GALLOC: A GeoAnnotator for Labeling LOCation descriptions from disaster-related text messages

Kai Sun[1], Yingjie Hu[1,2], Kenneth Joseph[2], and Ryan Zhenqi Zhou[1]

[1] GeoAI Lab, Department of Geography, University at Buffalo, Buffalo, NY, USA
[2] Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA

Abstract: During a natural disaster, people post text messages on various platforms, such as social media and short message service (SMS) platforms, to share urgent information and seek help. Many text messages contain location descriptions about victims and accidents. Accurately extracting these location descriptions can help disaster responders reach victims more quickly and even save lives. These location descriptions, however, are often more complex than simple place names (e.g., city names), and cannot be extracted using typical named entity recognition approaches. While new machine learning models could be trained, they require labeled training data that are time-consuming to create without an effective data annotation tool. To fill this gap, we develop GALLOC, a GeoAnnotator for Labeling LOCation descriptions from disaster-related text messages. GALLOC is an open-source and Web-based tool that provides a variety of functions for supporting location description annotation, such as artificial intelligence powered pre-annotation and automatic spatial footprint identification. It also supports multilingual data annotation, and can be used by a group of users to collaboratively create a dataset. We present the design considerations and functions of GALLOC, and evaluate it via a comparison with previous tools and an experiment to annotate a small set of disaster-related messages.

Keywords: Geo-annotation; location description; machine learning; disaster response; GeoAI.

1. Introduction

During a natural disaster, people post text messages on various platforms, such as social media and short message service (SMS) platforms, to share urgent information and seek help (Huang & Xiao, 2015; Yu et al., 2019; Suwaileh et al., 2022; Zhou et al., 2022; Zou et al., 2022). Many of these text messages contain important location descriptions about victims and accidents. Here is one example message posted on Twitter during Hurricane Harvey (the message content is slightly modified for privacy protection): "*2799 7th Avenue, Port Arthur, Texas. Please help this family out! They're stuck in their home with children! #HurricaneHarvey*". Another message said "*Can someone help? My friend has been stuck at Lyons Ave & Gregg St for hours. #Harvey*." In addition to the two examples, other forms of location descriptions have also been seen, such as: (i) highway exit, as in the message "*Flooding at I-45 Exit 47A. The exit is closed. Take a detour if you are going in that direction #houstonflood*"; (ii) road segment, as in "*Streets Flooded: Almeda Genoa Rd. from Windmill Lakes Blvd. to Rowlett Rd. #HurricaneHarvey #Houston*"; and (iii) adjacent

This is a preprint. The official version is published in the International Journal of Geographical Information Science at: https://doi.org/10.1080/13658816.2025.2464643

neighborhoods, as in "*Anyone doing high water rescues in the Royalwoods/Wilson Park area? My daughter has been stranded in a parking lot all night.*"

While people are generally advised to call 911 rather than posting on social media under emergency situations, previous disasters have shown that social media has become an important information channel when 911 is unreachable due to the disaster. A National Public Radio (NPR) report, titled "Facebook, Twitter Replace 911 Calls For Stranded In Houston" (Silverman, 2017), described how social media platforms were used by many Houston residents during Hurricane Harvey to request for help when 911 could not be reached. The facts that the storm damaged multiple emergency call centers and that there were a large number of 911 calls at the same time were two major reasons for the failure of the 911 system. Similar stories were also reported by The Wall Street Journal (Seetharaman & Wells, 2017) and Time Magazine (Rhodan, 2017). In addition to Hurricane Harvey, people also used social media to seek help in other more recent disasters, such as Hurricane Ian (Karimiziarani & Moradkhani, 2023) and the Buffalo blizzard in 2022 (Tsujimoto, 2023). While we can always strengthen 911 and other emergency response infrastructures, there will be situations when these infrastructures are paralyzed by the disaster or are overloaded by help requests. In these situations, social media becomes an important channel for people to seek help. Accurately extracting location descriptions from disaster-related messages on social media and other platforms can help disaster responders reach victims more quickly and even save lives.

Existing research on location extraction from texts often uses named entity recognition (NER) approaches (Gelernter & Mushegian, 2011; Gelernter & Balaji, 2013; Karimzadeh et al., 2019; Wang et al., 2020; X. Hu et al., 2022; Suwaileh et al., 2022; Berragan et al., 2023). Such approaches consider location extraction as a sub-problem of general NER, and use off-the-shelf NER tools or train deep learning based NER models to extract location entities from texts. While these approaches can effectively extract certain location entities, such as city names and state names, many location descriptions used by people during natural disasters consist of multiple entities, such as door number addresses, road intersections, and highway exits (Y. Hu & Wang, 2021; Fernández-Martínez, 2022; Chen et al., 2022; Suwaileh et al., 2023). These more complex location descriptions cannot be effectively extracted using typical NER approaches which are designed to extract individual entities. Figure 1 illustrates this issue using the two example messages discussed previously. Typical NER approaches extract individual and separate entities, such as "7th Avenue", "Port Arthur", and "Texas", while we need a complete location description, such as "2799 7th Avenue, Port Arthur, Texas", to identify the precise location of the victims. This issue can have serious consequences for disaster response, e.g., by erroneously geo-locating the victims to the center of "Port Arthur" or even to the center of "Texas".

(a) Separate entities typically output from NER approaches

(b) Complete location description needed for geo-locating victims

Figure 1. A comparison of the separate entities output by typical NER approaches and the complete location descriptions needed to geo-locate victims.

Theoretically, we can overcome the above issue by training new machine learning models using new datasets labeled with complete location descriptions rather than individual entities. However, there is a lack of datasets labeled in this manner, and it is time-consuming and labor-intensive to create new datasets without an effective data annotation tool. There exist data annotation tools for general natural language processing (NLP) tasks, such as BRAT (Stenetorp et al., 2012) and GATE Teamware (Bontcheva et al., 2013), but they lack a map interface that is highly important for assigning geographic coordinates to location descriptions. There also exist tools specifically developed for annotating toponyms from texts, such as WOTR GeoAnnotator (DeLozier et al., 2016), PSU GeoAnnotator (Karimzadeh & MacEachren, 2019), and GeoViz (McDonough et al., 2019). However, these tools have limitations in supporting the annotation of location descriptions not contained in typical gazetteers (e.g., door number addresses and road intersections) and do not provide effective support for the annotation of location categories.

In this paper, we present GALLOC: a *G*eo*A*nnotator for *L*abeling *LOC*ation descriptions from disaster-related text messages. GALLOC is designed as an open-source and Web-based geo-annotation tool that can facilitate the creation of datasets with labeled location descriptions and location categories. GALLOC can be used by machine learning practitioners to create labeled datasets for training new machine learning models. It can also be used by disaster researchers and emergency managers to analyze disaster-related location descriptions used by people in a geographic region and support future disaster response efforts. The contributions of this paper are as follows:

- We develop GALLOC as a geo-annotation tool for supporting the creation of labeled datasets with disaster-related location descriptions. We also present its design considerations which may help inform the development of similar geo-annotation tools.

- We design a variety of functions for GALLOC to facilitate the geo-annotation process. These functions include artificial intelligence (AI) powered pre-annotation, automatic spatial footprint identification, and multilingual data annotation support.

- We share related resources of GALLOC, including an online demo, open source code, a Docker image, and a user manual. These resources can help interested readers deploy and use GALLOC more easily, and also increase the reproducibility of this research.

The remainder of this paper is organized as follows. Section 2 reviews related work on data annotation tools for general NLP research as well as geo-annotation tools for annotating place names from texts. Section 3 presents the design considerations of GALLOC, and then Section 4 presents its overall architecture and implementation. Section 5 presents the major functions of GALLOC that we develop to meet the design considerations. In Section 6, we evaluate GALLOC via a comparison with previous tools and an experiment to annotate a small set of disaster-related messages. Section 7 discusses the implications of GALLOC and data annotation in the context of AI and disaster response. Finally, Section 8 concludes this research.

2. Related work

Data annotation tools are often developed in machine learning research to facilitate the creation of labeled datasets. In the field of natural language processing, a number of data annotation tools have been developed (Neves &Ševa, 2021). Examples include BRAT (Stenetorp et al., 2012), GATE Teamware (Bontcheva et al., 2013), and WebAnno (Yimam et al., 2013). These tools can be used to create annotated datasets for supporting a wide range of NLP tasks, from named entity recognition and dependency parsing to coreference resolution and sentiment analysis. While highly useful, these annotation tools were developed for general NLP tasks and typically do not provide a map-based user interface that is important for linking textual descriptions to their geographic locations. Consequently, it is difficult to directly use these general NLP-oriented data annotation tools for labeling location descriptions from disaster-related text messages.

There also exist geo-annotation tools specifically developed for creating datasets to support toponym-related tasks. An early example is the Toponym Annotation Markup Editor (TAME) developed by Leidner (2006). TAME was developed as a Web-based tool for the task of toponym resolution, i.e., given an ambiguous toponym, such as “Paris”, identifying its referred place instance, such as “Paris, Texas” (Jones & Purves, 2008; Purves et al., 2018; Ju et al., 2016; Gritta et al., 2018). TAME assumes that toponyms have already been pre-identified in text and the job of human annotators is to assign a pre-identified toponym to its place entry in two gazetteers developed by U.S. agencies, i.e., the Geographic Names Information System (GNIS) gazetteer by the U.S. Geological Survey and the GEOnet Names Server (GNS) gazetteer by the U.S. National Geospatial Intelligence Agency. Surprisingly, TAME does not provide a map interface, probably due to the limited Web mapping technologies back in 2006. The Edinburgh GeoAnnotator (Alex et al., 2014) is another annotation tool also developed for the task of toponym resolution. While it does provide a map-based interface, it makes the same assumption as TAME, i.e., toponyms are already pre-identified from the text, and the tool is only used to assign the pre-identified toponyms to gazetteer entries in the GeoNames gazetteer. The DLRGeoAnnotator is a recent tool developed also for assigning pre-identified toponyms to place entries in OpenStreetMap (X. Hu et al., 2024). It is a command-line based tool without a graphic user interface, and does not support the annotation of toponyms from texts. DLRGeoAnnotator was developed as a preliminary tool for creating the DLRGeoTweet dataset (X. Hu et al., 2024), and does not seem to be intended for general use currently.

Three geo-annotation tools have been developed for supporting both the annotation of toponyms from texts and linking the annotated toponyms to their geographic representations or gazetteer entries. WOTR GeoAnnotator is a geo-annotation tool developed by DeLozier et al. (2016) for annotating toponyms in a historical corpus, called the War Of The Rebellion (WOTR), which contains a set of historical archives about the American Civil War. The WOTR GeoAnnotator does not provide much automation: users need to manually annotate toponyms and manually draw their geographic representations; users are also expected to do many external searches on Google Maps and other websites in order to find the locations of toponyms (DeLozier et al., 2016). GeoViz is a geo-annotation tool that was developed for annotating a historical French corpora (McDonough et al., 2019). It uses the Edinburgh Geoparser (Alex et al., 2015) to pre-annotate toponyms, and allows human annotators to further edit the pre-annotated toponyms and assign them to entries in the GeoNames gazetteer. The PSU GeoAnnotator was developed by Karimzadeh and MacEachren (2019) at the Pennsylvania State University (PSU) to annotate toponyms from a corpus of tweets (Wallgrün et al., 2018). The tool embeds several off-the-shelf NER models, such as GATE ANNIE (Cunningham, 2002), CogComp NER (Ratinov & Roth, 2009), and Stanford NER (Manning et al., 2014), to pre-annotate toponyms, and allows human annotators to further edit these toponyms and assign them to entries in GeoNames.

