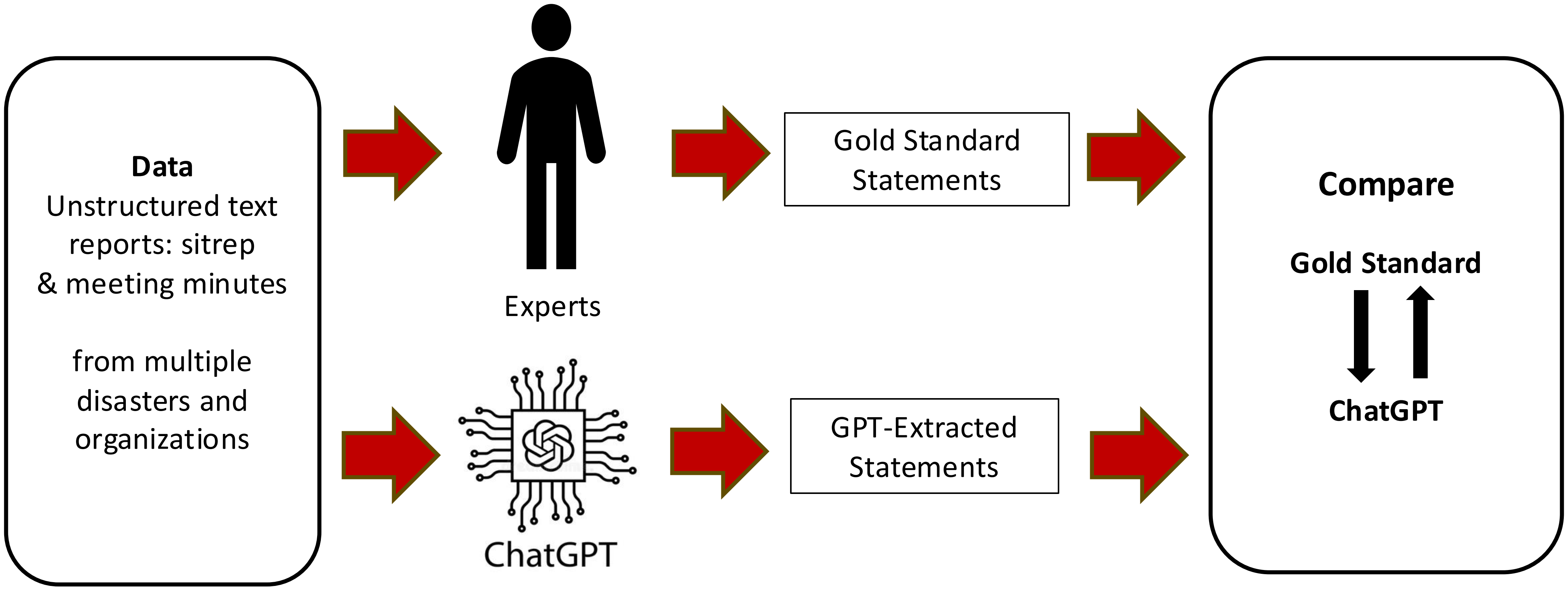
Location	Infrastructure	Status
Kahramanmaraş, Türkiye	Airport	Closed
Hatay, Türkiye	Airport	Closed
Gaziantep, Türkiye	Airport	Open
Şanlıurfa, Türkiye	Airport	Open
Malatya, Türkiye	Airport	Open
Adana, Türkiye	Airport	Open
Diyarbakır, Türkiye	Airport	Open
Adıyaman, Türkiye	Airport	Open


ie
Physical Access Constraints (PAC)
Concept of Operations (CONOPS)
Partners-managed Storage Capacity (PSC)
Waste Management & Recycling (WMR)
Search on LogIE
Labels
Show
Hide
Share information
UGANDA
RWANDA
Kisangani
Ubundu
Kindu
Goma
Bukavu
Gatumba
Kampala
Mitumbar Mountains

