

Article

# An LLM-Based Inventory Construction Framework of Urban Ground Collapse Events with Spatiotemporal Locations

Yanan Hao <sup>1,2</sup>, Jin Qi <sup>1,2,\*</sup>, Xiaowen Ma <sup>3</sup>, Sensen Wu <sup>1,2</sup>, Renyi Liu <sup>1,2</sup> and Xiaoyi Zhang <sup>4</sup>

<sup>1</sup> School of Earth Sciences, Zhejiang University, Hangzhou 310027, China; haoyn@zju.edu.cn (Y.H.); wusensengis@zju.edu.cn (S.W.); liurenyi@zju.edu.cn (R.L.)

<sup>2</sup> Zhejiang Provincial Key Laboratory of Geographic Information Science, Hangzhou 310028, China

<sup>3</sup> Research Center for Intelligent Technology Standardization, Zhejiang Lab, Hangzhou 311121, China; maxw@zhejianglab.com

<sup>4</sup> School of Spatial Planning and Design, Hangzhou City University, Hangzhou 310015, China; zhangxiaoyi@hzcw.edu.cn

\* Correspondence: qijin@zju.edu.cn

**Abstract:** Historical news media reports serve as a vital data source for understanding the risk of urban ground collapse (UGC) events. At present, the application of large language models (LLMs) offers unprecedented opportunities to effectively extract UGC events and their spatiotemporal information from a vast amount of news reports and media data. Therefore, this study proposes an LLM-based inventory construction framework consisting of three steps: news reports crawling, UGC event recognition, and event attribute extraction. Focusing on Zhejiang province, China, as the test region, a total of 27 cases of collapse events from 637 news reports were collected for 11 prefecture-level cities. The method achieved a recall rate of over 60% and a precision below 35%, indicating its potential for effectively and automatically screening collapse events; however, the accuracy needs to be improved to account for confusion with other urban collapse events, such as bridge collapses. The obtained UGC event inventory is the first open access inventory based on internet news reports, event dates and locations, and collapse co-ordinates derived from unstructured contents. Furthermore, this study provides insights into the spatial pattern of UGC frequency in Zhejiang province, effectively supplementing the statistical data provided by the local government.

**Keywords:** large language model (LLM); event extraction; road collapse; event recognition; geo-hazard; spatiotemporal location

**Citation:** Hao, Y.; Qi, J.; Ma, X.; Wu, S.; Liu, R.; Zhang, X. An LLM-Based Inventory Construction Framework of Urban Ground Collapse Events with Spatiotemporal Locations. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 133. <https://doi.org/10.3390/ijgi13040133>

Academic Editors: Wolfgang Kainz, Peng Peng, Feng Lu, Shu Wang, Maryam Lotfian and Yunqiang Zhu

Received: 7 February 2024

Revised: 1 April 2024

Accepted: 15 April 2024

Published: 16 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, urban ground collapse (UGC) [1,2] events have become more frequent with the increased intensity of urban underground space construction, presenting a continuous expansion pattern parallel with urbanization; this has serious consequences for urban economic development and the safety of citizens [1,3]. The contradiction between underground space development and urban safety has intensified so much that it has become urgent that the government formulates better management and control strategies and conducts in-depth research on collapse risk [4–6]. News reporting is one of the important data sources for understanding and studying UGC events. Therefore, relevant researchers have collected different unstructured information sources, such as local news media and social network media, and then processed these to construct event databases [7]. Although the analysis of a single event can help to put forward specific measures for the prevention and control of ground collapse regarding collapse mechanisms [8,9], event databases can provide better management references through revealing the spatiotemporal patterns and complex causes of collapses [10,11]. Consequently, combining geographic information technology and text recognition technology to extract UGC events

from a large number of news reports and analyzing their spatiotemporal patterns can quantitatively evaluate geological risks at a regional scale, which is of great significance [12–14].

Event detection and monitoring have long been the focus of government public affairs management, as the timely understanding of the beginning and evolution of popular social events helps authorities to respond quickly [15,16]. Existing event extraction methods can be generally divided into three categories: pattern matching methods [17], machine learning methods [18], and deep learning methods [19]. Traditional pattern matching methods rely on manually set rules, whereas machine learning methods are based on statistical patterns. In comparison, deep learning models for event extraction have been more widely used in recent years. Examples of these applications can be found in construction accidents [20], improvement [21], football games [22], and so on. To date, large language models (LLMs) have achieved significant breakthroughs in numerous automated learning tasks through unsupervised training on extensive corpora and data sets using generative pre-trained transformers (GPTs). These tasks include machine translation, text classification, intelligent question-answering, and more, and they have even surpassed human performance (on average) in areas such as semantic understanding, common-sense reasoning, and dialogue generation [23,24].

Considering the fact that the urban ground collapse problem involves scattered, irregular distributions and complex characterizations in terms of events, LLMs are an excellent choice for the task of automating the construction of an inventory of UGC events, along with the statistical analysis of their spatiotemporal distribution [25]. Their natural ability to understand language helps LLMs to fully consider contextual words and sentence information [26], such that the extraction of entities and entity relationship is more accurate [27,28]. Through event extraction, historical events can be analyzed more scientifically to effectively prevent future collapse events. Another challenge in UGC data collection is the precise location and time of a collapse event, which has seen growing research interest in the GIS science field [29,30]. For instance, linguistic and text information have been explored using different information extraction methods to transform narratives into spatiotemporal data [31–33].

Therefore, to address the challenges of UGC events, this paper presents an LLM-aided framework that is used to crawl and extract unstructured information from news reports to gather structured information for inventory construction. In China, Zhejiang province is a typical region with a significant number of UGCs and other geological hazards. As in other provinces (except for Shenzhen) [34], UGCs in Zhejiang also occur due to a lack of management and accountability, with a lack of official data on historical events. Therefore, the data-processing framework discussed in this paper takes Zhejiang as the test area, but it is also applicable to other provinces. The contributions of this research are summarized as follows: (1) we propose an applicable framework for the recognition and extraction of news information with geographic and time information; (2) based on the framework, we construct the first UGC event database for the provincial level in China, which is representative and complete as the relevant news data were comprehensively manually checked. In addition, the model's precision and accuracy can be used as a benchmark for other researchers studying the identification of UGC events.

The rest of the paper is structured as follows: Section 2 introduces the geological background and geo-hazard events of Zhejiang province. Section 3 provides details of the methodology of our framework for UGC events extraction. In Section 4, the results of the case study of Zhejiang province are analyzed, and the precision and accuracy of our approach are calculated. Section 5 discusses the sources of error, as well as the differences between events of the same type and different types. The conclusions and future works are summarized in Section 6.

## 2. Study Area

Zhejiang province is located on the southeast coast of China, with mountains accounting for 74.6% of the land area and diverse landform types (Figure 1). It is a vibrant and economically dynamic province, known for its rich cultural heritage, picturesque landscapes, and significant contributions to the country's economic growth. Zhejiang has 11 prefecture-level cities, each of which is an administrative unit between province and county. In terms of population, Hangzhou is the capital city of Zhejiang province and ranks first with more than 10 million people. Wenzhou, Ningbo, Jinhua, Taizhou, Jiaxing, and Shaoxing are cities that have a population of more than 5 million, and the remaining four cities have a population of more than 1 million.

The topography of Zhejiang province is complex, and the terrain is in the form of steps from southwest to northeast (tilt). In the southwest of Zhejiang province, the mountains are more than a kilometer high and the highest peak, Huang-mao-jian, is 1929 m above sea level. Dozens of small and large red bed basins are scattered and distributed among these mountains [13]. The northeast plain is a part of the Yangtze River Delta. The outcrop layer in the north Zhejiang Plain mainly consists of quaternary loose deposits. Sedimentary clastic rocks and carbonate rocks are widely distributed in the mountainous area of northwest Zhejiang province. Pyroclastic rock series and granites are widely distributed in the south mountain area of southwest Zhejiang province. Continental clastic rocks are widely distributed in the hilly basin of central Zhejiang province, and volcanic rock series form a low mountainous area around the basin. The hilly and mountainous areas along the southeast coast of Zhejiang province are mainly volcanic series, and the plain areas are quaternary loose sedimentary layers. The Zhoushan and coastal islands consist of volcanic rocks and granites.

The sudden geological disasters in Zhejiang province consist of four main kinds of mountain disasters [35,36]—landslides, debris flows, collapse, and ground collapse—which are mainly distributed in the hilly and mountainous areas of northwest, southwest, central, and southern Zhejiang [37]. Land subsidence is the most important geological hazard in Zhejiang Plain, and is the most serious in Hang-jia-hu Plain and Ning-Feng Plain.