While existing geo-annotation tools facilitate the creation of labeled datasets, they have two major limitations. First, these tools were primarily designed for annotating toponyms contained in gazetteers (e.g., city names and country names), and do not support the annotation of location descriptions that are not contained in a typical gazetteer. Descriptions like "2799 7th Avenue, Port Arthur, Texas" and "Lyons Ave & Gregg St" are not contained in a typical gazetteer, and cannot be easily annotated using these tools. While the WOTR GeoAnnotator does allow users to manually select a location description and draw its geographic representation without a gazetteer, its lack of automation largely hinders the efficiency of data annotation. Second, existing tools do not provide effective support for annotating the categories of location descriptions. Having the categories of location descriptions, such as *door number addresses*, *road intersections*, and *administrative units*, is important for geo-locating them, since different categories of location descriptions often require different geo-locating techniques. For example, we may want to use linear geocoding for *door number addresses*, line-based intersection identification for *road intersections*, and place name matching for *administrative units*. However, existing geo-annotation tools either consider all annotated locations as one single category, or use pre-defined categories from a gazetteer without supporting customized location categories based on a user-defined scheme. GALLOC is developed to overcome these two major limitations. In addition, GALLOC provides a variety of functions to further increase data annotation efficiency.

3. Design considerations

Before developing GALLOC, we formalize a set of design considerations based on literature review, the features of existing data annotation tools, and the needs of labeling location descriptions from disaster-related text messages. In the following, we present these design considerations (DCs).

DC 1: *Annotating complete location descriptions, not separate entities*. This consideration is based on the fundamental need of labeling disaster-related location descriptions, such as “2799 7th Avenue, Port Arthur, Texas” and “Lyons Ave & Gregg St”, into one complete piece rather than multiple separate entities. Annotations of complete location descriptions are critical for accurately geo-locating them. Given this consideration, we also cannot require a location description to be linked to a gazetteer entry, since many descriptions, such as door number addresses, road intersections, and highway exits, are not contained in typical gazetteers.

DC 2: Annotating the categories of location descriptions based on a user-defined scheme. Different categories of location descriptions may need different geo-locating techniques. Knowing the category of a location description is critical for choosing the most suitable geo-locating technique. Depending on the ultimate purposes of the created datasets, users may have their own preferences on the specific location categories and the number of categories. With this consideration, the annotation tool should allow users to define their own location category scheme and annotate data based on the defined scheme.

DC 3: *Providing pre-annotation of location descriptions from text*. Research resulting in the development of the PSU GeoAnnotator (Karimzadeh & MacEachren, 2019) and GeoViz (McDonough et al., 2019) suggested that having location descriptions pre-annotated from text helps increase the efficiency of data annotation. We adopt the same view here while also taking into account the fast advancement of AI technologies. Since new AI models are being developed constantly, we expect GALLOC to be able to connect to newly-developed AI models for pre-annotation, rather than using a fixed pre-annotation model that can become outdated quickly.

DC 4: Providing automatic identification of spatial footprints for location descriptions. The spatial footprints of location descriptions can be in different geometry types, such as points (e.g., door number addresses), lines (e.g., roads), and polygons (e.g., city districts). When such spatial footprints are available from external sources, e.g., OpenStreetMap and Google Maps, this data annotation tool should provide these spatial footprints automatically to reduce potential external searches and increase data annotation efficiency. This consideration follows a lesson provided in the research of WOTR GeoAnnotator (DeLozier et al., 2016).

DC 5: Supporting customized drawing of spatial footprints. Also learning from the WOTR GeoAnnotator (DeLozier et al., 2016), we aim to provide users with the ability to manually draw spatial footprints for some location descriptions when it is necessary. While automatic spatial footprint identification can help users find and create spatial footprints in many cases, there will be situations when a location description cannot be directly geo-located by external sources or when the automatically geo-located spatial footprint is unsatisfactory. In these situations, users of the geo-annotation tool should be able to manually create or edit spatial footprints.

DC 6: Supporting multi-user collaborative data annotation. The creation of labeled datasets, especially large labeled datasets, often requires collaboration among multiple human annotators who work together to annotate data and resolve annotation disagreements. The need for collaborative annotation was also suggested in the research of PSU GeoAnnotator (Karimzadeh & MacEachren, 2019) and general annotation tools, such as GATE Teamware (Bontcheva et al., 2013). With this consideration, the data annotation tool should support collaborative annotation by multiple users who may have different roles (e.g., the roles of *administrator* and *annotator*) for creating the dataset.

DC 7: Supporting the annotation of multilingual datasets. Depending on the geographic region affected by a disaster, location descriptions and text messages can be in languages beyond English, such as Spanish messages posted during Hurricane Maria in 2017 and Chinese messages posted during the Beijing flooding in 2023. In addition, location descriptions in different languages tend to have different linguistic patterns, and the ability to create labeled datasets for different languages will enable future research to study location descriptions across languages. With this consideration, the data annotation tool should be able to support the annotation of multilingual datasets.

In addition to the seven design considerations above, we also expect the geo-annotation tool to be Web-based and open-source. A Web-based platform facilitates the deployment and use of the developed tool, and much previous research, such as the PSU GeoAnnotator (Karimzadeh & MacEachren, 2019), WOTR GeoAnnotator (DeLozier et al., 2016), GeoViz (McDonough et al., 2019), and many general data annotation tools (Stenetorp et al., 2012; Yimam et al., 2013), were all designed as Web-based tools. In addition, the data annotation tool should be open-source to allow future extensions and modifications by interested researchers and users.

4. Overall architecture and implementation

Based on the design considerations, we design the overall architecture of GALLOC as shown in Figure 2. It has three major modules: *User Module*, *Project Module*, and *Annotation and Resolution Module.*

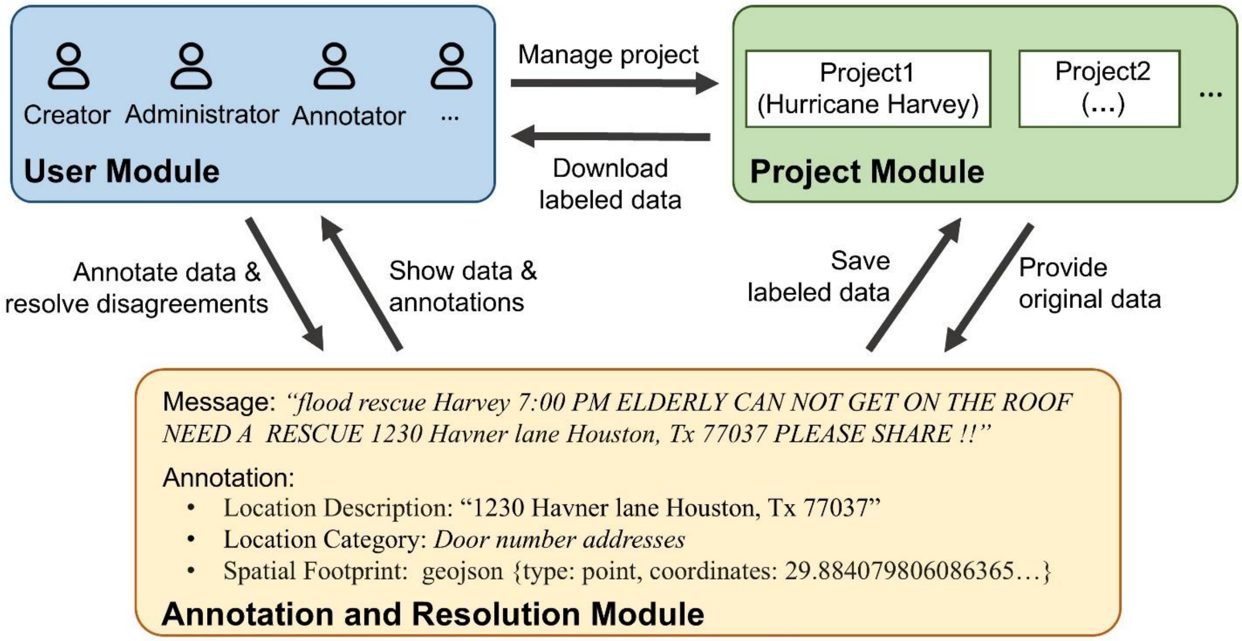

Figure 2. The overall architecture of GALLOC.

User Module. The User Module provides functions for supporting multiple users involved in the data annotation process. It provides three user roles with different privileges, which are: *creator*, *administrator*, and *annotator*. The *creator* is a super user who creates a data annotation project and has all the privileges provided by GALLOC, including adding other *administrators* and *annotators*, changing project settings, and all the privileges of an *administrator*. An *administrator* is a senior user who has most privileges, including changing project settings, resolving potential annotation disagreements, and all the privileges of an *annotator*. An *annotator* is a regular user who can make annotations and view projects, but does not have other privileges to change project settings. GALLOC also automatically adjusts its user interface by showing only those functions relevant to the role of a user. The User Module can help a group of users coordinate their data annotation work and facilitate the creation of a labeled dataset, especially when the size of the dataset is large.

Project Module. The Project Module provides functions for managing data annotation projects. Each project is linked to the context about the geographic region affected by the disaster (provided by the project *creator*), and allows human annotators to focus on the target geographic region during the annotation process. This geographic context reduces the potential place name ambiguity issue for data annotation, since the same place name is less likely to be used repeatedly in a focused geographic region. The Project Module also allows *creators* and *administrators* to set critical project parameters, such as the number of human annotators needed for annotating each message as well as the location categories defined for the project. The Project Module provides other fundamental functions as well, such as uploading new datasets for annotation and downloading annotated datasets.

Annotation and Resolution Module. This is the main module of GALLOC for data annotation. It supports two major tasks: annotating individual text messages with location descriptions and resolving potential disagreements from different human annotators. We develop a variety of functions to facilitate the data annotation process, such as AI-powered pre-annotation of location descriptions from text and automatic spatial footprint identification. After a text message is annotated by two or more human annotators, GALLOC will automatically compare the annotations and detect disagreements; these disagreements will then be displayed in a juxtaposed manner to facilitate their comparison and resolution. Only senior users in the roles of *creator* and *administrator* can resolve disagreements, but experienced *annotators* could be "promoted" to *administrators* at a later stage of the project to help with disagreement resolution.

GALLOC is implemented as a Web-based and open-source platform. It uses a technology stack of open-source libraries and Web development frameworks, such as HTML5, JavaScript, CSS3, and multiple Web libraries and frameworks such as Bootstrap and jQuery. A self-contained, serverless, and public domain database, SQLite 3, is employed to further simplify the deployment of GALLOC. To create an annotated dataset, a senior user in the role of *creator* or *administrator* first uploads a dataset containing a list of text messages to be annotated. Each message is formatted as a JavaScript Object Notation (JSON) object in its own line, and the content of the message is put in the "text" attribute. Figure 3 shows two input messages organized as two JSON objects in one file. Additional attributes about the message, such as when this message was posted and how many times it was reposted by others, could be included into the JSON object as well and will be included in the annotated result. The list of messages is prepared into a simple text file, such as a *.txt or *.json file, and is uploaded to GALLOC for annotation.


```
{"text": "2799 7th Avenue, Port Arthur, Texas. Please help this family out! They're stuck in their home with children! #HurricaneHarvey“}
{"text": "Quick update: pregnant woman in labor on San Angelo Street has been finally transported by ambulance. #Harvey "}
…
```


Figure 3. The format of the input text messages to be uploaded to GALLOC.

