Analyzing the Capability of Large Language Models in Extracting Landslide Event Data

HIRES Interns: Megha Nakarmi, Romalyne Jean - Julien; **Mentor:** Mame Diarra Gueye, Seon-Ho Kim; **Faculty Advisors/Researchers:** Naresh Devineni

Address/Affiliation - CUNY CREST High School Initiative in Remote Sensing of Earth System Engineering and Sciences (HIRES)
The City College of New York, NY 10031

BACKGROUND

- Landslides are mass movements of materials such as rocks, earth, or debris down a slope. It can happen suddenly or over a long period of time. Landslides can happen anywhere and can be very destructive and unpredictable.

- Recording landslide data is important to predict future disasters.

- NASA spent 2007-2019 creating a Global Landslide Catalog retrieving information about landslide dates, location, and more.

- The work NASA was doing was very time consuming and costly because the information was extracted by humans.

- Nowadays, LLMs are more immediate and are advancing to understand human generated prompts.

RESEARCH OBJECTIVE

- Motivated by recent advancements in AI and the need for research in this field, we have decided to explore how effective Large Language Models (LLMs) are at identifying and retrieving real-world landslide events.

METHODS AND STUDY AREA

- This research focused on landslide events in New York during 2012-2016.

- NASA's Global Landslide Catalog was used as a comparative tool to determine the LLMs accuracy.
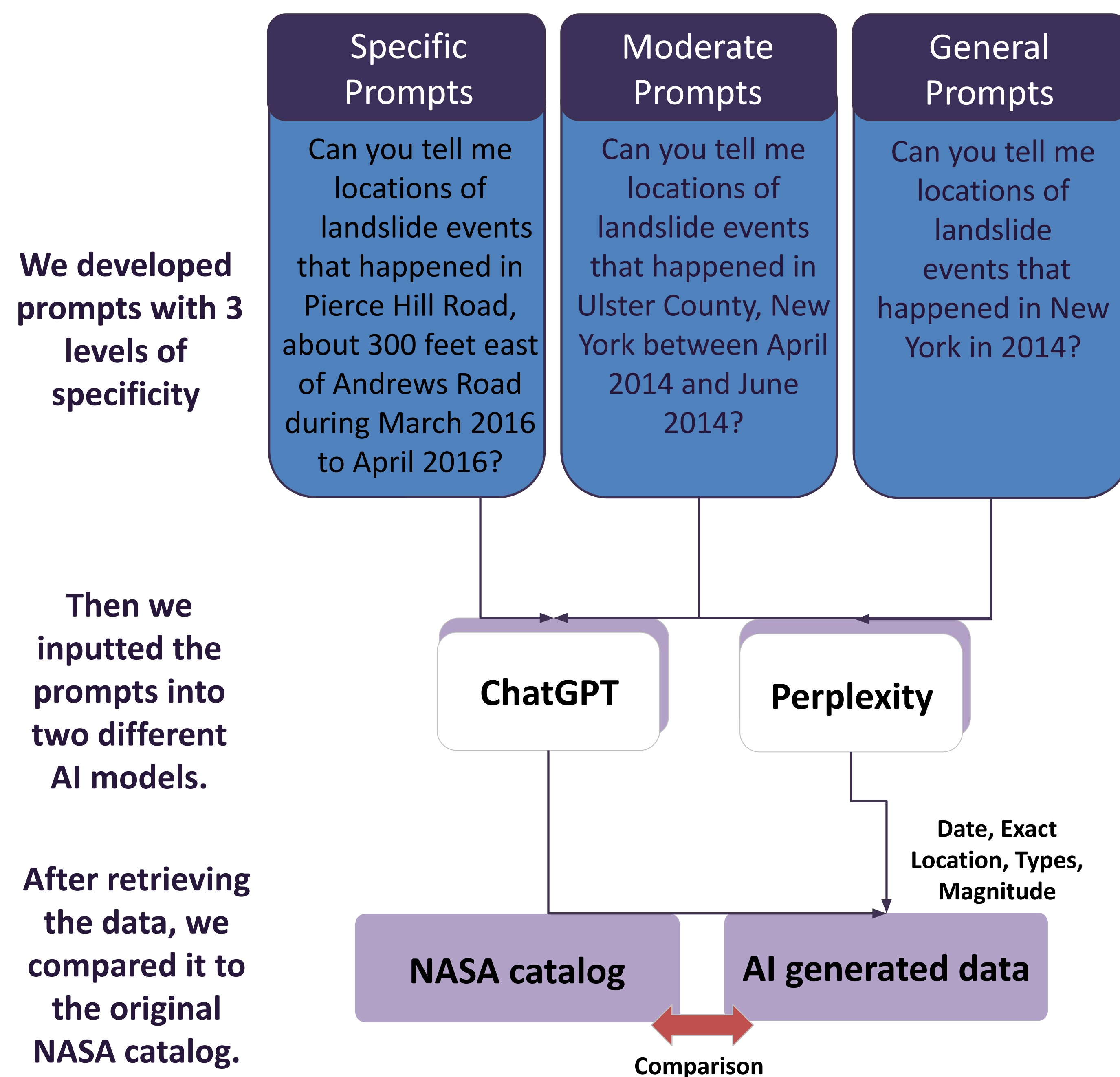