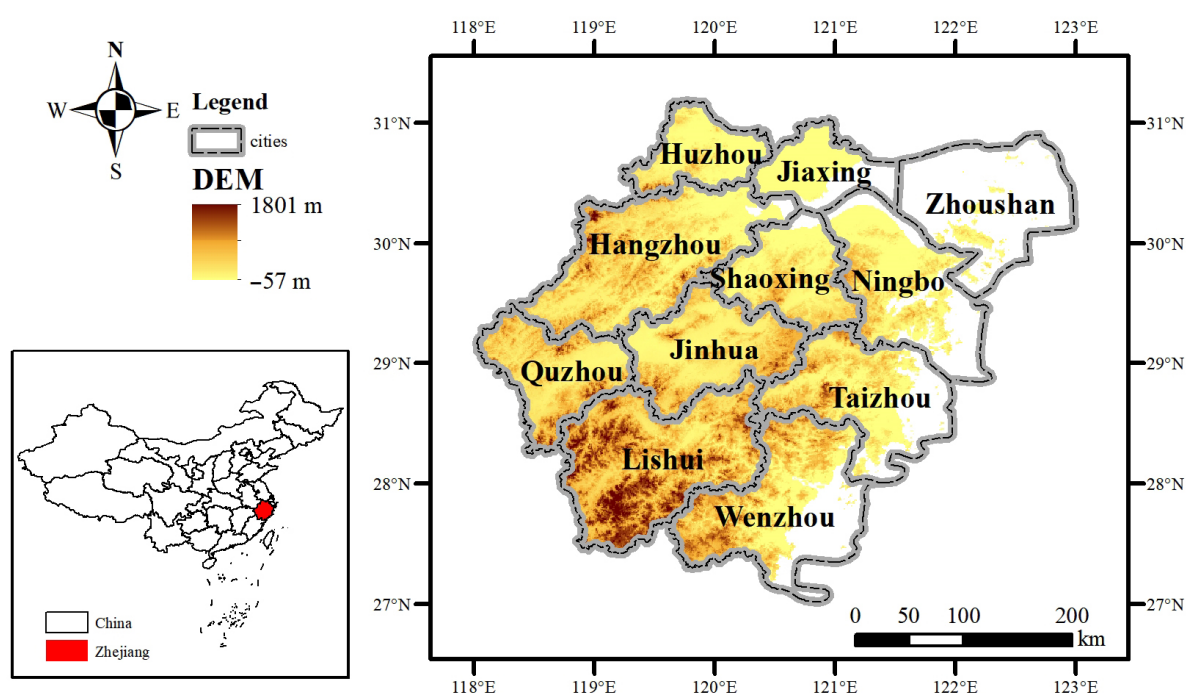
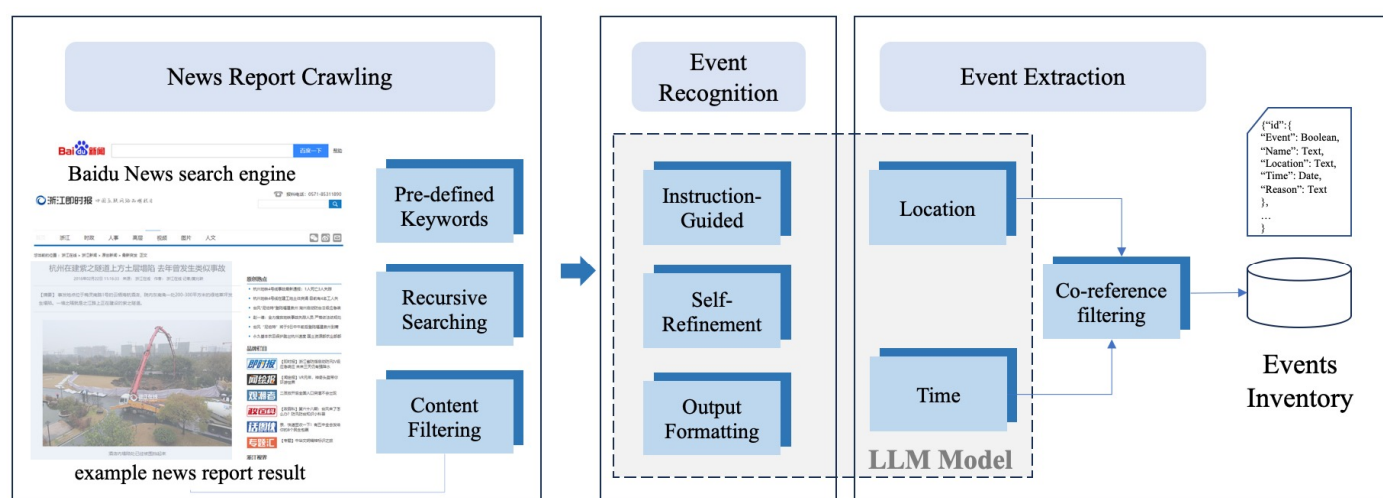


Figure 1. Map of Zhejiang province, China.

Zhejiang province is one of the most economically developed provinces in China. Rapid economic development in the area has been accompanied by rapid urbanization, and the exploitation and utilization of underground space are becoming more and more frequent. At present, four cities in Zhejiang province, including Hangzhou, Ningbo, Shaoxing, and Wenzhou, have metro systems, while the other seven cities are still in the planning stages in this aspect. The scale of human engineering activities is huge, which not only brings huge benefits to humans but also produces many serious environmental engineering geological problems, such as surface subsidence and ground collapses caused by underground mining. This is also a significant feature of UGCs, which differs from collapse events in traditional geological hazards. A typical city is Hangzhou, located in the eastern coastal area, which has experienced severe UGC events [2,38]. Mountain tunnelling [39] and metro system construction [40,41] have contributed to the occurrence of UGC events in Zhejiang province.

### 3. Materials and Methods

In order to implement the extraction of UGC event information, we integrated LLM technologies into a framework (Figure 2). The framework primarily comprises three modules: the news report crawling module, which can recursively request a news search service and collect news content; the event recognition module, which can identify and categorize events within news reports; and the event extraction module, which extracts specific arguments or relationships from unstructured text and formulates the report into an events inventory. The LLM technology mainly participates in the latter two modules to assist in the task of structuring event information.



**Figure 2.** The framework for UGC inventory construction.

#### 3.1. News Report Crawling

The news report crawling module serves as the foundational component in the UGC inventory construction framework. Its primary function is to collect relevant information from the internet and prepare the necessary content for subsequent analysis. This module functions through a web crawler designed to download web pages from pertinent links. Through employing HTTP requests with predefined keywords such as ‘daolu’ (road), ‘dimian’ (ground), ‘taxian’ (collapse), and specific city names, the news search engine effectively retrieves any pertinent web links. Each request is tailored to a specific city, initiating a targeted search to obtain a comprehensive list of web links. During the request and crawling process, we employed a recursive mechanism to traverse web links and gather comprehensive information. This iterative approach allows us to delve deeply into each web page, extracting not only the provided title, thumbnail, and abstract from the news

engine but also accessing the original webpage to gather additional detailed information. This process enables us to obtain a more thorough understanding of news content, including the media sources and news supplementary details that may not be readily available through the initial news engine interface.

Upon accessing the original webpage, various elements are encountered, including the logo, news media name, more news images, advertisements, related news links, and in some cases, a discussion board. In order to extract the true content of the news report from this diverse array of elements, a content filtering algorithm must be implemented. This algorithm is crucial for discerning and isolating the core news content from the surrounding web page elements, ensuring that the retrieved information accurately represents the news report itself. In our model, a simple noise removal strategy is used according to the structure and pattern of the webpage. Due to the concentrated and lengthy nature of text content in news reports, a contextual window with a distance of  $n$  lines can be utilized to compute the average number of characters and the variance of characters in each line. In doing so, any text within the window that exhibits an average character count lower than a predetermined value and variance exceeding a specific threshold can be identified as noise and subsequently filtered out.

### 3.2. LLM-Aided Event Recognition

When utilizing news search engines, a wide array of web pages is returned, encompassing news reports, event briefings, rescue tracking reports, and other related content. However, among these, less relevant content, such as disaster prevention policy comments or disaster prevention product introductions, may also be included. Leveraging LLM technology, we aimed to identify the most relevant instances of UGC news reports and subsequently transform this information into a structured data format. This process aligns with the classic task in the natural language processing (NLP) field known as event extraction. Event extraction involves two primary subtasks: event-type detection and event-argument extraction. Event-type detection focuses on identifying the core words that signify the occurrence of an event.