The output of GALLOC is the annotated messages which are automatically organized into a *.json file. Figure 4 shows two annotated messages. Similar to the input data file, the output file contains a list of JSON objects with each object representing one annotated message in its own line. A JSON object contains the original text message, the annotated location descriptions, the annotator information, and other attributes of the message included in the input data file. Each annotated location description contains its start and end positions in the message (via the "startIdx" and "endIdx" attributes), the location description text (via the "locationDesc" attribute), the location category (via the "locationCate" attribute), and its spatial footprint (via the "spatialFootprint" attribute). In addition, GALLOC offers two options to download the annotated dataset: one can choose to download all data annotations (i.e., the raw version of the labeled data) or only the agreed and resolved annotations (which can be considered as a cleaned version of the data).


```
1   {"Annotator":"John","AnnotationID":"qhxmyglfolyukhllgumurucvetpfgzcy","text":"
    2799 7th Avenue, Port Arthur, Texas. Please help this family out! They're stuck
    in their home with children! #HurricaneHarvey","Method":"Annotate",
    "Annotation":[{"endIdx":35,"startIdx":0,"spatialFootprint":[{"geometry":[{"coordin
    ates":[-93.932893,29.914454],"type":"Point"}],"type":"Point"}], "locationDesc":
    "2799 7th Avenue, Port Arthur, Texas","locationCate":"C1: Door number
    addresses"}],"MessageID":"fbddiuycdivaffmpshxepcpbijlhvcah"}
2   {"Annotator":"John","AnnotationID":"sxdbjotrdbqmtxsoktlrgocwvhzvbeeu","text":
    "Quick update: pregnant woman in labor on San Angelo Street has been finally
    transported by ambulance. #Harvey","Method":"Annotate","Annotation":
    [{"endIdx":58,"startIdx":41,"spatialFootprint":[{"geometry":[{"coordinates":[[-
    95.301243,29.770777],[-95.297735,29.770814],[-95.297745,29.770451],[-
    95.292982,29.770553],[-95.292681,29.770516]],"type":"LineString"}],
    "type":"LineString"}],"locationDesc":"San Angelo Street","locationCate":"C2:
    Street names"}],"MessageID":"agifqbpsazwupgmsolkiyztletatrfma"}
... ...
```


Figure 4. The format of the annotated text messages output by GALLOC.

There are other implementation and usage details about GALLOC, such as updating the roles of different users and checking the annotation progress of a project. Due to their technical nature, we put these details into the User Manual rather than this current paper. Interested readers can check this User Manual, our online demo, and source code repository of GALLOC at the following links:

- User Manual: https://github.com/geoai-lab/GALLOC/blob/master/User_Manual_GALLOC.pdf
- Online demo: https://geoai.geog.buffalo.edu/GALLOC/
- Source code repository: https://github.com/geoai-lab/GALLOC

5. Major functions of GALLOC

In this section, we present the major functions of GALLOC that are developed to meet the seven design considerations.

5.1. Annotating complete location descriptions and categories (Meeting DC 1 and DC 2)

GALLOC supports the annotation of location descriptions as complete pieces rather than separate entities. Figure 5 provides a screenshot of GALLOC demonstrating this function. To annotate a location description, the user selects the corresponding text span from the message, e.g., "*2799 7th Avenue, Port Arthur, Texas*" as highlighted in yellow in the screenshot. In cases when a text span is selected mistakenly, it can be canceled by clicking the cross icon at the upper right corner of the selected span.

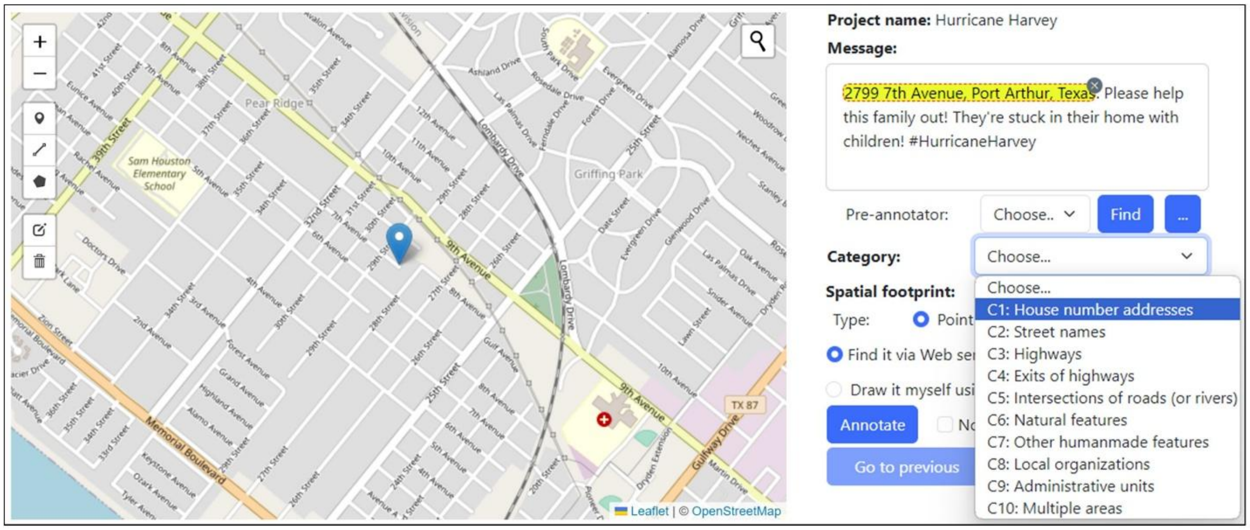

Figure 5. A screenshot of annotating complete location descriptions and categories.

After selecting the text span of the location description, the user can move on to specifying its category from a drop-down list with pre-populated location categories defined by the project *creator* or *administrators*. If detailed location categories are not necessary for a project, it can be left empty in the project setting and all location descriptions will be put under a default category of “Location”. Figure 5 shows an example of selecting a location category from a list.

After specifying the location category, the user can move on to annotating the spatial footprint using automatic footprint identification or manual drawing functions (more details are provided in Section 5.3). With these three steps, a location description is annotated with its complete textual description, location category, and spatial footprint. The user can repeat this process if multiple location descriptions exist in one text message, or can move on to the next message if all location descriptions have been annotated.

5.2. AI-powered pre-annotation of location descriptions from text (Meeting DC 3)

While users can manually select the text span of a location description, a pre-annotation tool can speed up this process. GALLOC provides a connection interface that allows users to connect their preferred pre-annotation tool to the system. Currently, we have added the NeuroTPR model to GALLOC for pre-annotation, which is a deep neural network based NER model (Wang et al., 2020). Other AI-powered NER tools, such as GazPNE (X. Hu et al., 2022) and Topobert (Zhou et al., 2023), can also be configured and connected to GALLOC. To connect a pre-annotation tool, the user provides the name and the URL of the tool as a Web service, as shown in Figure 6. GALLOC then automatically connects to this pre-annotation tool based on the information provided. We note that previous research has already used AI-powered models for pre-annotation (Karimzadeh & MacEachren, 2019; McDonough et al., 2019). The novelty of GALLOC is in the designed connection interface, rather than using embedded and fixed pre-annotation tools, which allows the pre-annotation tools to be updated as technology advances in the coming years. This connection interface allows more powerful AI tools likely to be developed in the future to be connected to GALLOC, and also enables different users to choose pre-annotation tools based on their preferences and project needs.

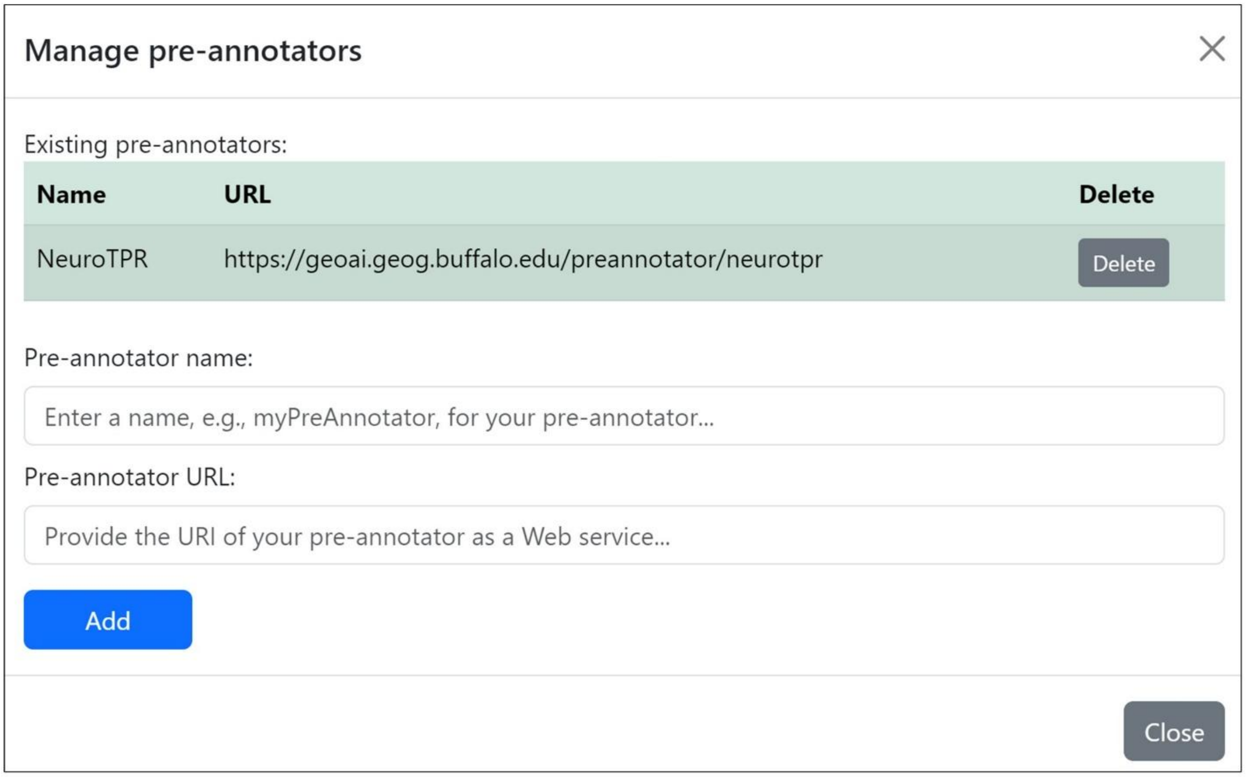

Figure 6. Interface for connecting a new pre-annotation tool to GALLOC.

While the connection interface supports a wide range of pre-annotation tools, their outputs need to be organized in a format that can be interpreted by GALLOC. Figure 7 shows this format: the output is formatted as one JSON object with an attribute "*Annotation"* that contains a list of the location descriptions identified by the tool. Each pre-annotated location description is a JSON object that has at least three attributes: (1) "*startIdx*": the starting position of a location description in the message; (2) "*endIdx*": the ending position; (3) "*locationDesc*": the text of the location description.

```
{
        Annotation:
        [
                {
                        startIdx: 32,
                        endIdx: 45,
                        locationDesc: "215 Logan St."
                },
                {
                        … # another recognized location description
                },
                …
        ]
}
```

Figure 7. The format for the output of a pre-annotator to be connected to GALLOC.

5.3. Supporting automatic and manual creation of spatial footprints (Meeting DC 4 and DC 5)

GALLOC supports both automatic and manual creation of spatial footprints for location descriptions. For automatic spatial footprint identification, we integrate the Web services from Nominatim (which is based on OpenStreetMap) and Google Maps Geocoding API. These two Web services cover a rich number of geographic features throughout the world, and are likely to continue their service in the coming years. Nominatim is open-source and can be used free of charge (a local installation of Nominatim is recommended). Google Maps Geocoding API provides a credit of $200 per month which supports about 40,000 free queries and is sufficient for many general use cases. In terms of their abilities to create spatial footprints, Google Maps Geocoding API only returns point-based geometries, while Nominatim can return more detailed geometries in polylines and polygons; however, Google Maps Geocoding API does a very good job in locating door number addresses. Thus, these two Web services can be used in a complementary manner. Figure 8 shows an example of using Nominatim to identify the spatial footprint of “Brays Bayou”.