We developed prompts with 3 levels of specificity

Specific Prompts	Moderate Prompts	General Prompts
Can you tell me locations of landslide events that happened in Pierce Hill Road, about 300 feet east of Andrews Road during March 2016 to April 2016?	Can you tell me locations of landslide events that happened in Ulster County, New York between April 2014 and June 2014?	Can you tell me locations of landslide events that happened in New York in 2014?

Then we inputted the prompts into two different AI models.

ChatGPT   Perplexity

Date, Exact Location, Types, Magnitude

After retrieving the data, we compared it to the original NASA catalog.

NASA catalog   AI generated data

Comparison

Figure 1. Methodology Flow Chart

RESULTS

- This word map displays the key words that helped LLMs retrieve the landslide events.

Figure 2. Keywords Used by LLM's for Landslide Search

- The AI models were able to retrieve the most events from specific prompts with ChatGPT retrieving above 50% of NASA's events.

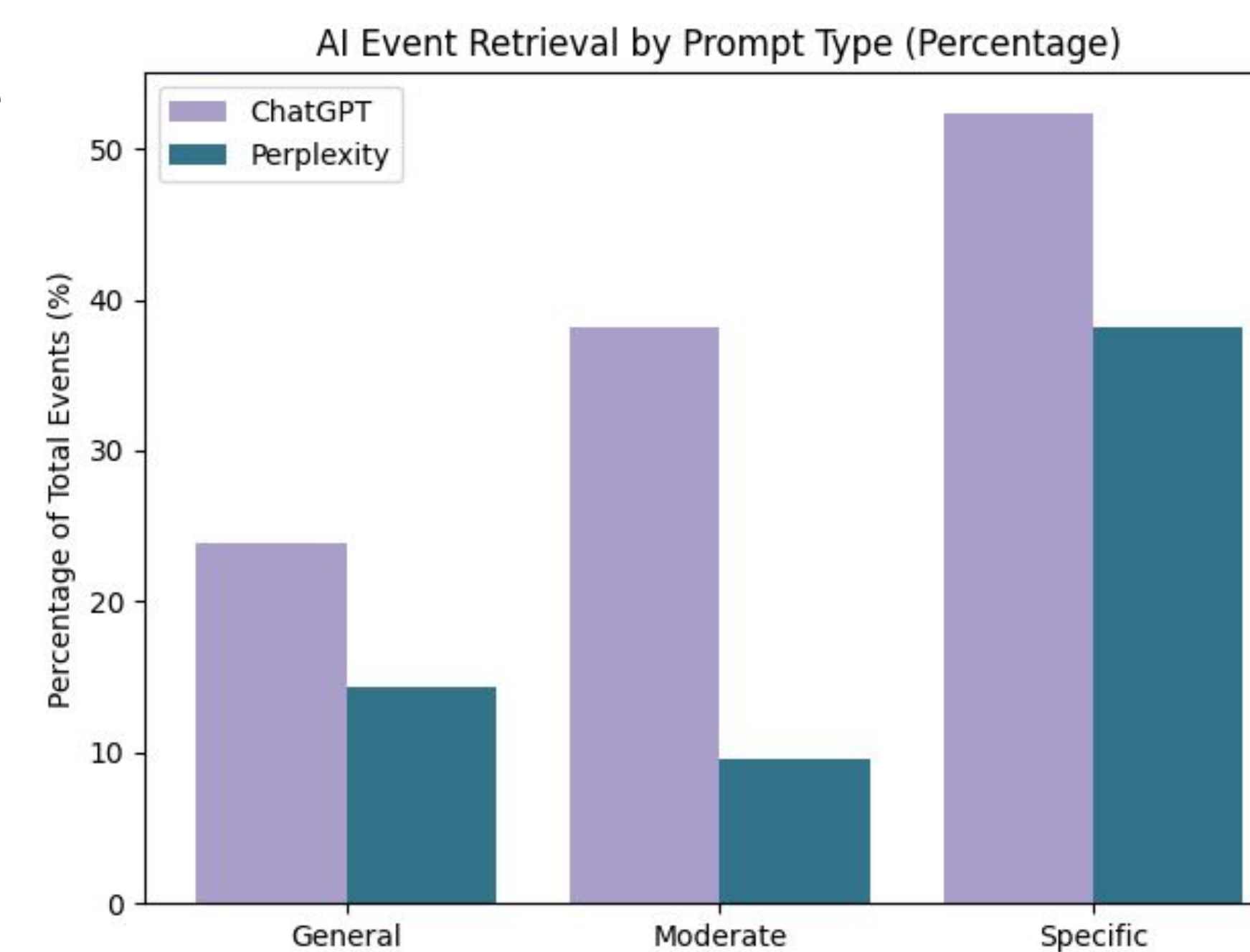- ChatGPT obtained more events than Perplexity in all prompt types overall.