An LLM learns human natural language logic and background knowledge from large amounts of text data through pre-training, gaining the ability to understand and generate human language. Prompt engineering is needed to guide the LLM when solving complex problems in specialized fields. It refers to the process of the strategic design of the LLM input–prompt word (prompt), which usually needs to specify the LLM task role, task conditions, and task-related expertise. The prompt has obvious characteristics, such as zero-shot or a few examples. In this framework, we used ChatGLM-3b [42], a deep learning model based on transformer architecture, to solve the event extraction task. ChatGLM-3b is widely used in the construction of natural language dialogue systems and has the characteristics of smooth dialogue. Therefore, when designing a prompt, we can take advantage of the interactivity and contextual guidance of ChatGLM-3b.

First, we asked the LLM model some questions about UGC events, and then we further refined or adjusted the extraction requirements based on the model's answers. Second, the appropriate examples were selected for contextual information about a UGC event to help the model better locate and identify the event. Finally, prompts (Figure 3a) were designed based on (1) making the task objective clear through specifying (in the prompt) that the goal of the task was to determine whether a given news report text was a collapse event and (2) providing an example; in order to help the model better understand the task, a news report of a UGC was given to present the model with a true result. At the same time, a news report about a non-UGC was given to present the model with a false result. (3) For output formatting, we specified (in the prompt) whether the output result was true or false.

**(Event Extraction, Prompt 0)**

For the task of "UGC Event Recognition", given a UGC sample and a non-UGC sample, one should identify whether a similar event can be recognized from the content.

Given the following list of samples:

```
{{shuffle(selected_sample[learn['event_type']].values()) | join("\n- ")}}
identify related events "{{chat[content ']'}}
({{selected_sample[learn['event_type']]}})" in the following report:
{{content}}
Extractions: ||| {{output_format(results, "") | join(" ")}}
```

(a) Event recognition prompt.

**(Event Extraction, Prompt 1)**

For the task of "UGC Argument Extraction", given a UGC sample with its arguments, one should extract the same arguments from the content.

In the following passage, an argument with the type "{{UGC\_arg}}" is related to the event "{{selected\_sample[learn['event\_type']].values()}}":

{{content}}

The argument should be (copy from the context if you find it; if not, do not generate): ||| {{Output\_Json(filter\_type(results, query\_arg))}}

(b) Event argument extraction prompt.

**Figure 3.** Prompt templates for the LLM model.

### 3.3. LLM-Aided Event Extraction

The final process is an event argument extraction task in order to identify and extract specific components or elements related to an event from a given text. These attributes often include entities such as the participants (subjects and objects involved in the event), time expressions, locations, and other relevant details. The task can be approached using techniques such as named entity recognition (NER), part-of-speech tagging, syntactic parsing, and semantic role labeling. In our framework, we also used ChatGLM-3b to extract the event attributes, including the name of the event, when and where it occurred, and the area and depth of the collapse. The design of the prompt is similar to the event-recognition process, where a collapse report and its associated attributes and values are provided as a demonstration case, and the model is asked to return results in a specified format (Figure 3b).

Time is an ambiguous and complex concept that can be expressed in many ways, such as absolute time (1 May 2023) and relative time (yesterday, next year, etc.), which presents a challenge for time extraction. The LLM model can better deal with the ambiguity of time extraction through understanding context, normalization, developing an understanding of a time reference frame, multi-modal information fusion, and knowledge introduction.

Another challenge is the geographical location extraction, including geographical name, location description, etc. The name of the place may be ambiguous, for example, "Beijing" can be Beijing City, a place within Beijing City, or a specific neighborhood. In our framework, we utilized a geocoding API to translate the place name into the corresponding geographic coordinate information [43].

An important post-processing task after event attribute extraction is co-reference filtering. Co-reference refers to the phenomenon of multiple news reports or multiple sentences in a news article referring to the same subject. Co-reference problems will affect the accurate extraction of event information, so it is necessary to conduct co-reference ablation. A variety of methods can be used for co-reference filtering. In this framework, we

adopted a rule-based approach. As the UGC event is accidental, the co-reference relationship can be identified by judging the spatiotemporal similarity of different extraction results. The events for which the spatiotemporal similarity value is higher than the threshold can be merged through formulating a similarity calculation policy for event attributes.

#### 4. Results and Analysis

##### 4.1. UGC Event Recognition

By crawling the response contents using keywords from the Baidu news portal, we obtained 637 links in total (Table 1). Hangzhou had 220 news pieces, Wenzhou had 79 news pieces, and Ningbo had 83 news pieces, all ranked in the top three, while the number of response content pieces from other cities had an average number of 31. There are great differences in the number of news events reported in different cities. On the other hand, although Baidu's news search engine was used, it can be seen that in the results returned by the search engine in Table 2, except for the news from Hangzhou, 61.3% of the proportion were reports of collapse events. For the news from the remaining 10 prefecture-level cities, less than 15.7% covered real collapse events. Only a small proportion of the news reports warranted further content extraction.

**Table 1.** Results comparison between LLM model and manual checking.

City Name	News Counts	LLM Result			Manual Result				Model Evaluation		
		True	False	Fail	True	Percentage	False	Percentage	R	P	F1
Hangzhou	220	127	43	50	135	61.3%	85	38.7%	74.07%	78.74%	76.34%
Huzhou	30	15	12	3	0	0%	30	100%	0.00%	0.00%	0.00%
Jiaxing	40	19	11	10	4	10.0%	36	90.0%	100.00%	21.05%	34.78%
Jinhua	35	16	11	1	0	0%	35	100%	0.00%	0.00%	0.00%
Lishui	26	9	9	8	3	11.5%	23	88.5%	66.67%	22.22%	33.33%
Ningbo	83	32	36	15	3	3.6%	80	96.3%	100.00%	9.38%	17.14%
Quzhou	19	10	7	2	3	15.7%	16	84.2%	66.67%	20.00%	30.77%
Shaoxing	32	11	13	8	3	9.3%	29	90.6%	66.67%	18.18%	28.57%
Taizhou	49	22	11	16	6	12.2%	43	87.8%	50.00%	13.64%	21.43%
Wenzhou	79	43	22	14	5	6.3%	74	93.7%	40.00%	4.65%	8.33%
Zhoushan	24	7	17	0	0	0.0%	24	100%	0.00%	0.00%	0.00%

**Table 2.** The precision of event attribute extraction on the LLM model.

City Name	True Positive	Time			Location			Model Precision
		Right	Wrong	Fail	Right	Wrong	Fail	
Hangzhou	100	34	57	9	49	42	9	34%
Huzhou	0	0	0	0	0	0	0	0
Jiaxing	4	2	2	0	4	0	0	50%
Jinhua	0	0	0	0	0	0	0	0
Lishui	2	1	1	0	2	0	0	50%
Ningbo	3	3	0	0	3	0	0	100%
Quzhou	2	1	1	0	2	0	0	50.0%
Shaoxing	2	0	2	0	2	0	0	0%
Taizhou	3	1	1	1	1	1	1	33.3%
Wenzhou	2	1	1	0	1	1	0	50.0%
Zhoushan	0	0	0	0	0	0	0	0

The computational power required by the LLM prompt and inferencing is not large, and a server with an NVIDIA A100 GPU was used for event recognition and extraction.



In order to measure the performance of the LLM model, we calculated the recall, precision, and F1\_score. *Recall* is a measure used to evaluate the ability of a model to identify all relevant instances within a dataset. It is also known as the true-positive rate. *Precision* is a measure of the accuracy of the positive predictions obtained from the model. The *F1\_score* measures the overall correctness of the model's predictions. They can be calculated using Equations (1)–(3):

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

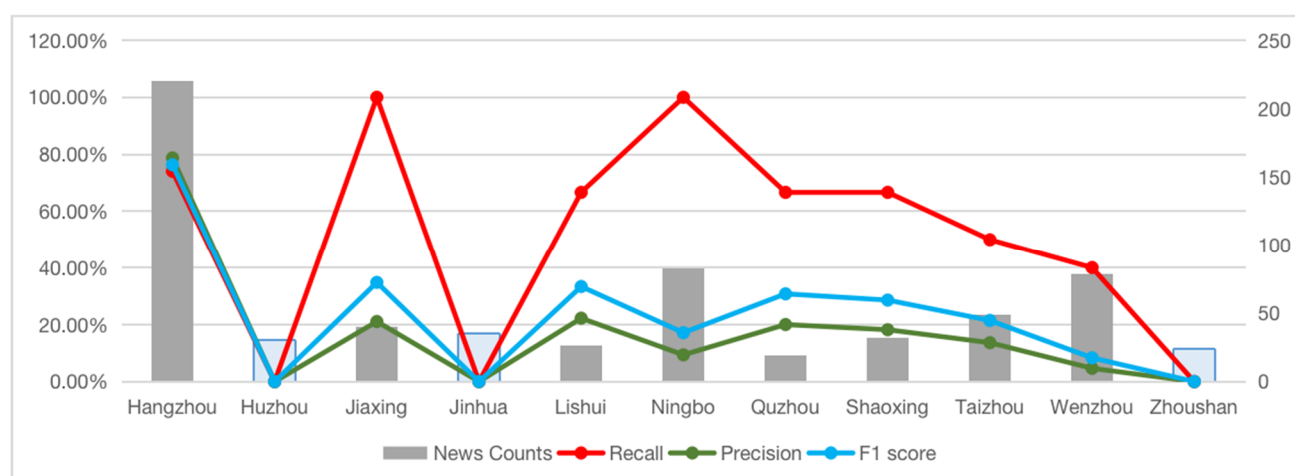
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1\_score = \frac{2 \times recall \times precision}{recall + precision} \quad (3)$$

where true positives (*TP*) are the number of correctly identified positive instances; false negatives (*FN*) are the number of positive instances that were incorrectly classified as negative; false positives (*FP*) are the number of negative instances that were incorrectly classified as positive; and true negatives (*TN*) are the number of correctly identified negative instances.