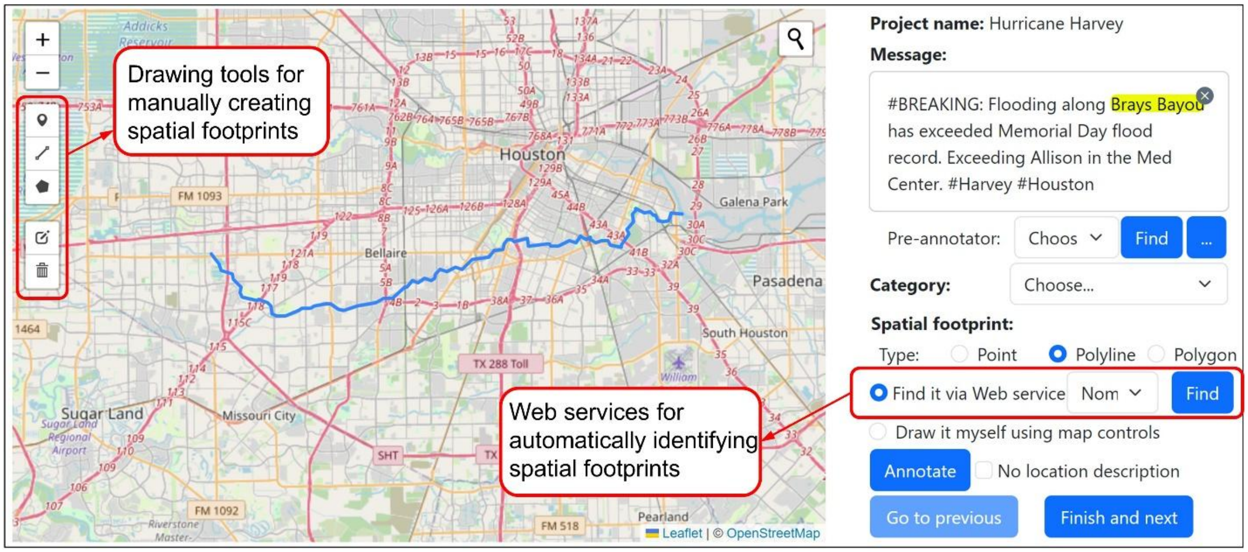

Figure 8. Functions for supporting automatic and manual creation of spatial footprints.

While automatic spatial footprint identification is highly useful, GALLOC also provides manual drawing functions for creating and editing spatial footprints. These functions are shown in the upper left corner of the map interface in Figure 8. Manual drawing could introduce errors in the created spatial footprints; however, such functions are still necessary when a location description cannot be directly geo-located by Nominatim and Google Maps, or when an automatically returned spatial footprint is not satisfactory. In these situations, the manual drawing functions can be used to create a spatial footprint from scratch, or to edit the initial footprint returned by a map service. To reduce the potential errors of manual drawing, it can be helpful to have some user training on these drawing functions before a formal data annotation task begins.

GALLOC supports the use of different geometry types in the same dataset. Thus, users can choose to use *points* to represent door number addresses and road intersections, *polylines* to represent roads and rivers, and *polygons* to represent parks and neighborhoods. Depending on the project needs, users also have the flexibility to simply use one type of geometry, e.g., using points only. The created spatial footprints are saved using the standard GeoJSON format in the annotated dataset.

5.4. Multi-user collaborative data annotation (Meeting DC 6)

To facilitate the creation of a large labeled dataset, GALLOC supports collaborative data annotation by a team of users. With different user roles provided by GALLOC, a collaborative data annotation can be completed in the following three main steps: (1) project setup by the project *creator* and *administrators*; (2) data annotation by all users, i.e., *annotators*, *administrators*, and *creator*; and (3) disagreement resolution by the project *creator* and *administrators*. As discussed before, the role of a user is editable and experienced *annotators* could also be promoted to *administrators* to help with disagreement resolution.

To resolve annotation disagreements, GALLOC automatically compares the annotations of different users on the same messages. The comparison focuses on the labeled text span of the location description, location category, the type of geometry, and location coordinates (with an acceptable distance range, currently set as within 1000 meters). Annotations that are determined as different by GALLOC are shown to users for disagreement resolution. Figure 9 provides a screenshot of two different annotations for the same text message. While both annotations select the same text span and the same location category, the first annotation has a polyline spatial footprint and the second annotation uses a point. With these two annotations, the user can then decide whether to directly accept one of these annotations, or to revise based on one annotation if neither annotation is satisfactory. The number of annotations needed for one text message is determined by a project administrator. For example, the administrator may decide that each message needs to be annotated by 3 different human annotators for quality assurance. If the number of annotations is set to “3”, then three annotations will be shown in a similar juxtaposed manner as shown in Figure 9 for comparison and resolution. We note that the function of collaborative data annotation has also been provided by the PSU GeoAnnotator (Karimzadeh & MacEachren, 2019), and the contribution of GALLOC is in its abilities to automatically detect annotation disagreements and to facilitate disagreement resolution via visual comparison and functions to edit current annotations.

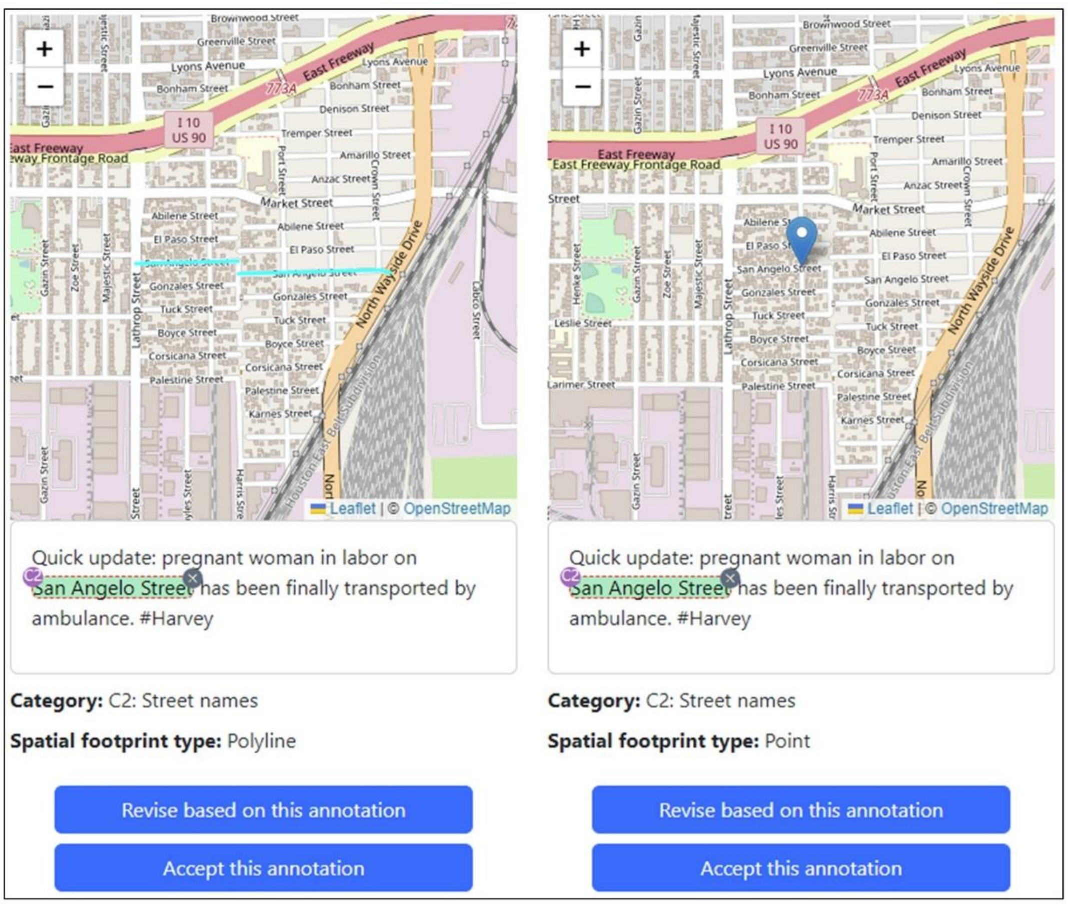

Figure 9. GALLOC shows two different annotations in juxtaposition for disagreement resolution.

5.5. Annotating multilingual datasets (Meeting DC 7)

GALLOC supports the annotation of disaster-related text messages in other languages beyond English. The dataset may contain messages in one single non-English language (e.g., Chinese); this is often the case for disasters happening in countries and regions where one primary non-English language is used. The dataset may also have a mixture of English and other languages (e.g., English and Spanish); this is often the case for disasters happening in countries and regions where English and other languages are both used by large populations. Users can also define location categories using their preferred non-English language.

Figure 10 provides two examples in which GALLOC is used to annotate text messages in Chinese and Spanish. Figure 10 (a) shows a Chinese message from Sina Weibo (a popular social media platform in China) during the flooding that happened in the Beijing-Tianjin-Hebei region in summer 2023. The message reported a road collapse at a location due to the heavy rainfall. As shown in the figure, we can use GALLOC to annotate this Chinese location description "*北京石景山京西大悦城*" (which refers to a large shopping mall on the side of the road collapse), and assign a Chinese location category "*人工建筑*" (i.e., human-made features) to this description. Figure 10 (b) shows a Spanish message posted during Hurricane Maria in 2017 on Twitter. The message reported a segment of Highway 184 that fell off the cliff. We can use GALLOC to annotate the location description, "Carretera 184 de Guavate hacia Patillas" (i.e., Highway 184 from Guanvate to Patillas) and then assign a Spanish category "Carretera" (i.e., Highway).

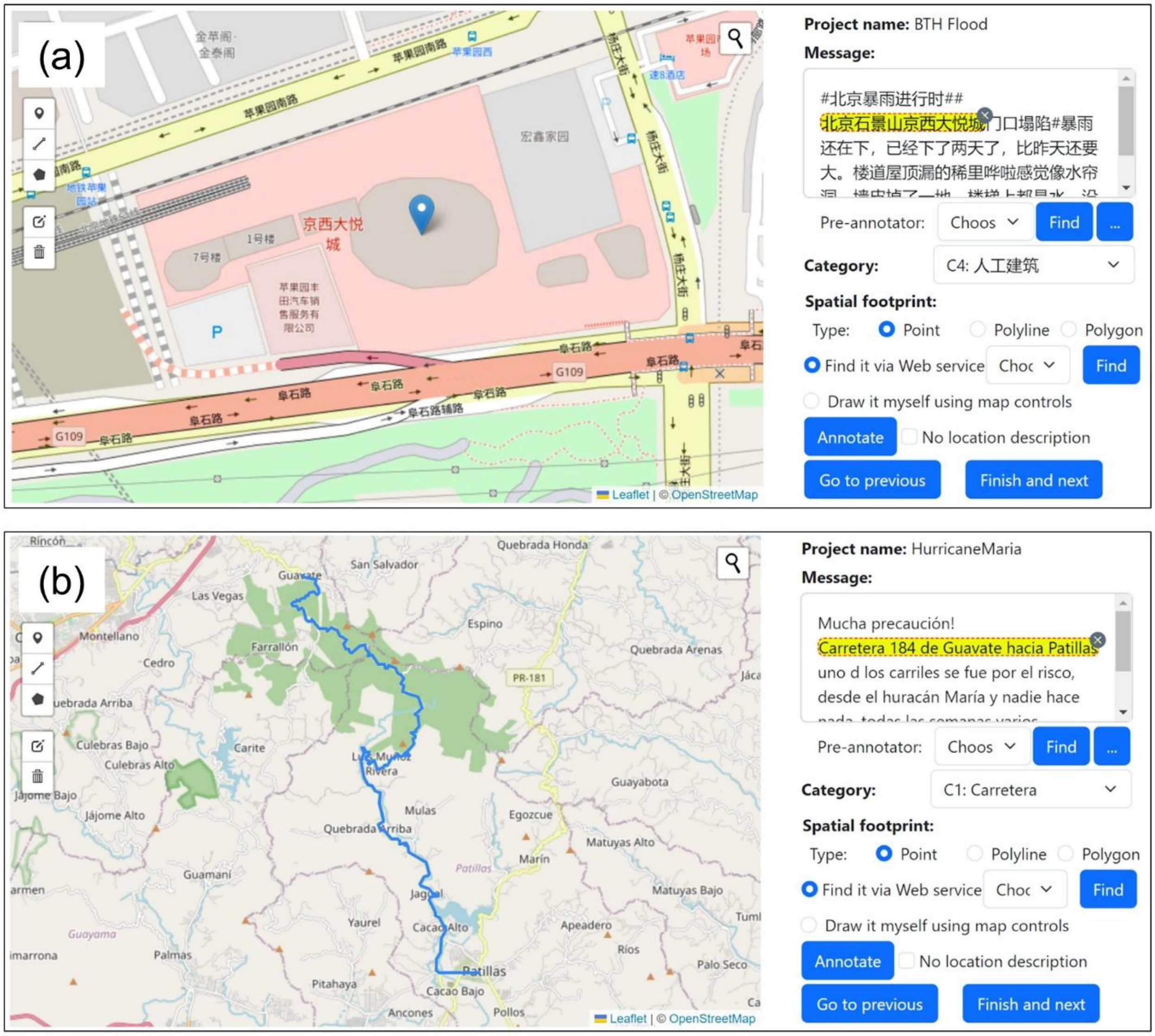

Figure 10. Using GALLOC to annotate disaster-related text messages in Chinese and Spanish.