Figure 3. Event Count Based off Prompt Specificity

- Out of the 21 landslide locations, ChatGPT matched 47.6% with NASA and mismatched 4.8%.

- Perplexity matched 28.6% locations and mismatched 9.5%.

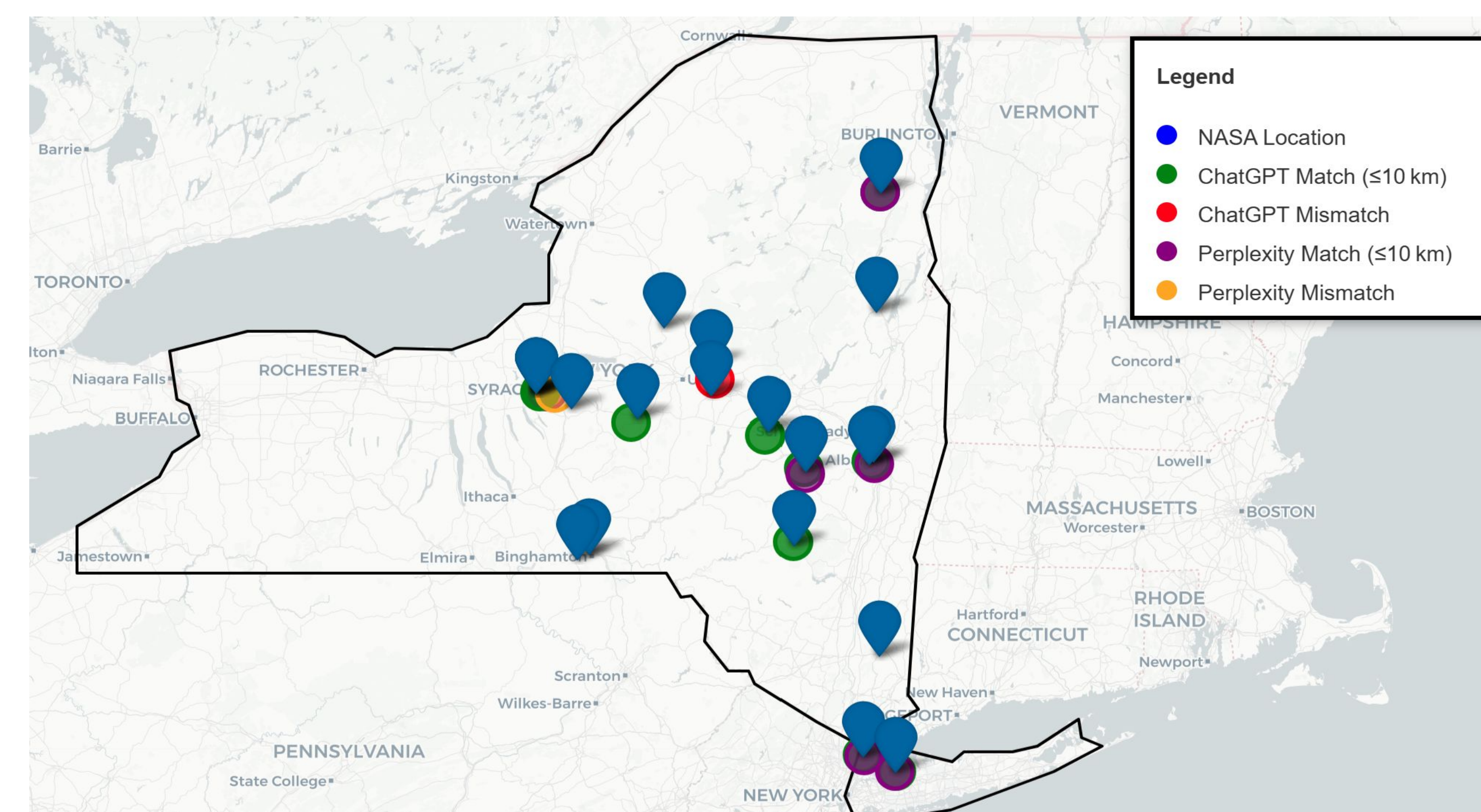- Both AI models matched locations at 28.6% and mismatched at 4.8%.