The results of our UGC identification model were compared with the manually checked results. If the UGC identification model and the manual check marked the news as a true event, then it was regarded as *TP*. However, if the result after manual checking was not a UGC event, then it was regarded as *FP*. If the UGC identification model marked the news as a false event, and so did the manual result, then it is regarded as *TN*; otherwise, it is regarded as *FN*. If the UGC identification model did not give a formatted result, then it was labeled fail. After manually checking the results of each city in Zhejiang province, we obtained the evaluation results of our UGC identification model, as can be seen in Table 1.

The result can also be seen in Figure 4. The curves represent the recall, precision, and F1 scores in red, green, and blue, respectively. Apart from Huzhou, Jinhua, and Zhoushan, which all had no UGC events, the majority of the F1 scores and precision values are relatively low (under 35%), which means our model is not an ideal tool for recognizing UGC events directly. However, the recall curve is significantly higher than the other two curves. There are two cities with 100% and four cities with more than 60%, which indicates that the LLM model can do well in terms of figuring out relevant events.



**Figure 4.** Evaluation of UGC event recognition model for each city.



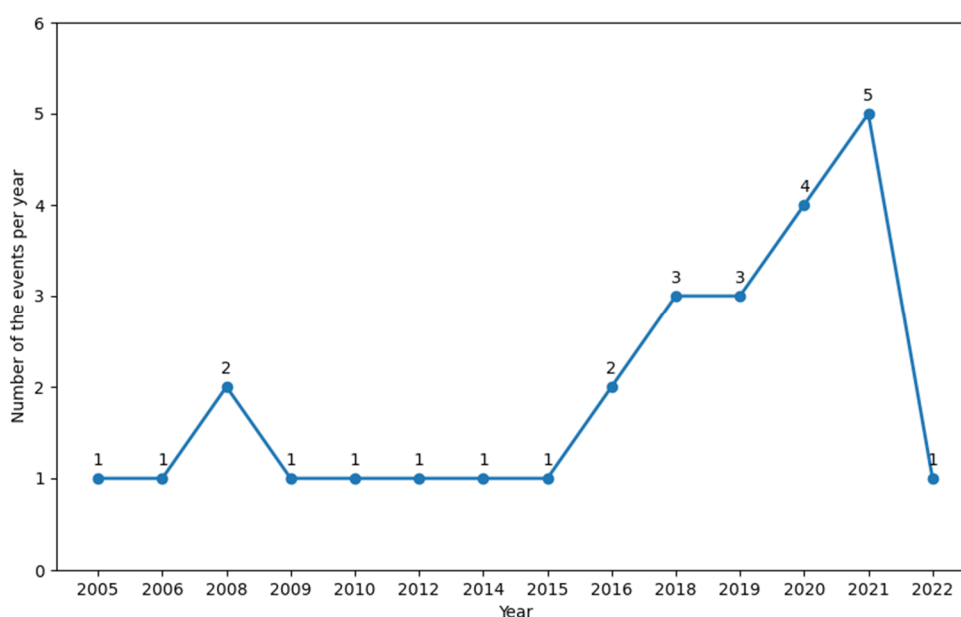
#### 4.2. UGC Event Extraction

The evaluation of the event attribute extraction results mainly sought to understand the correctness of the extraction of event occurrence time and place. During the precision calculations, if any attribute in time or place was extracted incorrectly, the extraction failed. In the case of time, the comparison is accurate to the day. For location, the comparison is appropriate for city names. The final results are listed in Table 2. Similarly, if the extraction result did not conform to the required JSON format, it was regarded as fail.

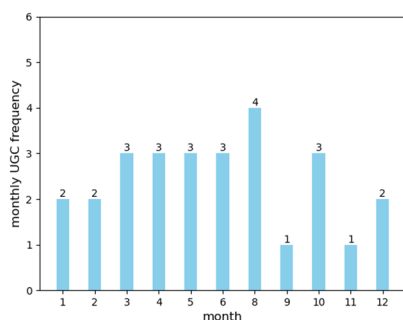
As there was only a single case of UGC in each city except for Hangzhou, the extraction results for Hangzhou can be considered as representative of the overall model performance. As one can see, the accuracy of time recognition is very low: it is only 34%. In contrast, location recognition was correct 49% of the time. This is due to a certain lag in news reports, and the time of the event is often expressed as a relative date, such as yesterday, the day before yesterday, or this afternoon, which leads to the incorrect extraction of the event time by the model; it is easy for the model to confuse the time of the news report with the time of the news event, resulting in extraction errors. As news reports have certain timeliness, when designing time filtering rules, events within 3 days can be defined as news entities with time similarity. Due to the sporadic nature of ground collapse events, very few events will occur in the same place and the same city (at present, only two collapse events occurred on the same day in Hangzhou). Therefore, the definition of place similarity rules according to the combination of city names and county names can effectively distinguish most co-referential events.

#### 4.3. UGC Inventory

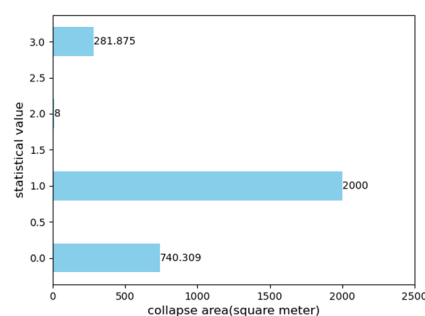
The inventoried UGC events correspond to the period of May 2005–May 2022 and consist of 27 events with known locations and dates derived from recognition results, attribute extraction results, and co-reference filtering results (Table 3). Annual and monthly UGC frequency (number of events per year and per month) is depicted in Figure 5a and b, respectively, and a statistical representation of the known collapse area for the UGC inventories is given in Figure 5c.



(a) Annual frequency of occurrence.



(b) Monthly frequency of occurrence.



(c) Morphometric characters.

**Figure 5.** Statistic characteristics of UGC events from 2005 to 2022.

The curve in Figure 5a is characterized by an increasing trend, especially a subsequent abrupt increase in the last decade. Since 2015, UGC events have accounted for 70.4%, which is similar to the conclusion of Chen's study [25]. Many studies have indicated that the collapse frequency in the rainy season is higher than that in the non-rainy season. Generally, the rainy season in China refers to April to September in Zhejiang province. Figure 5b shows the monthly distribution of ground collapse accidents in Zhejiang province from 2005 to 2022. Approximately 51.8% of the 27 statistical collapse events occurred from April to September, which is close to the proportion of the number of collapse events during the rainy season in China at 46.7% [2]. Figure 5c gives the statistical characteristics of the collapse area of UGC events; although most of the literature agrees that the area of urban ground collapse is smaller than that of traditional collapse [25], it can be seen that there are still large-scale collapses caused by 2000 square meters of engineering construction.

Of the 27 recorded events, 10 contained information about the collapse area, and six contained information about collapse depth. From the point of view of the collapse area, all of the events are small collapses of less than 1 square kilometer. The length of the collapse is greater than 30 m, indicating a tendency towards elongated forms. In the development and utilization of underground spaces, 0 to 3 m belong to the surface layer, and 3 to 15 m belong to the shallow layer. From the perspective of collapse depth, only one case of pavement collapse caused by heavy rain was 0.5 m deep, and the rest were shallow. Most of the collapses were caused by subway construction (six cases) and engineering foundation pits (six cases), followed by poor engineering quality (five cases) and underground pipeline leakage (three cases). These factors are all related to human activities, especially engineering construction and underground development projects. Therefore, measures to prevent UGC should focus on improving the standardization of engineering and construction projects, including strict supervision and code implementation, high-quality engineering design and construction, and comprehensive risk assessment and monitoring.