Airport is opened. UNHAS flights are now available

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Methodology Overview



Data
Unstructured text reports: sitrep & meeting minutes

from multiple disasters and organizations

Experts

ChatGPT

Gold Standard Statements

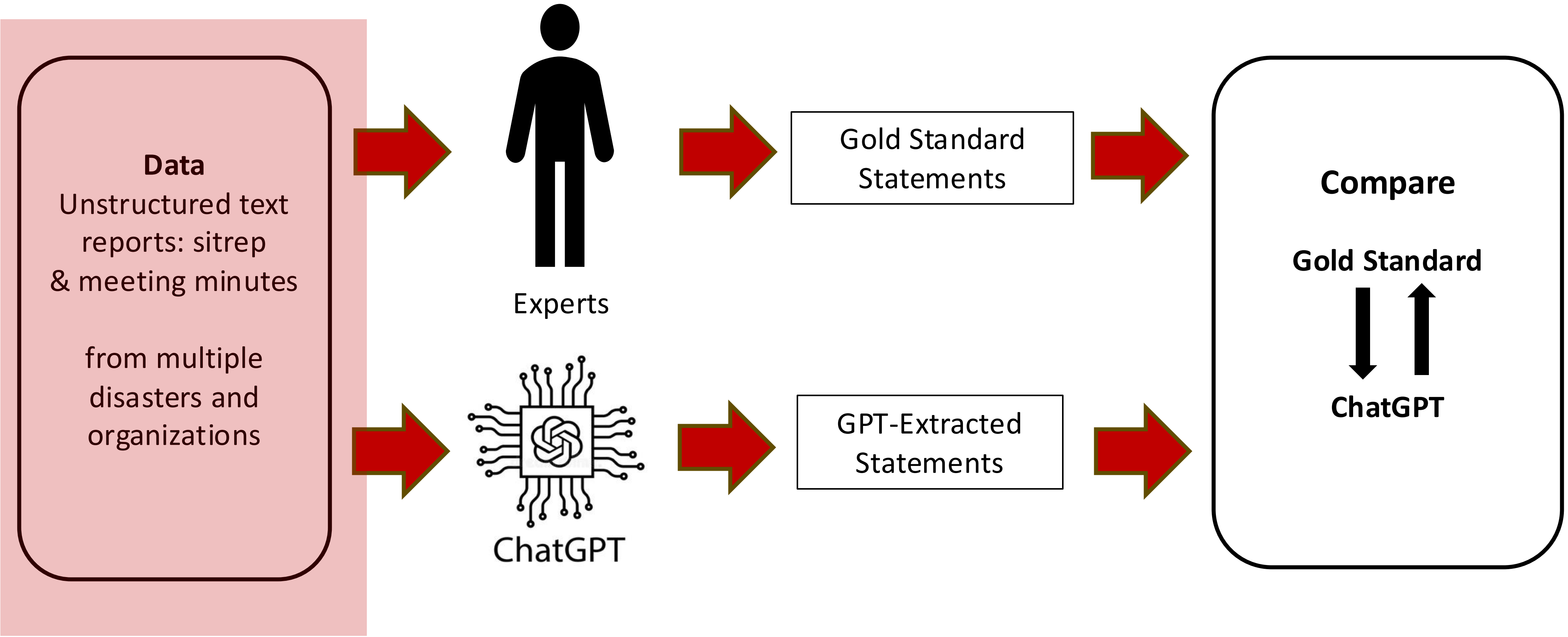
GPT-Extracted Statements

Compare

Gold Standard

ChatGPT

Methodology Overview



Data
Unstructured text reports: sitrep & meeting minutes

from multiple disasters and organizations

Experts

ChatGPT

Gold Standard Statements

GPT-Extracted Statements

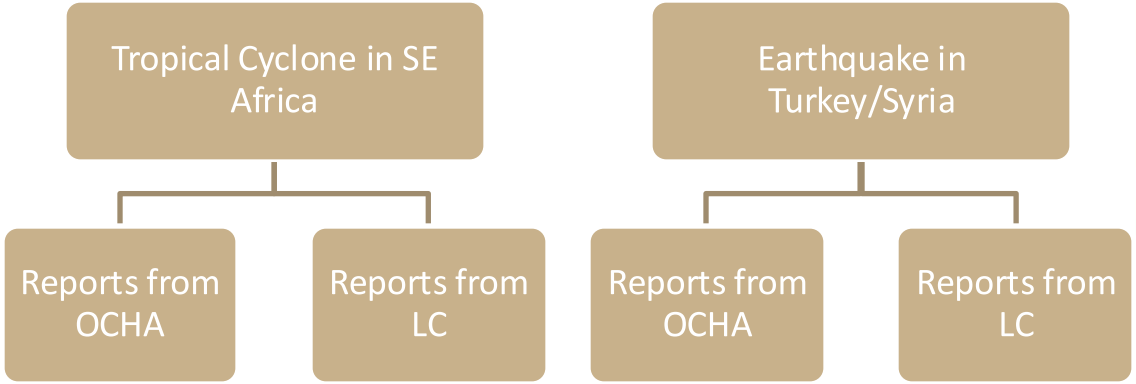
Compare

Gold Standard

ChatGPT

Data: 40 reports from two disasters

- Selected **two different disasters**
  - Tropical Cyclone Freddy in Southeast Africa (SEA)
  - Earthquake in Turkey/Syria (TRKY)
- Selected **two different organizations**:
  - Broad scope: United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA)
  - Narrow focus on logistics: Logistics Cluster (LC)
- Selected the **first 10 documents** produced by each organization in each disaster