Figure 4. AI model location accuracy based on NASA catalog

(Results continued)

- Temporal analysis reveals 63.6% of ChatGPT's events matched with NASA, while 57.1% of Perplexity's events matched with NASA.

- Although ChatGPT was able to compile more data, Perplexity's dates were slightly closer to NASA's as Perplexity had an average difference of 1.25 days, while ChatGPT had a 1.3-day difference.
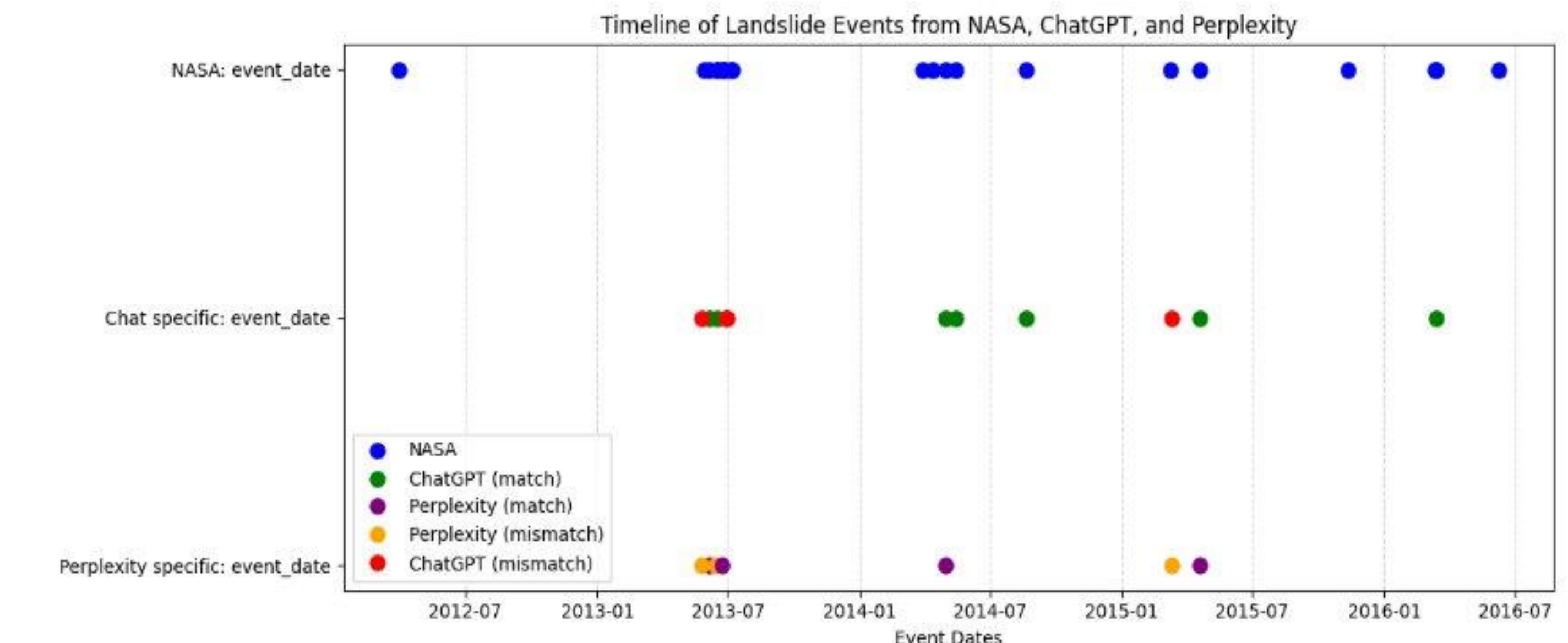