**Table 3.** UGC events that occurred in the period 2005–2022 in Zhejiang province.

No.	Date	Lon.	Lat.	City	Location	Factor	Length (m <sup>2</sup> )	Area (m <sup>2</sup> )	Depth (m)
1	18 May 2022	120.313	30.309	Hangzhou	Jinshahu Park Sunken Plaza	Pipe burst	\	\	\
2	30 August 2021	120.130	30.271	Hangzhou	Southeast corner of Academy Road and Tianmushan Road	\	\	12	\
3	29 December 2020	120.125	30.269	Hangzhou	Xixi Road Sidewalk	Subway construction	\	25	\
4	30 June 2020	120.153	30.161	Hangzhou	Dongguan Road Puyan Road Farmers' Market, Binjiang District	Heavy rain	\	8~9	0.5
5	30 June 2020	120.328	30.318	Hangzhou	Intersection of Shangsha Road and Xueyuan Street, Xiasha	Heavy rain	\	\	\
6	2 October 2021	120.280	30.442	Hangzhou	Intersection of Heyu Road and Xinzhou Road, Linping District	Collapse of foundation pit	\	8	4
7	15 November 2008	120.234	30.168	Hangzhou	Xianghu Station of Metro Line 1	Collapse of foundation pit	75	\	15
8	21 March 2021	120.183	30.311	Hangzhou	Resettlement housing site of ShenjiaNanyuan, opposite to Dongxinyuan, Changbang Road, Xiacheng District	Collapse of foundation pit	40	\	\
9	28 August 2019	120.180	30.268	Hangzhou	Bao-shan-qiao to North Jianguo Road station at Metro Line 5	Subway construction	\	\	\
10	16 August 2019	120.182	30.244	Hangzhou	Under the approach bridge to Hangzhou City Station Railway Station	Subway construction	\	\	\
11	7 May 2021	120.072	30.330	Hangzhou	Jinjiadu to Zijinhang Road Station Interval at Metro Line 4	Subway construction	\	\	\
12	11 June 2018	120.064	30.347	Hangzhou	Yunhe Lake construction project site in the north of SanDun	Collapse of foundation pit	\	2000	\
13	7 December 2018	120.180	30.252	Hangzhou	In front of KuiXiang Community, Jianguo Middle Road	Subway construction	\	\	\
14	21 April 2016	120.129	30.282	Hangzhou	Intersection of WenEr Road and XueYuan Road	Rainwater pipes leaking	\	20	\
15	22 February 2016	120.102	30.175	Hangzhou	Yunqi HNA Hotel, No. 1, Meiling South Road	Road construction	\	250	\
16	19 October 2015	120.162	30.339	Hangzhou	ChuXin Road No. 31	Collapse of foundation pit	\	10	3
17	1 April 2014	120.207	30.240	Hangzhou	ChengXing Road Station at Metro Line 4	Subway construction	\	\	\
18	25 February 2009	120.200	30.282	Hangzhou	Pu-jia-qiao section on the Canal Walking Trail	Engineering quality	\	\	\
19	7 January 2008	120.184	30.277	Hangzhou	GenShan Canal Bridge	Water pipe burst	\	30	3~4
20	13 March 2012	120.096	30.294	Hangzhou	Project site of Wuzhou International Plaza at the intersection of Gudun Road and Yuhangtang Road	Collapse of foundation pit	\	\	\
21	26 April 2021	120.569	30.045	Shaoxing	Intersection of Guandu Road and Henghu Road	Engineering quality	\	\	\

22	14 January 2018	120.575	30.047	Shaoxing	Jiayuan Plaza	Engineering quality	\	\	\
23	18 October 2010	119.888	28.982	Jinhua	Former Dongfeng Perspex Company in Wuyi County	Abandoned mines	\	\	\
24	10 September 2020	120.610	27.960	Wenzhou	Shanghui Road, Louqiao Street	Engineering quality	\	\	\
25	10 March 2019	121.631	29.883	Ningbo	Gaoxin District Baolong Plaza	Engineering quality	30	150	\
26	9 May 2005	120.938	30.040	Ningbo	Xiao-Yong Railway between Yuyao West and Yiting Interval	Earth cutting	150	\	10
27	9 August 2006	118.756	28.831	Quzhou	Houxi Street, Jiangshan City, Quzhou, Zhe-Gan Line	Geological cave collapse	\	\	\

Further, we conducted a distribution analysis of UGC events within the cities of Zhejiang province, considering both the events with known co-ordinates (Figure 6a) and the number of occurrences (Figure 6b). Using the kernel density model, the spatial distribution characteristics were reclassified into five different levels, from green to red, using the natural breaks classification method [1]. The core of the risk is concentrated in the eastern part of Hangzhou, while the rest are scattered in different prefecture-level cities and were not gathered together. As a rapidly developing city, Hangzhou's continuous construction activities may have had an impact on the stability of its underground structures. This effect may come from human activities such as subway construction, real estate development, and infrastructure construction. Another possible reason is that there may be a certain bias in the reports of internet news media, which pay more attention to Hangzhou than other cities, resulting in more concentrated coverage of its collapse events.

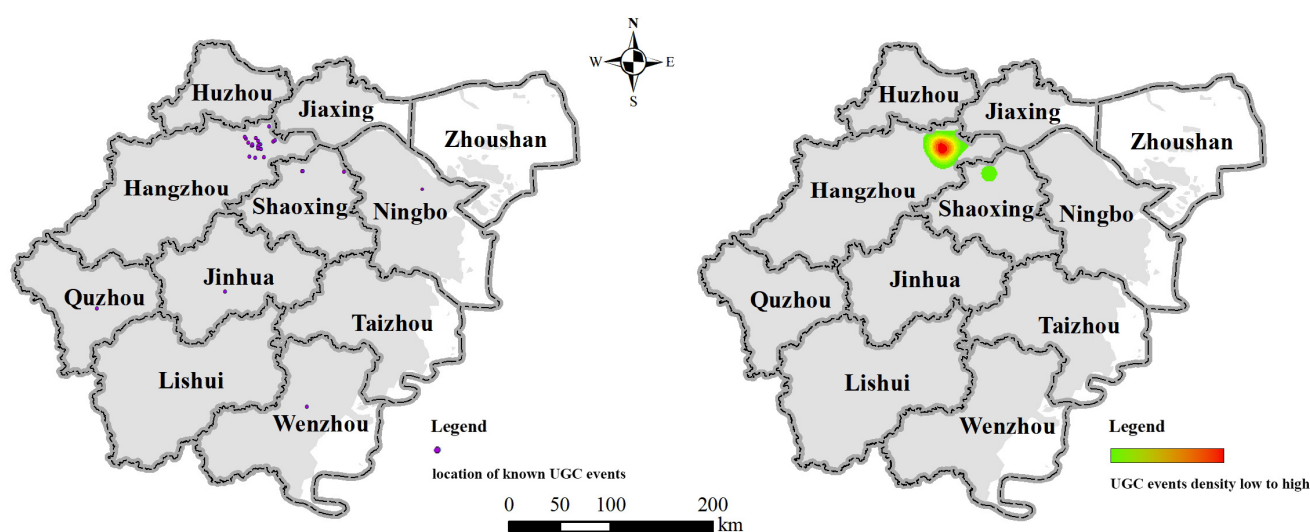


Figure 6. Probability density risk evaluation results.

## 5. Discussion

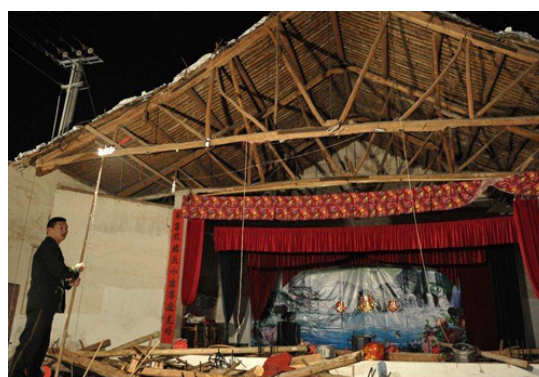
### 5.1. Comparison with Collapse-Related Accidents

In order to figure out the causes of a low accuracy rate but a high recall rate, we further analyzed the false-positive events and found that these misjudgments are basically caused by the misclassification of events. The most commonly misidentified events include tower crane collapse (and scaffold collapse), building collapse (especially the collapse of dangerous buildings), bridge collapse, and viaduct collapse, which are shown in Figure 7. Take Wenzhou as an example: 43 out of 79 reports were judged to be UGC events, but only 5 of them were true positives, and the remaining 38 (accounting for 88.3%) were all other collapse events, most of which were old residential buildings.