6. Evaluation

In this section, we evaluate GALLOC by comparing it with previous geo-annotation tools and using it to annotate a small set of disaster-related text messages.

6.1. A comparison with previous geo-annotation tools

We compare GALLOC with three most closely related tools, i.e., WOTR GeoAnnotator (DeLozier et al., 2016), PSU GeoAnnotator (Karimzadeh & MacEachren, 2019), and GeoViz (McDonough et al., 2019). We compare these tools based on the seven design considerations, and the result is summarized in Table 1.

Table 1. A comparison of GALLOC and three previous geo-annotation tools.

Design Considerations	GALLOC	WOTR Geoannotator	PSU Geoannotator	GeoViz
Annotating complete location descriptions not separate entities	✓	✓		
Annotating categories of location descriptions based on a user-defined scheme	✓			
Providing pre-annotation of location descriptions from text	✓ (using a connection interface; can be updated with new AI techniques)		✓ (using six embedded NER tools; cannot be updated)	✓ (using an embedded Edinburgh Geoparser; cannot be updated)
Identifying spatial footprints automatically	✓ (using Nominatim and Google Maps to geo-locate descriptions)		✓ (using place name matching based on GeoNames)	✓ (using place name matching based on GeoNames)
Drawing customized spatial footprints	✓	✓		
Supporting multi-user collaborative annotation	✓		✓	
Annotating multilingual datasets	✓			

Overall, GALLOC provides effective support for annotating complete location descriptions and user-defined location categories from disaster-related text messages. It also provides a connection interface for pre-annotation tools and allows new pre-annotation tools, especially those based on new AI techniques, to be dynamically connected to the system. By providing automatic spatial footprint identification and manual drawing functions, GALLOC facilitates the creation of spatial footprints for location descriptions, and supports the use of different types of geometries (i.e., points, polylines, and polygons) for spatial footprints. GALLOC also provides support for collaborative data annotation by offering different user roles and privileges. Finally, GALLOC is developed with supporting multilingual data annotation in mind, and can support the annotation of text messages in languages beyond English.

The comparison with previous tools also reveals an important technology aging issue. Given the fast advancements of technologies, it is very difficult to deploy previous tools that are based on older technologies. The lack of documentation further exacerbates the difficulty of deployment. We note that this technology aging issue is not limited to the previous geo-annotation tools only but applies to all software tools including GALLOC. To mitigate the impacts of this issue, we package GALLOC using Docker which is an open-source platform that facilitates the deployment of software applications by packaging an application in its separate running environment (called a "container"). We have prepared a Docker container for GALLOC with all the required dependencies and libraries, made a Docker image of this container, and shared this image on Docker Hub (an online repository with container images contributed by developers). Users can directly pull this image from Docker Hub and quickly run it on their own computers. By using Docker, we simplify the deployment process of GALLOC. We also provide step-by-step documentation on how to run the Docker image. Since GALLOC will be running in its own container separated from the outside technologies that are changing, we mitigate the issue of technology aging and allow GALLOC to continue being used in the coming years.

6.2. Using GALLOC to annotate disaster-related text messages

We further evaluate GALLOC through a user-based experiment and assess user experiences in data annotation with and without GALLOC. In the following, we present the experiment setting and results.

6.2.1. Experiment setting

Conducting this user-based experiment requires careful thinking on its setting. The core idea is to run the experiment in two groups: a test group and a control group. In the test group, human annotators will be asked to annotate a set of text messages using GALLOC; in the control group, the same human annotators will be asked to annotate the same number of text messages without using GALLOC. While this experiment setting controls individual difference and message number difference, there may exist message content differences that can still affect experiment results. First, different text messages may contain different numbers of location descriptions. Since each location description needs to be annotated with its text span, category, and spatial footprint, a message that contains two location descriptions may require two times the amount of work as a message that contains only one location description. This means that even after we have controlled the total number of text messages in the two sets, they may still involve different workloads depending on the actual location descriptions contained. Second, different categories of location descriptions (e.g., a door number address vs. a road intersection) may also require different amounts of time for data annotation, especially for identifying their spatial footprints. One way to overcome this content difference challenge is to use the exact same messages in the two groups. However, such an experiment setting will likely run into a memory effect, since the human annotators may still remember those messages when they annotate the same messages for the second time. This memory effect will likely affect the accuracy of the recorded data annotation time.

To mitigate the impacts of this content difference and to avoid the memory effect, we designed our experiment by making use of a dataset shared by a previous study (Y. Hu et al., 2023), which contains 1000 text messages (tweets) from Hurricane Harvey. Location descriptions in these text messages were previously annotated with their text spans and categories. As a result, we were able to control the numbers and categories of location descriptions contained in the text messages of the two sets. By writing a simple Python script, we selected two sets of messages from this dataset; each set contains 50 messages, and each message contains exactly one location description. We also ensured that messages in the two sets have the same distributions across different location categories (there are 11 location categories in total in the original dataset). With this experiment setting, we ensured that the two sets of messages have the same total number of messages, the same number of location descriptions, and the same distribution of location descriptions across categories. The actual contents of the messages in the two sets are still different (since they are different messages), but they are from the same disaster and are all within the same 140-character limitation used by Twitter in 2017. Thus, the messages from the two sets can be considered comparable, and they allow us to avoid the memory effect. We used only the textual content of the messages to prepare the two sets, and the annotations from the original data were not included.

The experiment was conducted as follows. Five human annotators with some background in GIS or disaster research were invited to participate in this experiment. These human annotators were asked to annotate the first set of 50 messages using GALLOC (the test group) and the second set of 50 messages without using GALLOC (the control group). For the second set, human annotators were asked to use a simple text editor, such as Notepad on Windows and TextEdit on macOS. In both groups, human annotators were allowed to use external websites, such as Google Maps, to search for location descriptions when necessary. A detailed data annotation guideline was created for this experiment (the created guideline is also shared in the repository). Before the data annotation began, a training session and a Q&A session were held to introduce the human annotators to this task and to answer questions. After these sessions, human annotators were given a flexible amount of time to get themselves familiar with this task and the tools used; after that, each human annotator was asked to take a test by annotating 10 messages using GALLOC and 10 messages without using GALLOC. Note that these 20 messages in the test were outside of the 100 messages used in the experiment. A human annotator was considered as ready to do the formal data annotation if they passed the test with over 80% annotation correctness. Human annotators then worked on annotating the two sets of messages, and a discussion session was held afterward to discuss the experience of the human annotators in doing data annotation with and without GALLOC.

6.2.2. Experimental results

6.2.2.1. Data annotation efficiency and accuracy

We first examine data annotation efficiency of the two groups. Human annotators were asked to record two time values for each message when doing data annotation: (1) time spent in annotating the location text and category; and (2) total time for annotating the message. We asked human annotators to record these two time values, instead of only the total time, because we noticed from the training sessions and the tests that human annotators tended to have different goals when they were creating the spatial footprints with and without GALLOC. When GALLOC was used, human annotators often wanted to create more detailed spatial footprints because they had the map services and the drawing tools of GALLOC that allowed them to do so. By contrast, when GALLOC was not used (i.e., when a simple text editor was used), human annotators often just settled with a very simple spatial footprint (e.g., a point or a bounding box) because they felt that it was practically impossible for them to create more detailed spatial footprints. Recording these two time values therefore allows us to further separate the potential impacts of having different goals for creating spatial footprints. We use the metric *Time Saved (TS)* in Equation (1) to quantify the improved data annotation efficiency:

$$TS = \frac{T_O - T_W}{To} \quad , \tag{1}$$

where To is the time spent without using GALLOC and T_W is the time spent with GALLOC.

Figure 11 shows the time difference for data annotation without and with GALLOC, and subfigures (a), (b), and (c) show the time for annotating text span and category only, spatial footprint, and total time respectively. The metric *Time Saved* is calculated using the median annotation time without and with GALLOC. As can be seen in Fig. 11(a), GALLOC substantially reduced the annotation time for annotating the text span and location category compared with not using GALLOC. For all five human annotators, their 75th percentile of the annotation time using GALLOC is even lower than the 25th percentile of the annotation time without GALLOC. In Fig. 11(b), annotation time for spatial footprints was calculated by subtracting the recorded text and category annotation time from the total annotation time. As can be seen, GALLOC also largely reduced the median time of annotating spatial footprints; however, the 75th percentile of the annotation times of the two groups are similar. We believe that the large reduction of median annotation time likely comes from the cases when spatial footprints can be automatically identified by GALLOC; in cases when spatial footprints cannot be directly identified, manually drawing them in GALLOC could take similar time as recording a simple bounding box without GALLOC. In terms of the total annotation time, GALLOC provided from 48% to 79% efficiency improvement across the five human annotators, which can be seen in Fig. 11(c).