Figure 5. Accuracy of Landslide Events by AI based on NASA Event Dates

- No events reported by AI were completely different from NASA

- 14.3% events reported by AI match both spatial and temporal data with NASA

- 28.6% events reported by AI match only spatial data with NASA
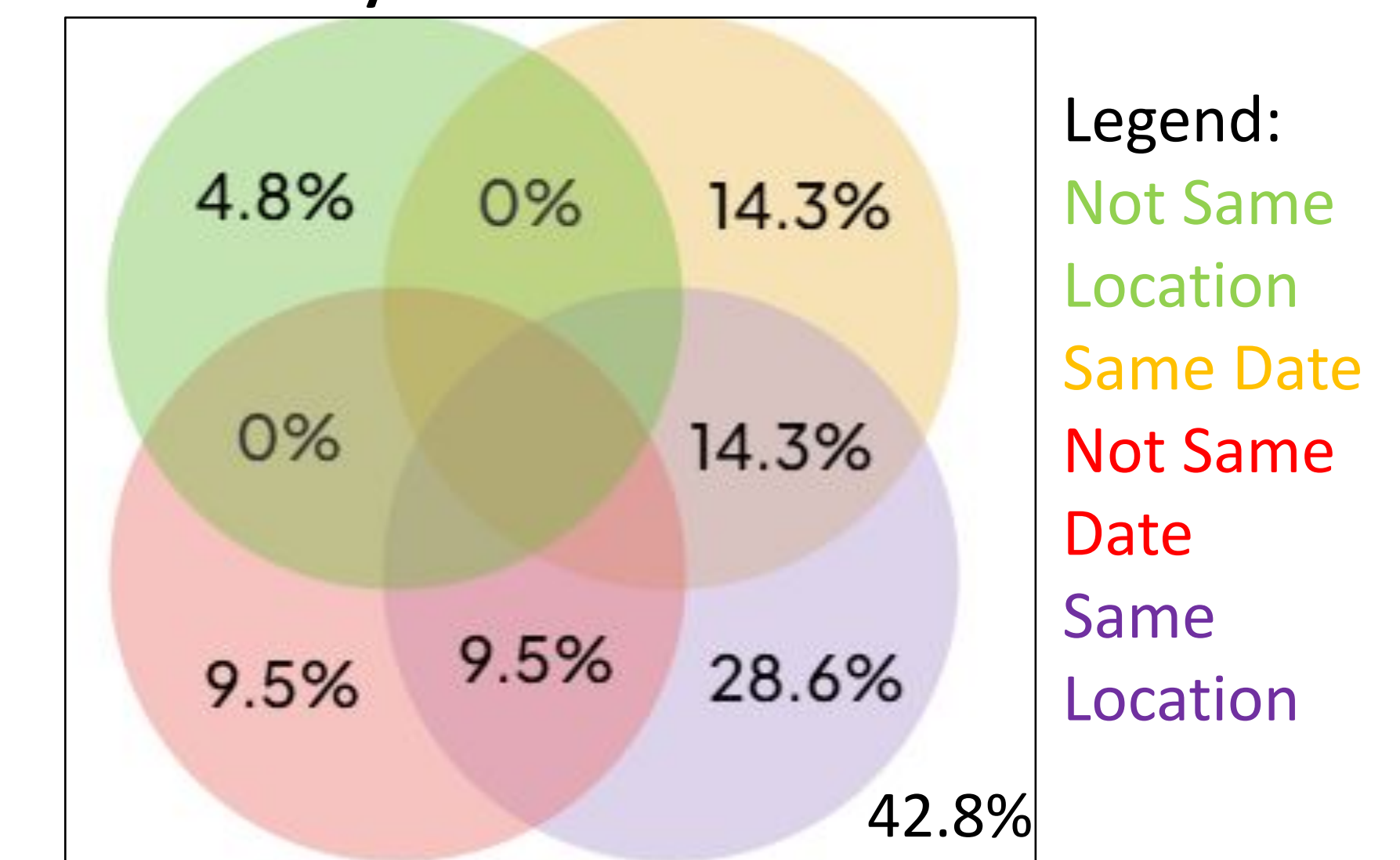

Figure 6. Comparing AI Reports of Spatial Data Against Temporal Data

CONCLUSION AND FUTURE STUDIES

- ChatGPT is better than Perplexity at reporting information on events overall.

- The more specific the prompt, the more data on events reported by ChatGPT.

- The AI models reported most location of the events accurately out of the ones found.

- There is certainty that the AI models did not report completely inaccurate data about landslide events.

- However, AI models only reported 14.3% of both spatial and temporal data accurately, which is relatively low compared to the sample size.

- Future studies should use ChatGPT Premium; free ChatGPT has limited tokens.

- Broader scope of time and location, or more recent sample events may impact the AI models' accuracy in reporting these events.

- Larger sample size of events and large trials need to be used to verify results.

REFERENCES

- NASA's Scientific Visualization Studio (2019). *Global Landslide Catalog* [Data set]. NASA. https://svs.gsfc.nasa.gov/4710