These false-negative events are mainly caused by semantic ambiguity, which is also a big challenge in NLP tasks. In some cases, the term *tanta* is used in news reports, which is semantically similar to *taxian* (collapse). In some cases, the two terms may be used interchangeably or as synonyms in news reports or related content, potentially leading the search engine to return content containing the keyword *tanta* as relevant reports for *taxian* (collapse). Another reason is the relevance algorithm of the search engine might consider various accidents and disasters related to UGC, some of which may involve collapses of buildings or structures. It is a fact that ground collapse events always lead to damage to buildings, structures, municipal facilities, and traffic facilities. As a result, search results may include these related incidents.



(a) Dacheng Road, Ningbo; 4 April 2014.



(b) Yazhuang Village, Jinhua; 13 February 2014.



(c) Jianguo Village, Jiaxing; 2 September 2008.



(d) Qiutao Road, Hangzhou; 18 May 2019.

**Figure 7.** Examples of false-positive events (collapse-related events).

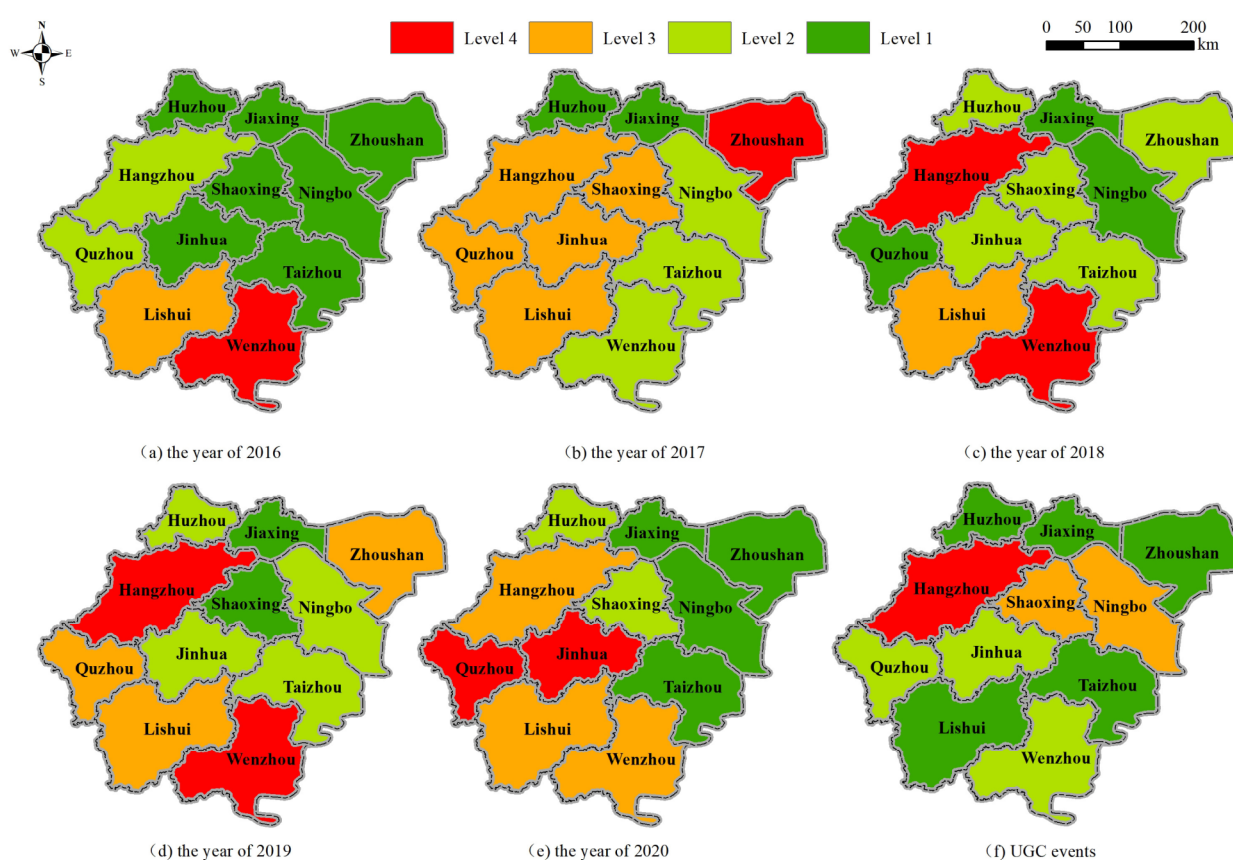
In summary, the misjudgments of our LLM model mainly arise from the combined effects of semantic ambiguity and relevance algorithms. In order to more accurately recognize and classify UGC events, further optimization of the LLM model and processing framework is needed. As we have not yet collected a sufficient number of samples for easily confused events, fine-tuning the LLM model is not an appropriate approach at present. A possible solution to this may be to integrate it with a knowledge graph; that is, the semantic representation of the output of the language model is associated and integrated with other semantic information, such as entity, relation, and attribute, in a knowledge graph. Different fusion strategies can be adopted in this step, such as rule-based fusion and similarity-based fusion. Through integrating the language model output and knowledge graphs, the richness and accuracy of semantic representation might be further enhanced. At the same time, the relational and attribute information in a knowledge graph can be used for inference and relational inference, so as to obtain a more complete and accurate classification result.

## 5.2. Comparison with Other Geo-Hazards Events

Zhejiang Natural Resources and Statistical Yearbook on the Environment [44] is an annual gathering of comprehensive statistical data that are jointly compiled by the Zhejiang Provincial Bureau of Statistics and other relevant departments and bureaus. It reflects the basic situation of Zhejiang's environment in various fields and summarizes and counts the relevant information on geological disasters, including landslides, rock collapse, debris flow, and ground collapse. In this section, we compare the sum number of geo-hazards with the results of our UGC inventory. Figures 8a–e show the hierarchical distribution of the number of geo-hazards occurring in prefecture-level cities each year from 2016 to 2020. The “natural-break” and “4-level” classification methods were used in the classification. Red indicates the highest level, orange is the second highest level, light green is the second lowest level, and dark green is the lowest level. Figure 8f shows the counting and grading results of the number of UGC events in different cities.



In 2016, the number of geological disasters in the 11 prefecture-level cities in Zhejiang was graded as seven cities at level 1, two cities at level 2, and one city at levels 3 and 4, respectively. Except for Wenzhou and Lishui, most cities are at a low-risk level. In 2017, the number of geological disasters in 11 prefecture-level cities in Zhejiang was graded as two cities at level 1, three cities at level 2, five cities at level 3, and one city at level 4. Most cities are at a moderate risk level, with Zhoushan being the highest. In 2018, the situation changed again, with most cities at low-risk levels, including three cities with a level 1 risk, five cities with a level 2 risk, and one city with a level 3 risk, with Wenzhou and Hangzhou at the highest risk level. The situation in 2019 was similar to that in 2018, except that the risk level in Zhoushan and Quzhou increased from level 2 to level 3, and in Ningbo, it increased from level 1 to level 2. In 2020, Quzhou and Jinhua were at a level 4 risk, and Hangzhou Lishui and Wenzhou were at a level 3 risk.



**Figure 8.** Geo-hazard frequency quantification of cities in Zhejiang province.

From the above situation over four years, the change in risk of geological disasters in different cities varies greatly between years. In summary, Wenzhou's risk level was at level 4 for 3 years and level 3 for 1 year, making it the city with the most intense geological disasters. This is followed by Hangzhou, which had 2 years at level 4 and 2 years at level 3. The four cities of Huzhou, Jiaxing, Ningbo, and Taizhou were at a low-risk level. Other cities presented some irregularities. In terms of UGC, Hangzhou is the city with the highest frequency of occurrence, followed by Ningbo and Shaoxing at level 3, Quzhou, Jinhua, and Wenzhou at level 2, and the remaining 5 cities at level 1.

It can be seen that the distribution of UGC events in the entire province differs significantly for geo-hazards. First, in terms of coverage, the number of cities at risk of geo-hazards far exceeds the number of cities at risk of UGC. Second, there are differences in high-risk cities. The cities with the highest geo-hazard risk are Wenzhou and Hangzhou, while Wenzhou's UGC risk is relatively low, and Hangzhou occupies a considerable



proportion. This significant difference also tells us that, although UGC is closely related to adverse geological conditions, the factors contributing to its occurrence are fundamentally different from those of geo-hazards, and human activities might be a deeper cause [38]. It can also be noticed that, in the years 2018 and 2019, the high-risk cities for UGC and geo-hazards were the same. The probable reason for this might be the heavy rainfall in summertime [45].