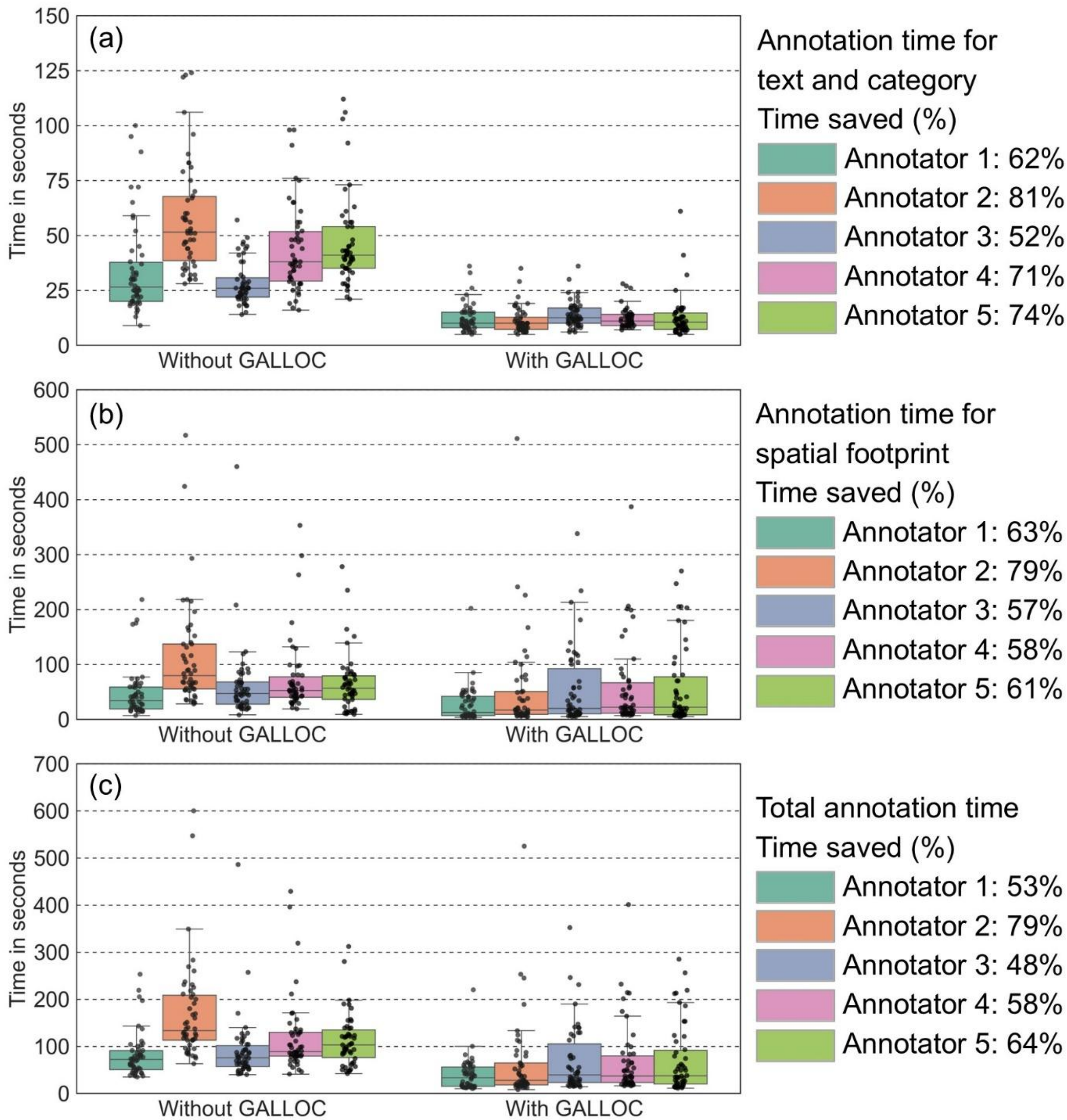

Figure 11. Time recorded for data annotation without and with GALLOC: (a) annotation time for location text and category only; (b) annotation time for spatial footprints; and (c) total annotation time.

We further examine the data annotation accuracy of the two groups. Because the original dataset has location descriptions and categories annotated, we are able to compare the annotations from the two groups with the ground truth. We assess annotation accuracy via three metrics, *precision*, *recall*, and *F-score*, based on Equations (2)-(4):

$$Precision = \frac{|Correctly\ annotated|}{|All\ annotated|} \quad , \qquad (2)$$

$$Recall = \frac{|Correctly\ annotated|}{|All\ correct|} \quad , \qquad (3)$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad . \tag{4}$$

One special situation in our experiment is that each message contains only one location description, and human annotators are informed about this experiment setting. Thus, |*All annotated*| in Equation (1) and |*All correct*| in Equation (2) have the same value of 50, which leads to the same values for precision, recall, and F-score. Given this special situation, we only report F-scores to reduce redundancy, which are shown in Figure 12. Note that an annotated location description is considered as correct only when it has the same text span and the same location category as in the ground truth. As can be seen, the human annotators all achieved high accuracy regardless of whether GALLOC is used. We note that data annotation accuracy is affected by both the tool and the human annotator. A human annotator with good attention to details is likely to achieve high accuracy for labeling the text span and category even without GALLOC. For Annotator 5, a slight decrease of F-score (from 0.92 to 0.9) was observed when GALLOC was used. For the other four human annotators, using GALLOC slightly increased their data annotation accuracy (e.g., an increase of F-score from 0.78 to 0.9 for Annotator 1), although the extent of increase varied across the annotators.

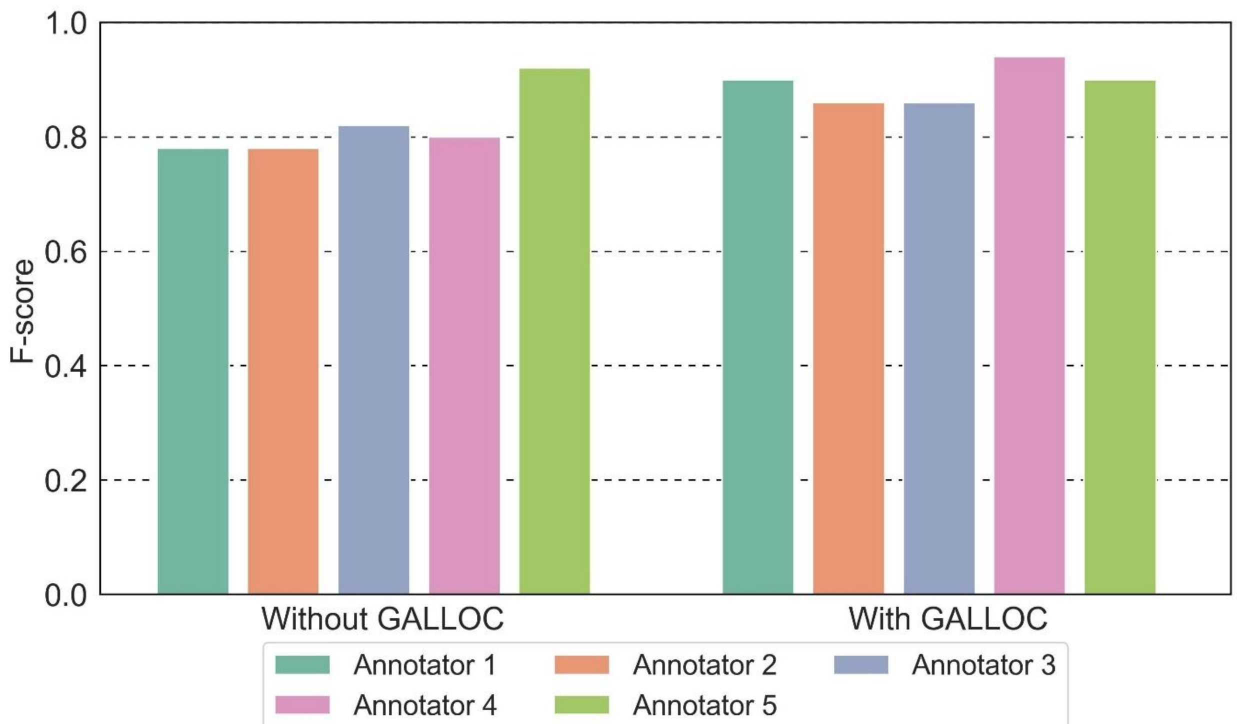

Figure 12. Data annotation accuracy (measured by F-score) without and with GALLOC.

6.2.2.2. Spatial footprints created for location descriptions

Because the original dataset does not contain labeled spatial footprints, we cannot directly assess the accuracy of the spatial footprints created by the human annotators in this experiment. However, we can assess the level of detail of the created spatial footprints in the two groups. Specifically, we measure the number of vertices of the spatial footprints associated with the correct annotations (i.e., correct text span and category). Table 2 shows the median and mean numbers of vertices of the spatial footprints in the two groups. Note that these median and mean numbers are calculated based on only the location categories that are likely to have non-point spatial footprints (e.g., roads and administrative units). Location categories whose spatial footprints should ideally be points, such as door number addresses and road intersections, are not included in the calculation. As can be seen in Table 2, the spatial footprints created with GALLOC generally have larger median vertice numbers and much larger mean vertice numbers than those created without GALLOC, suggesting a higher level of detail. The median and mean numbers of vertices without GALLOC are both small (both are smaller than 3), suggesting that human annotators only created simple geometries when GALLOC was not used. By contrast, the mean vertice number is much larger than the median vertice number when GALLOC was used, suggesting that some very detailed geometries were created and were likely directly returned by the Nominatim map service in GALLOC.

Table 2. Median and mean vertice numbers of spatial footprints for non-point location categories created without and with GALLOC.

Median / Mean	***Human Annotators***				
	Annotator 1	Annotator 2	Annotator 3	Annotator 4	Annotator 5
Without GALLOC	1 / 2.18	1 / 2.97	**2** / 2.68	1 / 1.42	1 / 2.54
With GALLOC	**2 / 492.82**	**2 / 507.72**	1 / **268.77**	**5 / 474.94**	**5 / 262.76**

The ability of GALLOC to support the creation of spatial footprints can also be seen through some concrete examples. Figure 13 provides three examples showing three common situations in which GALLOC made it easier for a human annotator to identify and create the spatial footprint. In Fig. 13(a), GALLOC helps quickly identify and create the precise and detailed footprint of "Deer Park" (which is a small city/town in the Houston metropolitan area) via the Nominatim map service; without GALLOC, it is very difficult, if not impossible, for one to create such a detailed spatial footprint. Since Nominatim is based on OpenStreetMap data, GALLOC can automatically identify and create detailed spatial footprints for many geographic features throughout the world. In Fig. 13(b), the spatial footprint of the location description "I-45 at Main Street" cannot be directly identified by a typical map service including Google Maps and Nominatim; however, GALLOC allows the human annotator to separately identify the footprints of "I-45" and "Main Street", and then quickly locate their intersection. In Fig. 13(c), the location description "Almeda Genoa Rd. from Windmill Lakes Blvd. to Rowlett Rd." again cannot be directly geo-located by a map service. With the help of GALLOC, one can find the footprint of this location description by first identifying the three roads involved and then quickly locating the specific road segment. Without GALLOC, it is very difficult to locate this type of road segment location description.

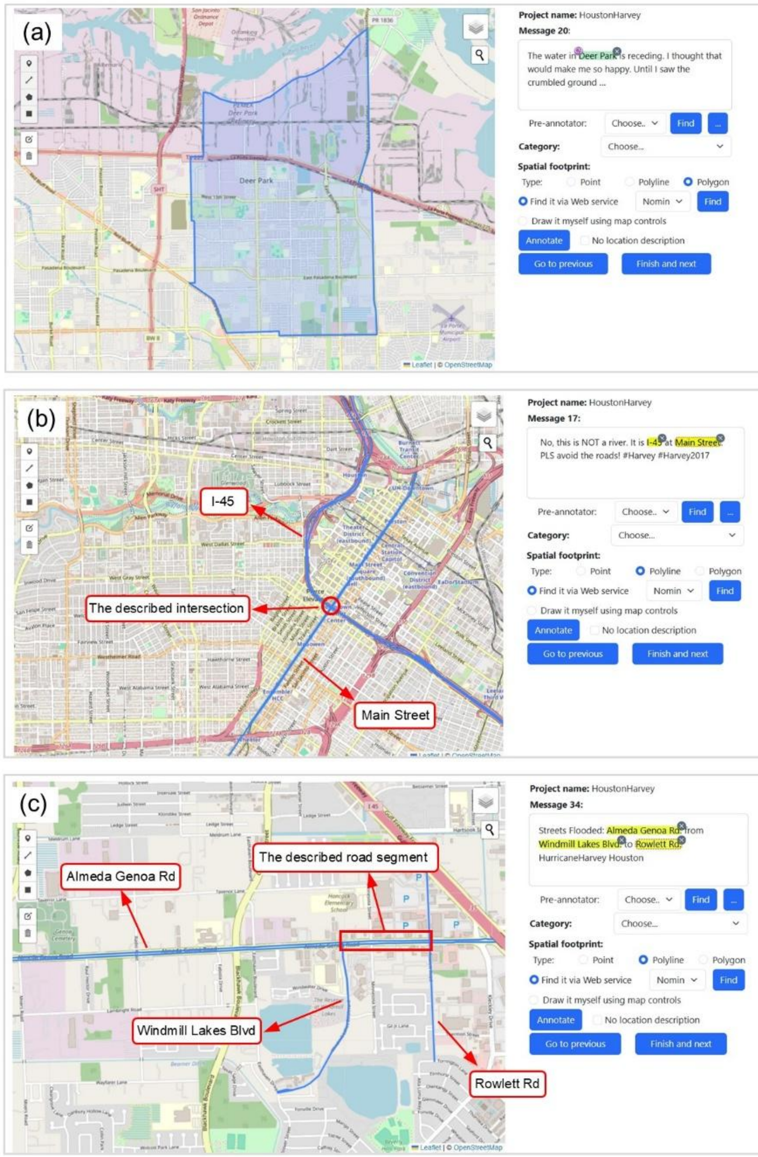

Figure 13. Examples of GALLOC supporting the identification and creation of spatial footprints.