Tropical Cyclone in SE Africa

Reports from OCHA

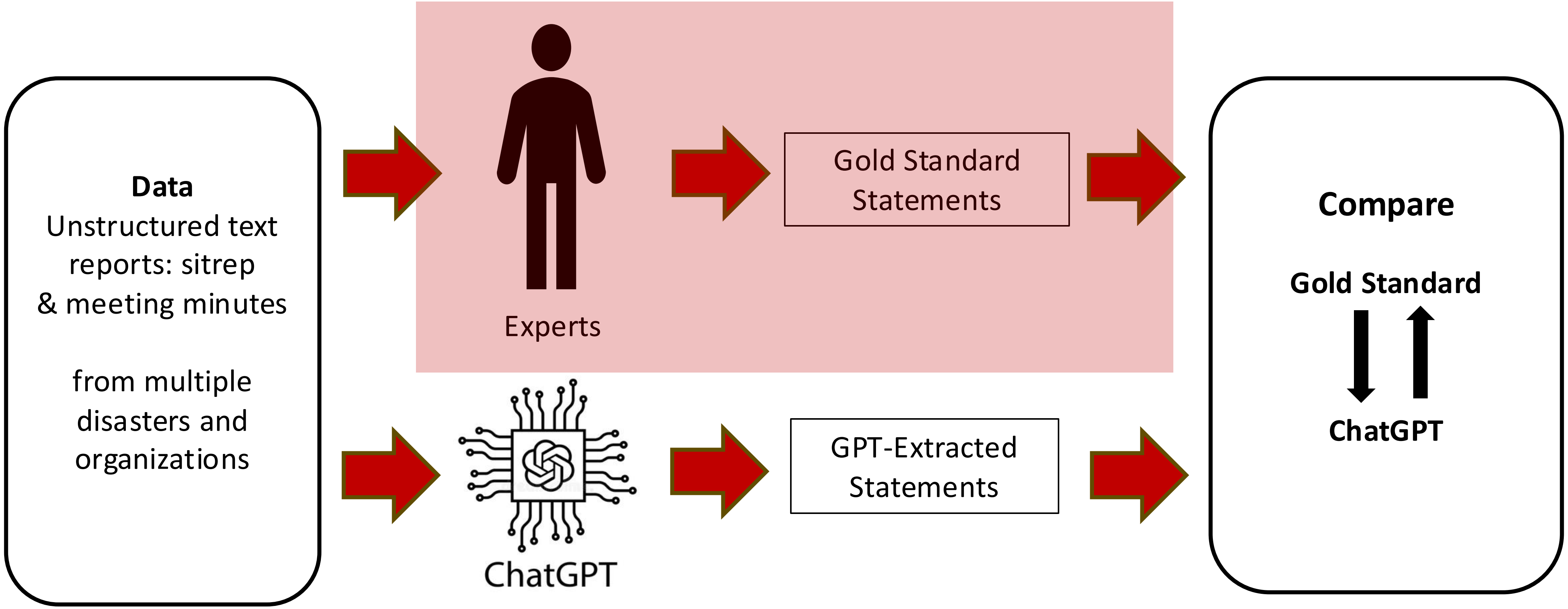
Reports from LC

Earthquake in Turkey/Syria

Reports from OCHA

Reports from LC

Methodology Overview



Data
Unstructured text reports: sitrep & meeting minutes

from multiple disasters and organizations

Experts

Gold Standard Statements

ChatGPT

GPT-Extracted Statements

Compare

Gold Standard

ChatGPT

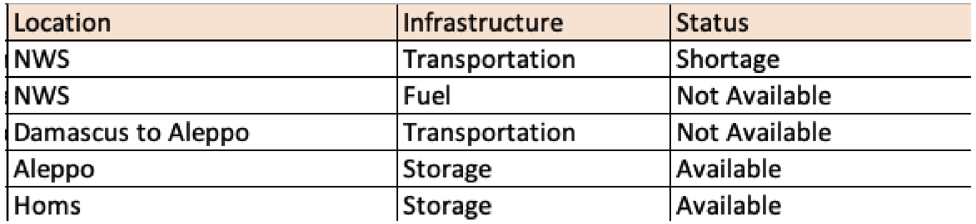
Experts identified all key logistics information in each report

- Guidelines defined what counts as “key logistics information”
- Consistent process
  - Two researchers independently labeled multiple documents; disagreements reconciled.
- Total number of statements extracted: 396

Information to Extract

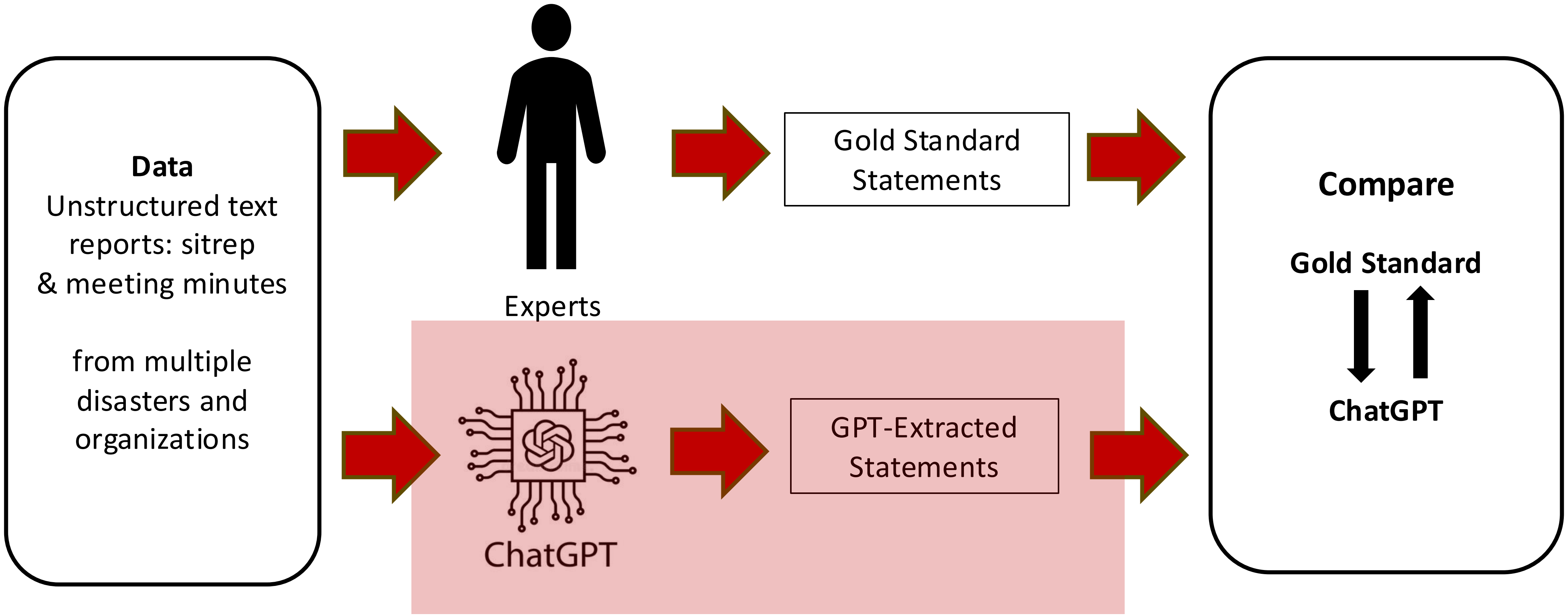
Status and location of:

- Airports
- Seaports
- Road
- Railways
- Transportation
- Storage
- Fuel
- Coordination services



Location	Infrastructure	Status
NWS	Transportation	Shortage
NWS	Fuel	Not Available
Damascus to Aleppo	Transportation	Not Available
Aleppo	Storage	Available
Homs	Storage	Available

Methodology Overview



Data
Unstructured text reports: sitrep & meeting minutes

from multiple disasters and organizations

Experts

Gold Standard Statements

ChatGPT

GPT-Extracted Statements

Compare

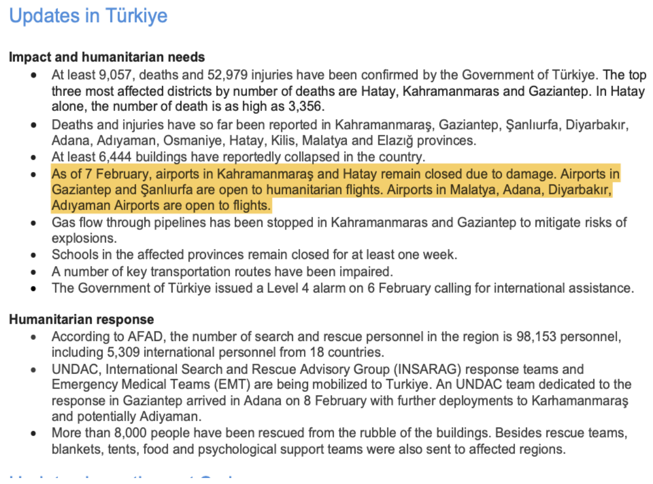
Gold Standard

ChatGPT

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

ChatGPT extracts statements based on same guidelines

- ChatGPT is prompted to extract information from each document
- Prompts were iteratively refined
- Evaluated multiple model versions and temperature settings to assess variability