Apart from the difference in distribution pattern, our UGC inventory has another advantage over the Zhejiang Natural Resources and Statistical Yearbook on Environment document. There were only two events recorded in the years 2016 to 2020 in the statistical data, while there are 13 events in our UGC inventory. Further, in the statistical yearbook, UGC events can only be found according to a sum count in each city, whereas we give the specific location, probable trigger factors, and the magnitudes of the events, which are all vital for disaster management.

## 6. Conclusions

In order to accurately identify UGC events from big data in news reports and social media and establish a database for geological disaster prevention and control, this paper discusses the construction of a working framework based on LLM models, which consists of three steps: news report crawling, UGC event recognition, and event attribute extraction. Leveraging the advantages and convenience of GPT in semantic understanding and information extraction, this framework utilized the Baidu search engine and the guidance of UGC event-related keywords, achieving the collection of 637 news reports and the identification and extraction of 27 cases of UGC in Zhejiang province, China, including the latitude, longitude, and time information within a time window of 3 days.

We obtained a UGC event inventory based on the proposed framework, which has been manually checked comprehensively (i.e., for each item). The UGC events in the inventory occurred from May 2005 to May 2022, including 27 cases. This is the first open access inventory of UGC events based on internet news reports, and event dates, locations, and collapse areas derived from unstructured content. The kernel density model shows that the core of UGC risk is concentrated in the eastern part of Hangzhou, while the rest are scattered over different prefecture-level cities and are not spatially concentrated. Compared with geo-hazard statistical data, the inventory has its specialty in reflecting the spatial pattern of UGC frequency in the cities of Zhejiang province which, in turn, proves that UGC is not only caused by adverse geological conditions, but also human activities. This historical database can be used as basic data by relevant departments for the prevention and control of urban geological hazards.

However, the framework and its utilized LLM models still need to be improved. Evaluation of the model showed that (1) the LLM model shows an advantage of being able to identify events, with a recall higher than 60%, yet it is still weak in distinguishing the collapse event type between urban ground collapse and the collapse of artificial buildings or structures. The use of an ontology-type recognition method might solve this problem; (2) the LLM model shows 34% precision in extracting the date of the UGC event and 49% precision in extracting the location of the UGC event. The frequent occurrence of time descriptors in news reports interferes with event extraction, and the time window of 3 days is an optimal choice for co-conference filtering. More work is needed to improve the accuracy of the proposed method in the future.

**Author Contributions:** Conceptualization, Jin Qi, Renyi Liu and Xiaoyi Zhang; Methodology, Jin Qi and Xiaoyi Zhang; Software, Yanan Hao; Validation, Jin Qi; Formal analysis, Yanan Hao; Investigation, Jin Qi and Xiaowen Ma; Resources, Renyi Liu; Writing—original draft, Yanan Hao and Xiaowen Ma; Writing—review & editing, Jin Qi and Xiaoyi Zhang; Visualization, Yanan Hao; Supervision, Xiaoyi Zhang; Project administration, Sensen Wu; Funding acquisition, Sensen Wu. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Key Research and Development Program of China (grant 2021YFB3900900); the Provincial Key R&D Program of Zhejiang (grant 2021C01031); the Deep-time Digital Earth (DDE) Big Science Program; the joint funds of the Zhejiang Provincial Natural Science Foundation of China (grant LHZY24A010001) and the Scientific Research Foundation of Zhejiang University City College (grant X-202208).

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to restrictions imposed by privacy approval and informed consent agreements regarding study participants.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhang, X.; Ruan, Y.; Xuan, W.; Bao, H.; Du, Z. Risk Assessment and Spatial Regulation on Urban Ground Collapse Based on Geo-Detector: A Case Study of Hangzhou Urban Area. *Nat. Hazards* **2023**, *118*, 525–543. <https://doi.org/10.1007/s11069-023-06016-8>.
2. Wang, X.-W.; Xu, Y.-S. Investigation on the Phenomena and Influence Factors of Urban Ground Collapse in China. *Nat. Hazards* **2022**, *113*, 1–33. <https://doi.org/10.1007/s11069-022-05304-z>.
3. Pellicani, R.; Spilotro, G.; Gutiérrez, F. Susceptibility Mapping of Instability Related to Shallow Mining Cavities in a Built-up Environment. *Eng. Geol.* **2017**, *217*, 81–88. <https://doi.org/10.1016/j.enggeo.2016.12.011>.
4. Liu, R.; Sun, H.; Qin, J.; Zheng, Z. A Multi-Geophysical Approach to Assess Potential Sinkholes in an Urban Area. *Eng. Geol.* **2023**, *318*, 107100. <https://doi.org/10.1016/j.enggeo.2023.107100>.
5. Zhang, C.; Kwon, S.; Oh, J.; Park, K. An Integrated AHP-PROMETHEE II Ranking Method to Evaluate the Resilience of Sewer Networks Considering Urban Flood and Ground Collapse Risks. *Water Sci. Technol.* **2023**, *87*, 1438–1453. <https://doi.org/10.2166/wst.2023.067>.
6. Zhou, H.; Liu, C.; Wang, G.; Kang, K.; Liu, Y. Study on Drilling Ground Collapse Induced by Groundwater Flow and Prevention Based on a Coupled CFD-DEM Method. *KSCE J. Civ. Eng.* **2022**, *26*, 2112–2125. <https://doi.org/10.1007/s12205-022-1061-x>.
7. Tufano, R.; Guerriero, L.; Annibali Corona, M.; Bausilio, G.; Di Martire, D.; Nisio, S.; Calcaterra, D. Anthropogenic Sinkholes of the City of Naples, Italy: An Update. *Nat. Hazards* **2022**, *112*, 2577–2608. <https://doi.org/10.1007/s11069-022-05279-x>.
8. Tan, F.; Tan, W.; Yan, F.; Qi, X.; Li, Q.; Hong, Z. Model Test Analysis of Subsurface Cavity and Ground Collapse Due to Broken Pipe Leakage. *Appl. Sci.* **2022**, *12*, 13017. <https://doi.org/10.3390/app122413017>.
9. He, Q.; Tan, F.; Peng, Z.; Tao, L.; Jiao, Y.; Liu, S.; Peng, H. Genesis Analysis of Ground Collapse in Wuhan Based on 3D Geological Model. *Front. Earth Sci.* **2022**, *10*, 934452. <https://doi.org/10.3389/feart.2022.934452>.
10. Zerhouny, M.; Fadil, A.; Hakdaoui, M. Underground Space Utilization in the Urban Land-Use Planning of Casablanca (Morocco). *Land* **2018**, *7*, 143. <https://doi.org/10.3390/land7040143>.
11. Lu, W.; Wu, Y.; Choguill, C.L.; Lai, S.-K.; Luo, J. Underground Hangzhou: The Challenge of Safety vs. Commerciality in a Major Chinese City. *Cities* **2021**, *119*, 103414. <https://doi.org/10.1016/j.cities.2021.103414>.
12. Gutiérrez, F.; Parise, M.; De Waele, J.; Jourde, H. A Review on Natural and Human-Induced Geohazards and Impacts in Karst. *Earth-Sci. Rev.* **2014**, *138*, 61–88. <https://doi.org/10.1016/j.earscirev.2014.08.002>.
13. Xie, H.; Zhang, Y.; Chen, Y.; Peng, Q.; Liao, Z.; Zhu, J. A Case Study of Development and Utilization of Urban Underground Space in Shenzhen and the Guangdong-Hong Kong-Macao Greater Bay Area. *Tunn. Undergr. Space Technol.* **2021**, *107*, 103651. <https://doi.org/10.1016/j.tust.2020.103651>.
14. Zhou, S.; Zhai, G. A Multi-Hazard Risk Assessment Framework for Urban Disaster Prevention Planning: A Case Study of Xiamen, China. *Land* **2023**, *12*, 1884. <https://doi.org/10.3390/land12101884>.
15. Yu, M.; Bambacus, M.; Cervone, G.; Clarke, K.; Duffy, D.; Huang, Q.; Li, J.; Li, W.; Li, Z.; Liu, Q. Spatiotemporal Event Detection: A Review. *Int. J. Digit. Earth* **2020**, *13*, 1339–1365. <https://doi.org/10.1080/17538947.2020.1738569>.
16. Wang, W.; Stewart, K. Spatiotemporal and Semantic Information Extraction from Web News Reports about Natural Hazards. *Comput. Environ. Urban Syst.* **2015**, *50*, 30–40. <https://doi.org/10.1016/j.compenvurbsys.2014.11.001>.
17. Peuquet, D.J.; Robinson, A.C.; Stehle, S.; Hardisty, F.A.; Luo, W. A Method for Discovery and Analysis of Temporal Patterns in Complex Event Data. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1588–1611. <https://doi.org/10.1080/13658816.2015.1042380>.
18. Eligüz, N.; Çetinkaya, C.; Dereli, T. Comparison of Different Machine Learning Techniques on Location Extraction by Utilizing Geo-Tagged Tweets: A Case Study. *Adv. Eng. Inform.* **2020**, *46*, 101151. <https://doi.org/10.1016/j.aei.2020.101151>.
19. Feng, Y.; Sester, M. Extraction of Pluvial Flood Relevant Volunteered Geographic Information (VGI) by Deep Learning from User Generated Texts and Photos. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 39. <https://doi.org/10.3390/ijgi7020039>.
20. Zhang, F.; Fleyeh, H.; Wang, X.; Lu, M. Construction Site Accident Analysis Using Text Mining and Natural Language Processing Techniques. *Autom. Constr.* **2019**, *99*, 238–248. <https://doi.org/10.1016/j.autcon.2018.12.016>.
21. Khalef, R.; El-adaway, I.H. Automated Identification of Substantial Changes in Construction Projects of Airport Improvement Program: Machine Learning and Natural Language Processing Comparative Analysis. *J. Manag. Eng.* **2021**, *37*, 04021062. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000959](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000959).