We also queried the location descriptions in this experiment against the GeoNames gazetteer to see how many of them can be found. Because the GeoNames gazetteer has been used in previous geo-annotation tools to locate toponyms, this query allows us to roughly assess the number of location descriptions that cannot be annotated using previous tools. For the 50 messages annotated using GALLOC (which contain 50 location descriptions), we found that 40 location descriptions (i.e., 80%) cannot be found in GeoNames. We note that this high percentage is likely due to the fact that this dataset focuses on fine-grained location descriptions used in a disaster context, such as door number addresses, road intersections, and road segments, which are not included in a typical gazetteer. For another dataset, such as a dataset of news articles containing mostly city names and country names, it could be that most toponyms can be found in GeoNames.

6.2.2.3. Data annotation workflow

After the data annotation task, we held a discussion session to learn the experience of the human annotators with and without GALLOC. The discussion session revealed several improvements provided by GALLOC on the data annotation workflow. First, for the data preparation step, human annotators almost did not need to do anything when using GALLOC; in comparison, they had to first download the text messages to be annotated, unzip the files, and organize the files for annotation when GALLOC was not used. We note that GALLOC also requires one person to set up the project; however, other human annotators do not need to do the data preparation work again once the project is set up. Second, for the process of doing data annotation, human annotators shared that they had to switch among several tools and applications (e.g., the text editor, Google Maps, and Google Search) when GALLOC was not used. While human annotators still needed to do external searches from time to time when using GALLOC, they could use GALLOC only for a majority of the text messages and experienced less hassle from switching apps. Third, for the data annotation results, while human annotators were not asked to do disagreement resolution in this experiment, we showed in the discussion session that GALLOC can facilitate disagreement resolution by automatically detecting disagreements and visualizing them in a juxtaposed manner for resolution. In comparison, it can be very difficult for human annotators to resolve annotation disagreements using a text editor.

6.2.2.4. Summary and further improvements

In sum, the experiment result suggests that GALLOC can increase data annotation efficiency and accuracy, support the creation of more detailed spatial footprints, and improve several aspects of the data annotation workflow. The enhancements brought by GALLOC are also seen across different human annotators. While the experiment showed that GALLOC is generally an effective tool, it also identified some limitations which have been addressed afterward. For example, human annotators suggested adding a choice of satellite-image basemap which can help them check the geographic features on the ground when the map-based visualization is unclear. This feature is now added. We also made some other small improvements based on user feedback, such as highlighting the search result of the search function provided by GALLOC. We hope that these and other improvements can help GALLOC better support future data annotation tasks.

7. Discussion

7.1. Data annotation gaps filled by GALLOC for AI and disaster management

There has been a lot of interest in using AI for enhancing disaster management (Kuglitsch et al., 2022; Suwaileh et al., 2022; Zhou et al., 2022). It has been recognized that AI, especially trustworthy AI, heavily relies on the availability of high-quality labeled training data (Liang et al., 2022), and data annotation tools play important roles in facilitating the creation of such high-quality data. In the era of large language models (LLMs), high-quality training data in different languages and from different geographic areas have become even more critical for reducing the biases of large AI models. In the field of geographic information retrieval (GIR) (Jones & Purves, 2008; Purves et al., 2018; Stock et al., 2022; X. Hu et al., 2023), researchers have also discussed the importance of geo-annotation tools for creating labeled datasets and have developed corresponding tools (Karimzadeh & MacEachren, 2019; McDonough et al., 2019; Y. Hu & Wang, 2021; X. Hu et al., 2024). Existing geo-annotation tools are effective in annotating data for typical GIR applications, such as recognizing and geo-locating city names and country names from Web pages. When it comes to location descriptions under a disaster context, gaps exist as these descriptions often cannot be linked to a gazetteer entry due to their multi-entity nature (e.g., door number addresses, road intersections, and highway exits) and may need to be annotated with user-defined location categories for geo-locating purposes. GALLOC, therefore, fills these gaps by enabling the annotation of complete location descriptions and user-defined location categories from disaster-related text messages. It also enables the annotation of text messages in different languages beyond English.

Another data annotation gap filled by GALLOC is the creation of spatial footprints beyond points. It has been long recognized that point-based geometry representation is often not suitable for linear and areal geographic features (Speriosu & Baldridge, 2013; Weissenbacher et al., 2015; Karimzadeh et al., 2019; Leppämäki et al., 2024). Yet, points are still used in most existing geoparsing datasets to represent the spatial footprints of toponyms. One possible reason is the difficulty of creating polyline- or polygon-based spatial footprints without an effective tool. Typical gazetteers, like GeoNames, only provide point-based spatial footprints (Acheson et al., 2017), and manually drawing a line for a river or a polygon for a city district is highly labor intensive. GALLOC facilitates the creation of polyline- and polygon-based spatial footprints by providing these more complex footprints from OpenStreetMap through Nominatim. In cases when the OpenStreetMap-based geometries are not accurate, GALLOC also allows users to edit those geometries rather than creating the spatial footprints from scratch. Meanwhile, GALLOC also supports the creation of point-based spatial footprints for suitable geographic features, such as road intersections. The need for annotated datasets with more complete spatial footprints for robust experiments has also been discussed in recent research (Leppämäki et al., 2024), and GALLOC fills such a data annotation gap.

7.2. Use of GALLOC, design considerations, and data privacy

GALLOC can be used by AI model developers, disaster researchers, and emergency managers for two major purposes. First, it can be used for creating labeled datasets for training AI models. These datasets, labeled with complete location descriptions, location categories, and spatial footprints, allow the trained AI models to acquire corresponding capabilities to identify complete location descriptions from text, classify location categories, and geo-locate their spatial footprints. The annotated datasets can also be used to guide LLMs, as shown in recent research (Y. Hu et al., 2023; Mai et al., 2024). Second, GALLOC can be used for studying typical location descriptions used by people in a geographic region. The ways that people describe locations likely vary from country to country due to language differences and different address systems. The multilingual support of GALLOC facilitates this type of studies across countries and languages. The obtained knowledge about location descriptions can then be used by emergency managers to increase community resilience in various ways. For example, emergency managers may identify inaccurate location descriptions used by people in their managed geographic area, and may conduct community outreach activities to help people learn more effective ways to communicate locations during a disaster.

Seven design considerations have been formalized before the development of GALLOC. These design considerations are formalized based on literature review, existing data annotation tools, and the needs of annotating location descriptions from disaster-related text messages. They have the potential to inform the development of similar data annotation tools. For example, the pre-annotation function can greatly increase data annotation efficiency; yet, pre-annotation tools and related AI techniques are changing rapidly, and people may have different preferences to pre-annotation tools as well. *Design Consideration 3* suggests the use of a connection interface, rather than an embedded and fixed pre-annotation tool, that allows users to connect their own pre-annotation tools and to replace the tools with newer versions as technology evolves. Similarly, to support the collaboration of a group of users, *Design Consideration 6* suggests the inclusion of different user roles with different privileges to facilitate this collaboration and coordinate annotation tasks. GALLOC also dynamically adjusts its user interface by showing only the functions relevant to a user to help them focus on their assigned tasks and reduce potential distractions and mistakes. These design considerations could be re-used for the development of similar data annotation tools that also need to provide pre-annotation functions and support multi-user collaboration.

Creating an annotated dataset could raise data privacy concerns that require our careful attention. For GALLOC, we provide an online demo that allows the annotation of a simple dataset without requiring the user to deploy the tool. Using the online demo, however, requires the dataset to be uploaded to our university server and could have data privacy issues when a network security breach happens. This can be especially of concern when the data to be annotated are sensitive, e.g., help requests from vulnerable population groups (Y. Hu & Wang, 2021; Zou et al., 2022). To better protect privacy, we believe that conducting data annotation in a secure local computing environment (without uploading data to any online server) is a safer approach. We have designed GALLOC in a way that only needs relatively simple steps for local deployment. In addition, the data annotation process can also trigger privacy concerns, since the text messages have to be read by human annotators. We see two possible approaches to mitigating the privacy issues. First, training sessions can be held for human annotators before a data annotation task to increase their awareness of the privacy issues that may be involved in the data annotation process and help them learn best practices for privacy protection. Second, algorithms could be developed to encrypt location descriptions before annotation and decrypt them afterward. The encryption could involve changing door numbers or shifting some characters of road names based on certain rules. Nevertheless, these algorithms should be used carefully, as they could fail to detect location descriptions or make text messages difficult for human annotators to read. A combination of training sessions for human annotators and careful use of encryption algorithms may help protect privacy and support the creation of a labeled dataset.

7.3. Potential research directions enabled by GALLOC

As a geo-annotation tool, GALLOC enables potential research directions by helping researchers create needed datasets more quickly. Several research directions could be explored with the help of GALLOC. First, we can create datasets to study location description differences across different types of disasters, such as hurricanes, earthquakes, wildfires, and snow storms. These annotated datasets can help us understand the similarities and differences of location descriptions across disasters, and help emergency managers potentially focus on certain location descriptions closely associated with a type of disaster. Second, we can create datasets to study the regional differences of disaster-related location descriptions. In addition to country and language differences, location descriptions could also be different across urban and rural geographic regions in which people may prefer to use man-made or natural geographic features to describe their locations under a disaster context. Finally, while GALLOC is designed with disaster management in mind, it can also be used to create datasets to study location descriptions in other contexts beyond disasters. With much flexibility in annotating text spans, location categories, and spatial footprints, GALLOC enables researchers to annotate the description text based on their own interests, create their own sets of location categories, and choose the types of spatial footprints that best meet their research needs.

8. Conclusions

People post text messages on various platforms during natural disasters to seek help and share urgent information. Many of these text messages contain important location descriptions about victims and accidents, and accurately recognizing and geo-locating these location descriptions can help disaster responders reach victims more quickly and even save lives. Location descriptions in disaster-related text messages often consist of multiple entities, such as door number addresses, road intersections, and highway exits, and cannot be effectively extracted using typical NER approaches. Training new machine learning models to recognize these location descriptions requires datasets labeled with complete location descriptions and location categories. Creating such datasets, however, is time-consuming and labor-intensive without an effective data annotation tool. In this work, we develop GALLOC, a GeoAnnotator for Labeling LOCation descriptions from disaster-related text messages. GALLOC is developed as a Web-based and open-source tool. It follows seven design considerations and supports the annotation of complete location descriptions and user-defined location categories. GALLOC also provides AI-powered pre-annotation, supports multi-user collaborative data annotation, and can be used to annotate text messages in other languages in addition to English. By facilitating the creation of labeled datasets, GALLOC can help train new AI models and answer new research questions related to disaster management, location descriptions, and beyond.