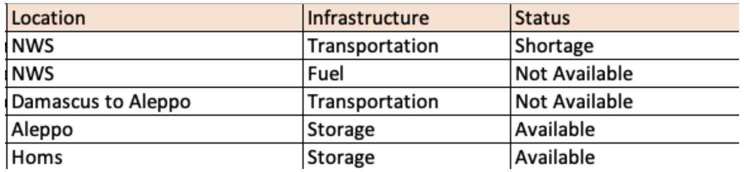
Updates in Türkiye

Impact and humanitarian needs

- At least 9,057, deaths and 52,979 injuries have been confirmed by the Government of Türkiye. The top three most affected districts by number of deaths are Hatay, Kahramanmaras and Gaziantep. In Hatay alone, the number of death is as high as 3,356.
- Deaths and injuries have so far been reported in Kahramanmaraş, Gaziantep, Şanlıurfa, Diyarbakır, Adana, Adıyaman, Osmaniye, Hatay, Kilis, Malatya and Elazığ provinces.
- At least 6,444 buildings have reportedly collapsed in the country.
- As of 7 February, airports in Kahramanmaraş and Hatay remain closed due to damage. Airports in Gaziantep and Şanlıurfa are open to humanitarian flights. Airports in Malatya, Adana, Diyarbakır, Adıyaman Airports are open to flights.
- Gas flow through pipelines has been stopped in Kahramanmaras and Gaziantep to mitigate risks of explosions.
- Schools in the affected provinces remain closed for at least one week.
- A number of key transportation routes have been impaired.
- The Government of Türkiye issued a Level 4 alarm on 6 February calling for international assistance.

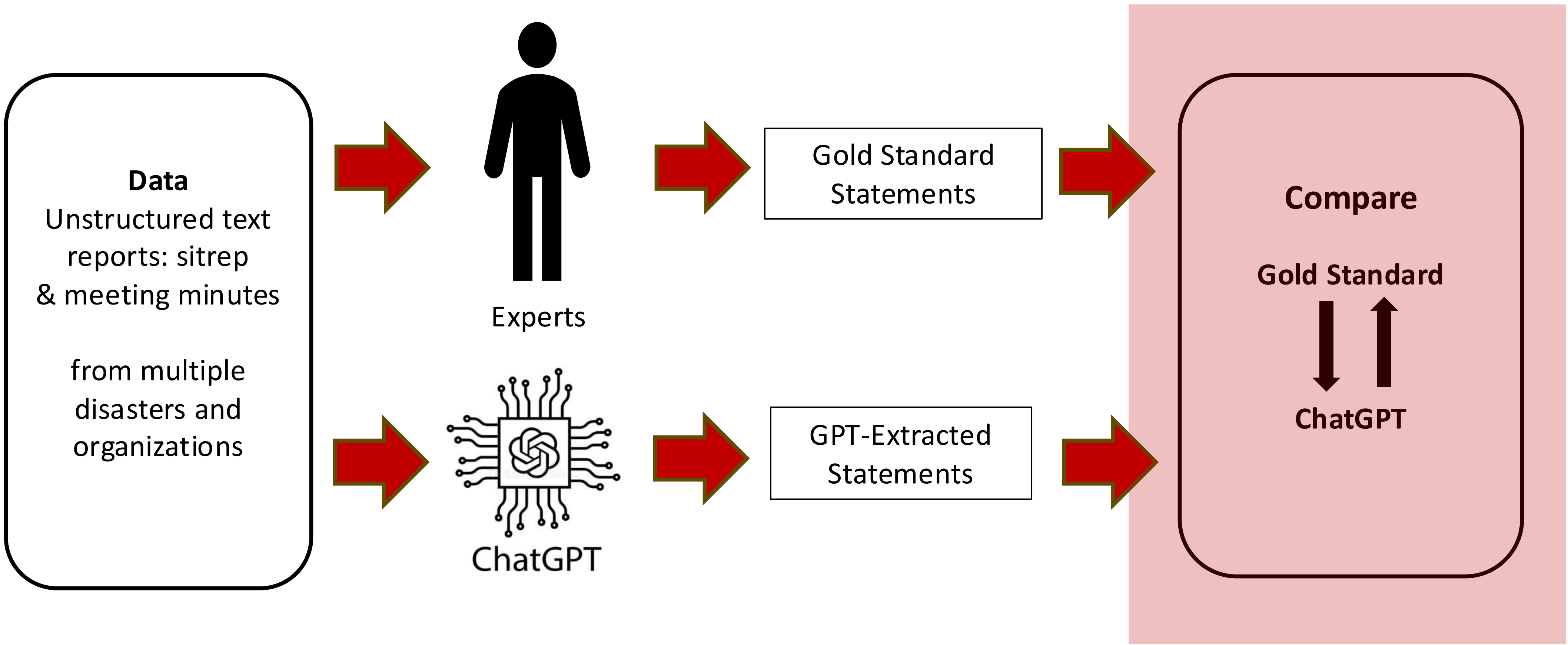
Humanitarian response

- According to AFAD, the number of search and rescue personnel in the region is 98,153 personnel, including 5,309 international personnel from 18 countries.
- UNDAC, International Search and Rescue Advisory Group (INSARAG) response teams and Emergency Medical Teams (EMT) are being mobilized to Turkiye. An UNDAC team dedicated to the response in Gaziantep arrived in Adana on 8 February with further deployments to Karhamanmaraş and potentially Adiyaman.
- More than 8,000 people have been rescued from the rubble of the buildings. Besides rescue teams, blankets, tents, food and psychological support teams were also sent to affected regions.