22. Huang, Y.; Li, Y.; Shan, J. Spatial-Temporal Event Detection from Geo-Tagged Tweets. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 150. <https://doi.org/10.3390/ijgi7040150>.
23. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learn. Individ. Differ.* **2023**, *103*, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>.
24. Kauf, C.; Ivanova, A.A.; Rambelli, G.; Chersoni, E.; She, J.S.; Chowdhury, Z.; Fedorenko, E.; Lenci, A. Event Knowledge in Large Language Models: The Gap between the Impossible and the Unlikely. *Cogn. Sci.* **2023**, *47*, e13386. <https://doi.org/10.48550/arXiv.2212.01488>.
25. Chen, Y.; Tang, R.; Fang, J.; Kong, F. Analysis on Spatio-Temporal Law and Causation of Urban Road Collapse in China from 2014 to 2018. *Water Resour. Hydropower Eng.* **2020**, *51*, 108–116. <https://doi.org/10.13928/j.cnki.wrahe.2020.07.014>.
26. Yang, S.; Feng, D.; Qiao, L.; Kan, Z.; Li, D. Exploring Pre-Trained Language Models for Event Extraction and Generation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5284–5294. <https://doi.org/10.18653/v1/P19-1522>.
27. Liang, Y.; Liu, Y.; Wang, X.; Zhao, Z. Exploring Large Language Models for Human Mobility Prediction under Public Events. *arXiv* **2023**, arXiv:2311.17351. <https://doi.org/10.48550/arXiv.2311.17351>.
28. Huang, F.; Huang, Q.; Zhao, Y.; Qi, Z.; Wang, B.; Huang, Y.; Li, S. A Three-Stage Framework for Event-Event Relation Extraction with Large Language Model. In Proceedings of the International Conference on Neural Information Processing, Changsha, China, 20–23 November 2023; Springer: Singapore, 2023; pp. 434–446. [https://doi.org/10.1007/978-981-99-8181-6\\_33](https://doi.org/10.1007/978-981-99-8181-6_33).
29. Abraham, S.; Mäs, S.; Bernard, L. Extraction of Spatio-Temporal Data about Historical Events from Text Documents. *Trans. GIS* **2018**, *22*, 677–696. <https://doi.org/10.1111/tgis.12448>.
30. Kokla, M.; Guilbert, E. A Review of Geospatial Semantic Information Modeling and Elicitation Approaches. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 146. <https://doi.org/10.3390/ijgi9030146>.
31. Hu, X.; Zhou, Z.; Li, H.; Hu, Y.; Gu, F.; Kersten, J.; Fan, H.; Klan, F. Location Reference Recognition from Texts: A Survey and Comparison. *ACM Comput. Surv.* **2023**, *56*, 1–37. <https://doi.org/10.1145/3625819>.
32. Sermet, Y.; Demir, I. Towards an Information Centric Flood Ontology for Information Management and Communication. *Earth Sci. Inform.* **2019**, *12*, 541–551. <https://doi.org/10.1007/s12145-019-00398-9>.
33. Wang, B.; Wu, L.; Li, W.; Qiu, Q.; Xie, Z.; Liu, H.; Zhou, Y. A Semi-Automatic Approach for Generating Geological Profiles by Integrating Multi-Source Data. *Ore Geol. Rev.* **2021**, *134*, 104190. <https://doi.org/10.1016/j.oregeorev.2021.104190>.
34. Guo, L.; Lei, C.; Xiong, C. Application Research of Ground Penetrating Radar Technology in Prevention and Control of Ground Collapse Disaster in Shenzhen. In Proceedings of the SEG 2019 Workshop: Geophysics for Smart City Development, Beijing, China, 29–31 July 2019; Society of Exploration Geophysicists: Houston, TX, USA, 2019; pp. 19–19. <https://doi.org/10.1190/smct-2019-16.1>.
35. Liao, X.; Xu, W.; Zhang, J.; Qiao, Y.; Meng, C. Analysis of Affected Population Vulnerability to Rainstorms and Its Induced Floods at County Level: A Case Study of Zhejiang Province, China. *Int. J. Disaster Risk Reduct.* **2022**, *75*, 102976. <https://doi.org/10.1016/j.ijdrr.2022.102976>.
36. Liu, B.; Siu, Y.L.; Mitchell, G. A Quantitative Model for Estimating Risk from Multiple Interacting Natural Hazards: An Application to Northeast Zhejiang, China. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 1319–1340. <https://doi.org/10.1007/s00477-016-1250-6>.
37. Ding, J.; Zhang, G.; Gao, J.; Wang, S.; Xue, B.; Wang, K.; Yang, J.; Jiang, R.; Lian, J. The Rainfall Thresholds and Soil Characteristics of Large Geological Disasters in Zhejiang, China. *Arab. J. Geosci.* **2022**, *15*, 559. <https://doi.org/10.1007/s12517-021-09303-5>.
38. Zhou, L.; Zhang, X.; Shen, H.; Zhang, D.; Bao, H. GIS Analysis of Urban Ground Collapse Susceptibility: A Case Study of Eastern Hangzhou, China. *Front. Earth Sci.* **2024**, *11*, 1338877. <https://doi.org/10.3389/feart.2023.1338877>.
39. Yan, F.; Qiu, W.; Sun, K.; Jiang, S.; Huang, H.; Hong, Y.; Hou, Z. Investigation of a Large Ground Collapse, Water Inrush and Mud Outburst, and Countermeasures during Subway Excavation in Qingdao: A Case Study. *Tunn. Undergr. Space Technol.* **2021**, *117*, 104127. <https://doi.org/10.1016/j.tust.2021.104127>.
40. Gong, X.; Zhang, X. Excavation Collapse of Hangzhou Subway Station in Soft Clay and Numerical Investigation Based on Orthogonal Experiment Method. *J. Zhejiang Univ. Sci. A* **2012**, *13*, 760–767. <https://doi.org/10.1631/jzus.A1200120>.
41. Chen, R.P.; Li, Z.C.; Chen, Y.M.; Ou, C.Y.; Hu, Q.; Rao, M. Failure Investigation at a Collapsed Deep Excavation in Very Sensitive Organic Soft Clay. *J. Perform. Constr. Facil.* **2015**, *29*, 04014078. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0000557](https://doi.org/10.1061/(ASCE)CF.1943-5509.0000557).
42. Zeng, A.; Liu, X.; Du, Z.; Wang, Z.; Lai, H.; Ding, M.; Yang, Z.; Xu, Y.; Zheng, W.; Xia, X.; et al. Glm-130b: An Open Bilingual Pre-Trained Model. *arXiv* **2022**, arXiv:2210.02414. <https://doi.org/10.48550/arXiv.2210.02414>.
43. Chen, W.; Zhou, Y.; Stokes, E.C.; Zhang, X. Large-Scale Urban Building Function Mapping by Integrating Multi-Source Web-Based Geospatial Data. *Geo-Spat. Inf. Sci.* **2023**, 1–15. <https://doi.org/10.1080/10095020.2023.2264342>.

44. Qin, C.; Tang, Z.; Chen, J.; Chen, X. The Impact of Soil and Water Resource Conservation on Agricultural Production-an Analysis of the Agricultural Production Performance in Zhejiang, China. *Agric. Water Manag.* **2020**, *240*, 106268. <https://doi.org/10.1016/j.agwat.2020.106268>.
45. Du, M.; Hou, Y.; Qi, P.; Wang, K. The Impact of Different Historical Typhoon Tracks on Storm Surge: A Case Study of Zhejiang, China. *J. Mar. Syst.* **2020**, *206*, 103318. <https://doi.org/10.1016/j.jmarsys.2020.103318>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.