Acknowledgement

We thank the anonymous reviewers for their constructive comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data and codes availability statement

The data and codes that support the findings of this study are available on figshare at: https://doi.org/10.6084/m9.figshare.26768089.

Funding

This work is supported by the U.S. National Science Foundation under Grant No. BCS-2117771 and BCS-2416886. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Notes on contributors

Kai Sun is a Postdoctoral Associate in the Department of Geography at the University at Buffalo. His research focuses on geospatial artificial intelligence (GeoAI) for disaster management. His contributions to this paper include conceptualization, data collection and curation, methodology, data analysis, result interpretation and discussion, visualization, writing – original draft, and writing – editing and revision.

Yingjie Hu is an Associate Professor in the Department of Geography at the University at Buffalo. His research interests include GIScience, GeoAI, and disaster resilience. His contributions to this paper include conceptualization, data collection and curation, methodology, data analysis, result interpretation and discussion, writing – original draft, and writing – editing and revision.

Kenneth Joseph is an Associate Professor in the Department of Computer Science and Engineering at the University at Buffalo. He is a computational social scientist who focuses on the measurement and modeling of complex social systems. His contributions to this paper include conceptualization and writing – editing and revision.

Ryan Zhenqi Zhou (Zhenqi Zhou) is a Ph.D. candidate in the Department of Geography at the University at Buffalo and a research assistant in the GeoAI Lab. His research interests include GeoAI, disaster resilience, public health, and human mobility. His contributions to this paper include data collection and curation, data analysis, and writing – editing and revision.

References

Acheson, E., De Sabbata, S., & Purves, R. S. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems*, *64*, 309–320.

Alex, B., Byrne, K., Grover, C., & Tobin, R. (2014). A web-based geo-resolution annotation and evaluation tool. *Proceedings of LAW VIII-The 8th Linguistic Annotation Workshop*, 59–63.

Alex, B., Byrne, K., Grover, C., & Tobin, R. (2015). Adapting the Edinburgh geoparser for historical georeferencing. *International Journal of Humanities and Arts Computing*, *9*(1), 15–35.

Berragan, C., Singleton, A., Calafiore, A., & Morley, J. (2023). Transformer based named entity recognition for place name extraction from unstructured text. *International Journal of Geographical Information Science*, *37*(4), 747–766. https://doi.org/10.1080/13658816.2022.2133125

Bontcheva, K., Cunningham, H., Roberts, I., Roberts, A., Tablan, V., Aswani, N., & Gorrell, G. (2013). GATE Teamware: A web-based, collaborative text annotation framework. *Language Resources and Evaluation*, *47*, 1007–1029.

Chen, P., Xu, H., Zhang, C., & Huang, R. (2022). Crossroads, Buildings and Neighborhoods: A Dataset for Fine-grained Location Recognition. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3329–3339.

Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities*, *36*, 223–254.

DeLozier, G., Wing, B., Baldridge, J., & Nesbit, S. (2016). Creating a novel geolocation corpus from historical texts. *Proceedings of the 10th Linguistic Annotation Workshop Held in Conjunction with ACL 2016 (LAW-X 2016)*, 188–198.

Fernández-Martínez, N. J. (2022). The FGLOCTweet Corpus: An English tweet-based corpus for fine-grained location-detection tasks. *Research in Corpus Linguistics*, *10*(1), 117–133.

Gelernter, J., & Balaji, S. (2013). An algorithm for local geoparsing of microtext. *GeoInformatica*, *17*(4), 635–667.

Gelernter, J., & Mushegian, N. (2011). Geo-parsing messages from microtext. *Transactions in GIS*, *15*(6), 753–773.

Gritta, M., Pilehvar, M. T., & Collier, N. (2018). *Which melbourne? Augmenting geocoding with maps*.

Hu, X., Al-Olimat, H. S., Kersten, J., Wiegmann, M., Klan, F., Sun, Y., & Fan, H. (2022). GazPNE: annotation-free deep learning for place name extraction from microblogs leveraging gazetteer and synthetic data by rules. *International Journal of Geographical Information Science*, *36*(2), 310–337.

Hu, X., Elßner, T., Zheng, S., Serere, H. N., Kersten, J., Klan, F., & Qiu, Q. (2024). DLRGeoTweet: A comprehensive social media geocoding corpus featuring fine-grained places. *Information Processing & Management*, *61*(4), 103742.

Hu, X., Zhou, Z., Li, H., Hu, Y., Gu, F., Kersten, J., Fan, H., & Klan, F. (2023). Location reference recognition from texts: A survey and comparison. *ACM Computing Surveys*, *56*(5), 1–37.

Hu, Y., Mai, G., Cundy, C., Choi, K., Lao, N., Liu, W., Lakhanpal, G., Zhou, R. Z., & Joseph, K.

(2023). Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science*, *37*(11), 2289–2318.

Hu, Y., & Wang, J. (2021). How do people describe locations during a natural disaster: An analysis of tweets from Hurricane Harvey. *Proceedings of the 11th International Conference on Geographic Information Science*. GIScience, Poznan, Poland.

Huang, Q., & Xiao, Y. (2015). Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information*, *4*(3), 1549–1568.

Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval. *International Journal of Geographical Information Science*, *22*(3), 219–228.

Ju, Y., Adams, B., Janowicz, K., Hu, Y., Yan, B., & McKenzie, G. (2016). *Things and Strings: Improving Place Name Disambiguation from Short Texts by Combining Entity Co-Occurrence with Topic Modeling* (E. Blomqvist, P. C. F. Poggi, & F. Vitali, Eds.; Vol. 10024, pp. 353–367). Springer.

Karimiziarani, M., & Moradkhani, H. (2023). Social response and Disaster management: Insights from twitter data Assimilation on Hurricane Ian. *International Journal of Disaster Risk Reduction*, *95*, 103865. https://doi.org/10.1016/j.ijdrr.2023.103865

Karimzadeh, M., & MacEachren, A. M. (2019). GeoAnnotator: A collaborative semi-automatic platform for constructing geo-annotated text corpora. *ISPRS International Journal of Geo-Information*, *8*(4), 161.

Karimzadeh, M., Pezanowski, S., MacEachren, A. M., & Wallgrün, J. O. (2019). GeoTxt: A scalable geoparsing system for unstructured text geolocation. *Transactions in GIS*, *23*(1), 118–136.

Kuglitsch, M. M., Pelivan, I., Ceola, S., Menon, M., & Xoplaki, E. (2022). Facilitating adoption of AI in natural disaster management through collaboration. *Nature Communications*, *13*(1), 1–3.

Leidner, J. L. (2006). An evaluation dataset for the toponym resolution task. *Computers, Environment and Urban Systems*, *30*(4), 400–417.

Leppämäki, T., Toivonen, T., & Hiippala, T. (2024). Geographical and linguistic perspectives on developing geoparsers with generic resources. *International Journal of Geographical Information Science*, *0*(0), 1–22. https://doi.org/10.1080/13658816.2024.2369539

Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence*, *4*(8), 669–677.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R., & Lao, N. (2024). On the Opportunities and Challenges of Foundation Models for GeoAI (Vision Paper). *ACM Transactions on Spatial Algorithms and Systems*, *10*(2), 1–46. https://doi.org/10.1145/3653070

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60.

McDonough, K., Moncla, L., & Van de Camp, M. (2019). Named entity recognition goes to old regime France: Geographic text analysis for early modern French corpora. *International Journal of Geographical Information Science*, *33*(12), 2498–2522.

Neves, M., & Ševa, J. (2021). An extensive review of tools for manual annotation of documents. *Briefings in Bioinformatics*, *22*(1), 146–163.
Purves, R. S., Clough, P., Jones, C. B., Hall, M. H., & Murdock, V. (2018). Geographic information retrieval: Progress and challenges in spatial search of text. *Foundations and Trends® in Information Retrieval*, *12*(2–3), 164–318.
Ratinov, L., & Roth, D. (2009). Design challenges and misconceptions in named entity recognition. *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, 147–155.
Rhodan, M. (2017). Hurricane Harvey: The U.S.'s first social media storm. *Time Magazine*. https://time.com/4921961/hurricane-harvey-twitter-facebook-social-/
Seetharaman, D., & Wells, G. (2017). Hurricane Harvey Victims Turn to Social Media for Assistance. *The Wall Street Journal*. https://www.wsj.com/articles/hurricane-harvey-victims-turn-to-social-media-for-assistance-1503999001
Silverman, L. (2017). Facebook, Twitter Replace 911 Calls For Stranded In Houston. *National Public Radio*. https://www.npr.org/sections/alltechconsidered/2017/08/28/546831780/texas-police-and-residents-turn-to-social-media-to-communicate-amid-harvey
Speriosu, M., & Baldridge, J. (2013). *Text-Driven Toponym Resolution using Indirect Supervision*. 1466–1476.
Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. (2012). BRAT: a web-based tool for NLP-assisted text annotation. *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 102–107.
Stock, K., Jones, C. B., Russell, S., Radke, M., Das, P., & Aflaki, N. (2022). Detecting geospatial location descriptions in natural language text. *International Journal of Geographical Information Science*, *36*(3), 547–584.
Suwaileh, R., Elsayed, T., & Imran, M. (2023). IDRISI-RE: A generalizable dataset with benchmarks for location mention recognition on disaster tweets. *Information Processing & Management*, *60*(3), 103340.
Suwaileh, R., Elsayed, T., Imran, M., & Sajjad, H. (2022). When a disaster happens, we are ready: Location Mention Recognition from crisis tweets. *International Journal of Disaster Risk Reduction*, 103107.
Tsujimoto, B. (2023). *How a 70,000-person Facebook group saved lives – and hope – during Buffalo's blizzard*. https://buffalonews.com/news/local/how-a-70-000-person-facebook-group-saved-lives-and-hope-during-buffalos-blizzard/article_d8991b2e-87be-11ed-8f49-c7f17f78555b.html
Wallgrün, J. O., Karimzadeh, M., MacEachren, A. M., & Pezanowski, S. (2018). GeoCorpora: Building a corpus to test and train microblog geoparsers. *International Journal of Geographical Information Science*, *32*(1), 1–29.
Wang, J., Hu, Y., & Joseph, K. (2020). NeuroTPR: A neuro-net toponym recognition model for extracting locations from social media messages. *Transactions in GIS*, *24*(3), 719–735.
Weissenbacher, D., Tahsin, T., Beard, R., Figaro, M., Rivera, R., Scotch, M., & Gonzalez, G. (2015). Knowledge-driven geospatial location resolution for phylogeographic models of virus migration. *Bioinformatics*, *31*(12), i348–i356.
Yimam, S. M., Gurevych, I., de Castilho, R. E., & Biemann, C. (2013). Webanno: A flexible, web-based and visually supported system for distributed annotations. *Proceedings of the*

51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 1–6.

Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C. (2019). Deep learning for real-time social media text classification for situation awareness – using Hurricanes Sandy, Harvey, and Irma as case studies. *International Journal of Digital Earth*, *12*(11), 1230–1247. https://doi.org/10.1080/17538947.2019.1574316

Zhou, B., Zou, L., Hu, Y., Qiang, Y., & Goldberg, D. (2023). TopoBERT: A plug and play toponym recognition module harnessing fine-tuned BERT. *International Journal of Digital Earth*, *16*(1), 3045–3063. https://doi.org/10.1080/17538947.2023.2239794

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., Cai, H., Abedin, J., & Mandal, D. (2022). VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. *Computers, Environment and Urban Systems*, *95*, 101824.

Zou, L., Liao, D., Lam, N. S., Meyer, M. A., Gharaibeh, N. G., Cai, H., Zhou, B., & Li, D. (2022). Social media for emergency rescue: An analysis of rescue requests on Twitter during Hurricane Harvey. *International Journal of Disaster Risk Reduction*, 103513.