Location	Infrastructure	Status
NWS	Transportation	Shortage
NWS	Fuel	Not Available
Damascus to Aleppo	Transportation	Not Available
Aleppo	Storage	Available
Homs	Storage	Available

Methodology Overview



Data
Unstructured text reports: sitrep & meeting minutes

from multiple disasters and organizations

Experts

ChatGPT

Gold Standard Statements

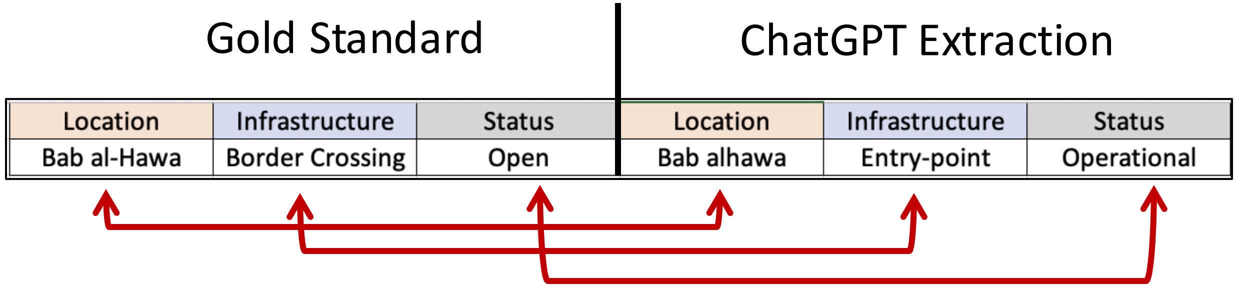
GPT-Extracted Statements

Compare

Gold Standard

ChatGPT

Comparison



Gold Standard			ChatGPT Extraction		
Location	Infrastructure	Status	Location	Infrastructure	Status
Bab al-Hawa	Border Crossing	Open	Bab alhawa	Entry-point	Operational

Matched statement

Errors

Mis-match in statement component(s)

Gold Standard			ChatGPT Extraction		
Location	Infrastructure	Status	Location	Infrastructure	Status
Bab al-Hawa	Border Crossing	Open	Bab alhawa	Entry-point	Closed

GPT did not catch the statement

Gold Standard			ChatGPT Extraction		
Location	Infrastructure	Status			
Bab Salama	Border Crossing	Open			

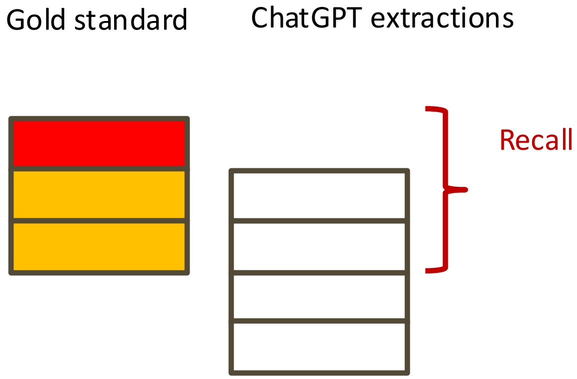
GPT suggested an incorrect statement

Gold Standard			ChatGPT Extraction		
			Location	Infrastructure	Status
			Border	Procedures	Controlled

Performance Metrics

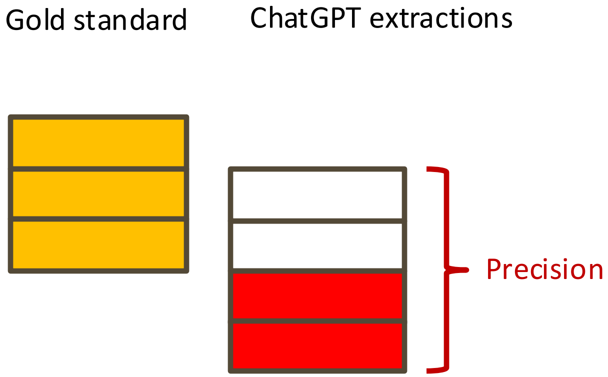
- Recall: How many of the gold standard statements does ChatGPT correctly identify?

$$\boldsymbol{Recall} = \frac{Matched\ Statements}{Gold\ Standard\ Statements}$$



- Precision: How many of ChatGPT's statements are in the gold standard (i.e., not "extra")?

$$\boldsymbol{Precision} = \frac{Matched\ Statements}{ChatGPT\ Statements}$$



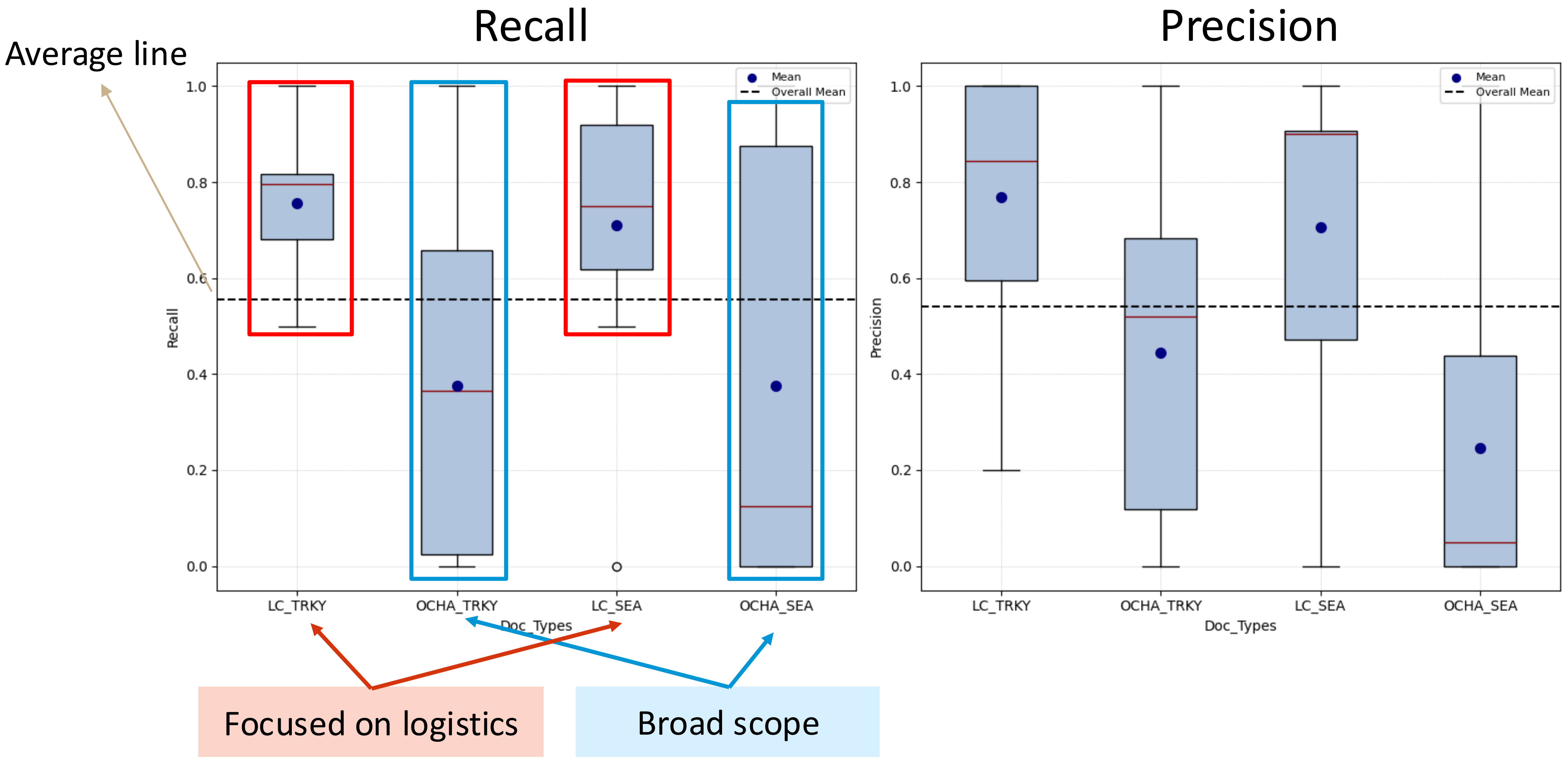
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Overall Performance

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Performance differs based on source document



Recall

Precision

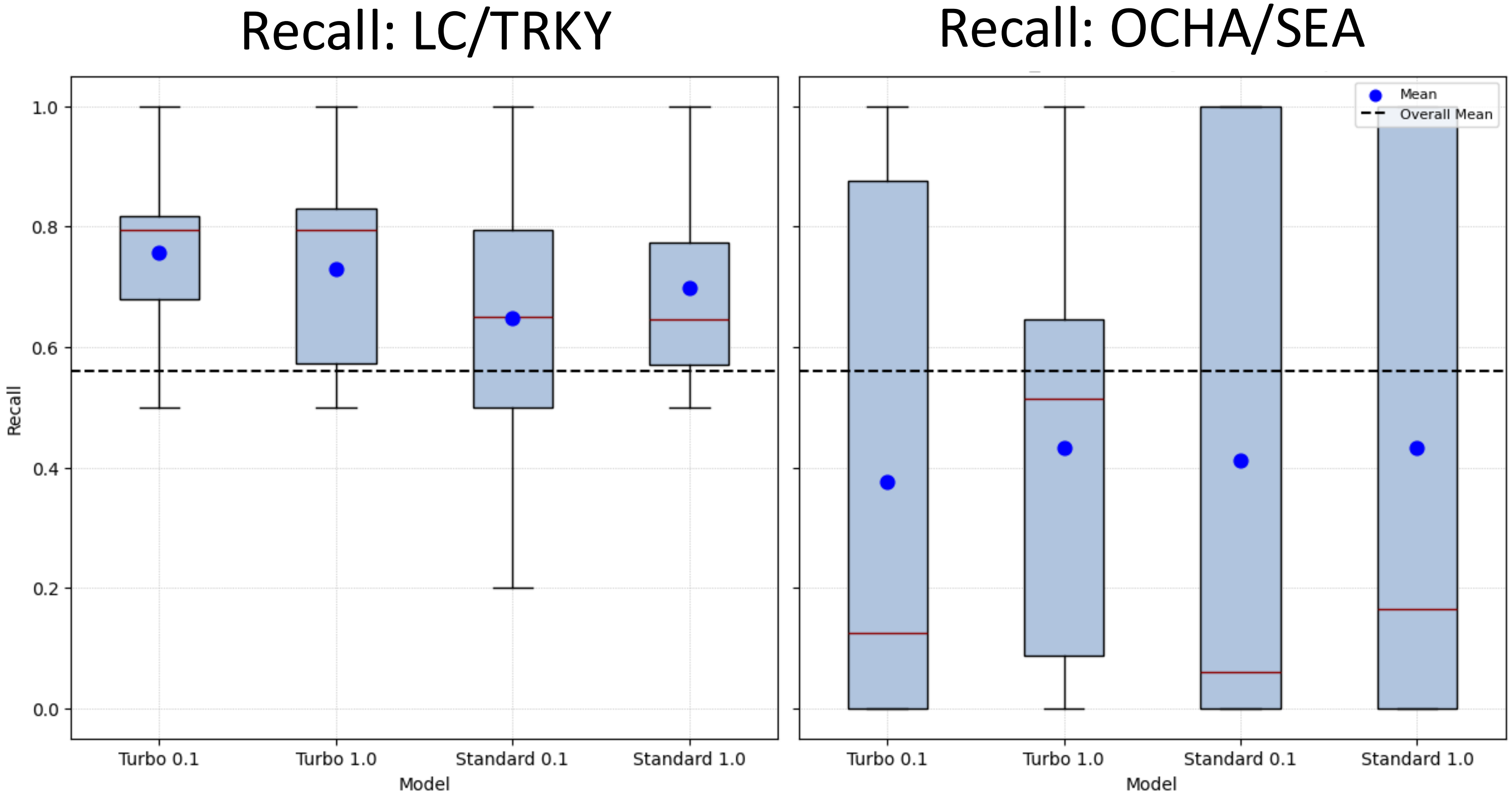
Average line

Focused on logistics

Broad scope

Performance depends on the **organization** that produced the report

Performance does not depend on model parameters



Performance is consistent across different **model versions** and **temperatures**

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

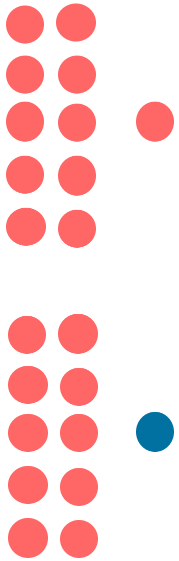
Improving Performance

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

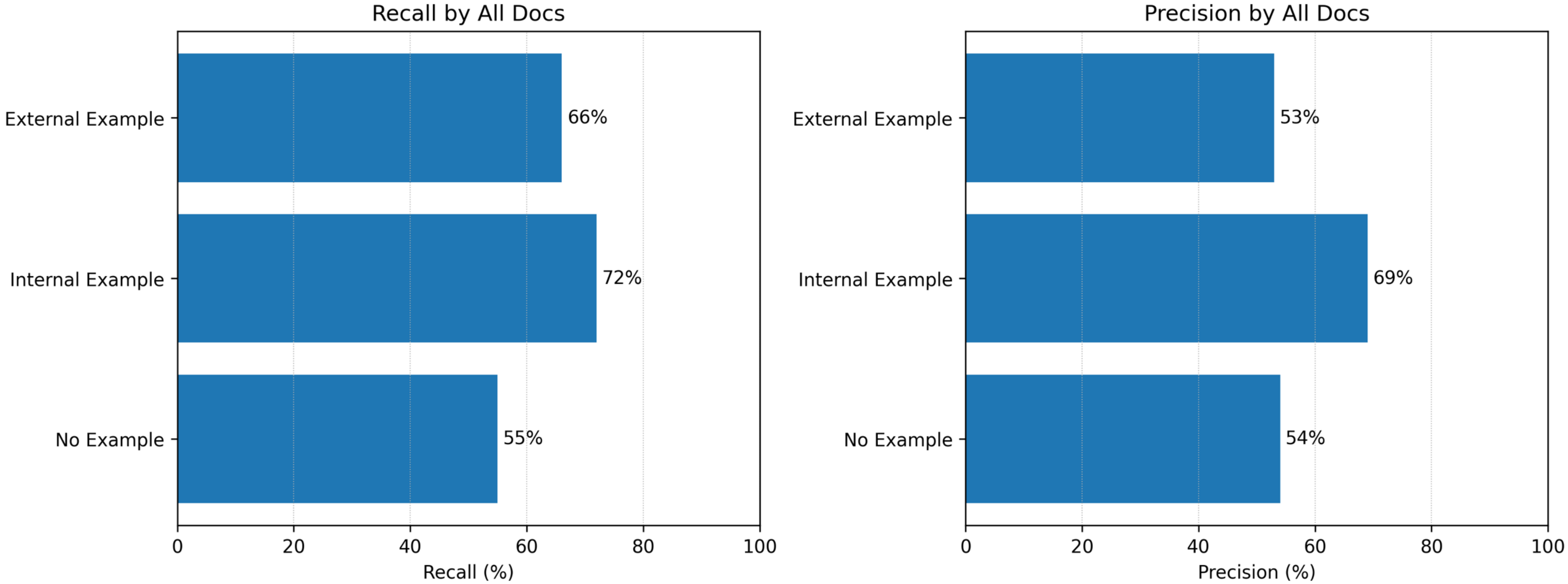
Attempt 1: Provide an example

- **Will performance improve if the prompt includes an example?**

- Tested two types of examples
  - **Internal:** from a document produced by the same organization in the same disaster
  - **External:** from a document produced by a different organization in a different disaster



Performance improves with internal examples



Presenting an example from the same dataset improves recall and precision

*External Example: an example from different disaster and organization

** Internal Example: an example from the same dataset

Attempt 2: Remove ambiguity

- **Will performance improve if we remove ambiguous categories of information?**
  - Information on coordination was very difficult to interpret in the documents
- Tested performance with coordination removed from the scope

Information to Extract

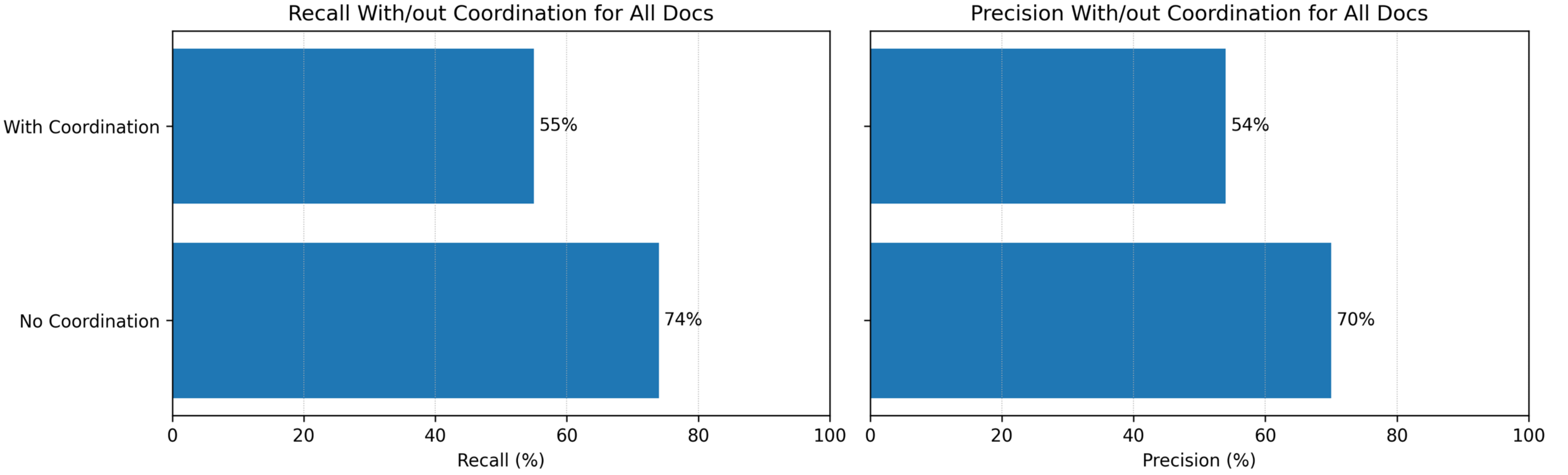
Status and location of:

- Airports
- Seaports
- Road
- Railways
- Transportation
- Storage
- Fuel
- ~~Coordination services~~

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

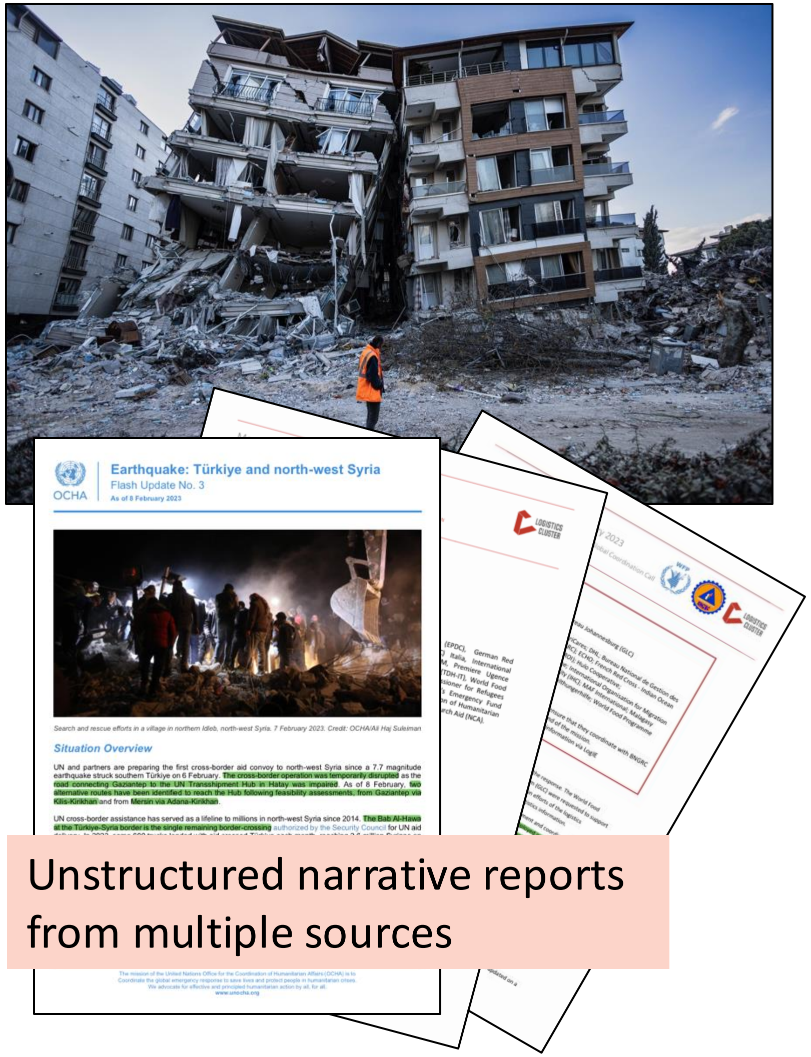
Performance improves when ambiguous categories are removed from consideration



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Can LLMs extract a logistics COP from narratives?



Earthquake: Türkiye and north-west Syria

Flash Update No. 3

As of 8 February 2023

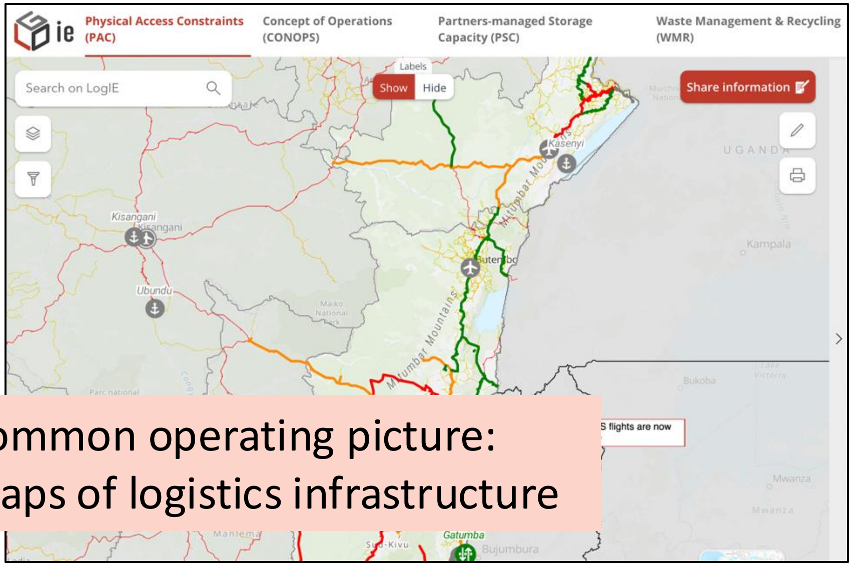
Search and rescue efforts in a village in northern Idleb, north-west Syria. 7 February 2023. Credit: OCHA/Ali Haj Suleiman

Situation Overview

UN and partners are preparing the first cross-border aid convoy to north-west Syria since a 7.7 magnitude earthquake struck southern Türkiye on 6 February. The cross-border operation was temporarily disrupted as the road connecting Gaziantep to the UN Transshipment Hub in Hatay was impaired. As of 8 February, two alternative routes have been identified to reach the Hub following feasibility assessments, from Gaziantep via Kilis-Kirikhan and from Mersin via Adana-Kirikhan.

UN cross-border assistance has served as a lifeline to millions in north-west Syria since 2014. The Bab Al-Hawa at the Türkiye-Syria border is the single remaining border-crossing authorized by the Security Council for UN aid

Unstructured narrative reports from multiple sources



ie
Physical Access Constraints (PAC)
Concept of Operations (CONOPS)
Partners-managed Storage Capacity (PSC)
Waste Management & Recycling (WMR)
Search on LogIE
Labels
Show
Hide
Share information
Kisangani
Ubundu
Kasenyi
Kampala
UGANDA
Bukoba
Mwanza
Gatumba
Bujumbura
Sud-Kivu
Maniema

Common operating picture: Maps of logistics infrastructure

Conclusions

- Performance depends on the **documents**!
  - Works well for more direct, structured narratives (0.75-0.8 precision, recall)
  - Works less well for broader, less focused narratives (highly variable performance)
- Providing **relevant examples** improves performance
- Performance also depends on the **ambiguity** of the information to be extracted
  - Performance was poor in extracting information on coordination services, better for airports, seaports, etc.
- **Temperature and model version** has little impact on performance

Ongoing Work

- Analyzing the nature of the errors ChatGPT makes
  - …and how they can be fixed
- Verifying the results by feeding the model with additional information from different sources

Thank you!

Contacts:
Zaid Kbah, PhD Candidate, zkbah@gwu.edu
Erica Gralla, Associate Professor, egralla@gwu.edu