Detection and Categorization of Needs during Crises Based on Twitter Data

Pingjing Yang[1], **Ly Dinh**[2], **Alex Stratton**[3], **Jana Diesner**[1, 4]

[1]University of Illinois Urbana-Champaign
[2]University of South Florida
[3]University of California Berkeley
[4]Technical University of Munich
py2@illinois.edu, lydinh@usf.edu, als217@berkeley.edu, jana.diesner@tum.de

Abstract

The Ukraine-Russia conflict has brought sizable detrimental impact to the global energy, food, finance, and manufacturing industries, as well to many affected people. In this paper, we use Twitter (now X) to automatically identify who needs what from text data and how the types of needs that we categorized and standardize evolved throughout this conflict. Our findings suggest that the Ukraine expresses a need for weapons, Russia for land, Europe for gas, and America for leadership. The majority of needs expressed on Twitter during this conflict are related to the categories transportation, military, health & medical, financial and money, energy, and essential items (food, water, shelter, non-food items). Stated needs changed as the conflict escalated or fell into stalemate. Needs also varied depending on the tweet's location, with tweets from Ukraine's neighboring countries being related to food and medicine, while tweets from non-neighboring countries stated needs for clothing and tents. Tweets written in Ukrainian and Russian shared similar need terms, such as medicines and kits, compared to English tweets, which expressed needs such as ammunition and humanitarian aid. Our comparison of needs across four different disaster events, namely this conflict, an earthquake, a major hurricane, and the COVID-19 pandemic, showed how needs differ depending on the nature of the crisis and how domain-adjustment of needs categories is necessary. We contribute to the crisis informatics literature by (1) validating a methodology for using tweets to study the demand and supply of things that different stakeholders need during crisis events and (2) testing, comparing, and improving the fit of widely used need classification schemas for studying crisis from different domains.

Introduction

Existing crisis response guidelines are designed to be of general use and applicable to all disasters such that they can provide a foundational framework for emergency management (U.S. Department of Homeland Security 2008; Alexander 2005). For instance, the Federal Emergency Management Agency (FEMA) created the National Response Framework to coordinate federal disaster response efforts across various types of disasters and offer guidance to federal, state, territorial, tribal, and local authorities in the United States (U.S. Department of Homeland Security 2008; Alexander 2005). However, prior research has shown that at least in some situations, it is important to distinguish between different types of crises, especially man-made versus natural ones, as they require different response strategies to effectively address unique characteristics and challenges (Alexander 2005; Palen et al. 2009). Natural crises such as hurricanes, earthquakes, and floods are often unpredictable and require immediate action to save lives and minimize damage (Sarol et al. 2020; Sarol, Dinh, and Diesner 2021). Man-made crises such as terrorist attacks or warfare often involve intentional harm or negligence and require approaches involving intelligence, law enforcement, security measures, and efforts to ensure the safety of affected populations (Chang, Diesner, and Carley 2012; Diesner and Carley 2010; Petrescu-Prahova and Butts 2005; Tierney and Trainor 2004). However, various types of crises may share common elements such as a need for rapid communication, coordination, and allocation of resources to address immediate and long-term impacts effectively.

Prior literature in crisis informatics highlights the need for understanding the distinct characteristics and dynamics of different types of crises, including their causes, impact, and response requirements (Olteanu, Vieweg, and Castillo 2015; Sarol et al. 2020). Detecting needs related to a crisis, a task driven by advances in natural language processing (NLP), has shown to be effective in reliably identifying specific challenges associated with different crisis events (e.g., Covid-19) and types (e.g., biological disaster) (Sarol et al. 2020; Basu et al. 2017; Sarol, Dinh, and Diesner 2021). In this paper, we analyze (utterances involving) mentions of needs and their text-based contextualization based on tweets, not needs that people or groups may actually have per se. Prior work on systematically detecting needs using NLP methods has used text summarization (Sarol, Dinh, and Diesner 2021; Rudra et al. 2015), word embeddings (Sarol et al. 2020; Basu et al. 2017), and classification using linguistic features (Olteanu, Vieweg, and Castillo 2015; Verma et al. 2011). For example, Sarol, Dinh, and Diesner (2021) applied extractive text summarization with term-frequency and keyword-in-context scores as word-level features to crisis texts, and extracted the most salient sentences that were crisis-related. Basu and colleagues (2017) leveraged Word2vec to extract terms that were closest to the terms "needs" and "availabilities" from tweets about an earthquake, and matched need-tweets with availabilities-tweets based on shared terms. Olteanu and colleagues (2015) used tweet classification based on informativeness to reveal the needs and concerns of individuals during a crisis. Their analysis indicated that among the considered tweets, 32% contained useful information, 20% focused on emotional support, 10% were related to donations and volunteer work, 10% provided caution and advice, and 7% were related to infrastructure damages. Verma and colleagues (2011) utilized various linguistic features to classify tweets that contain situational awareness information about a crisis. They found that tweets related to situational awareness tend to contain more factual information and use a more formal tone than tweets unrelated to situational awareness.

While these and other methods can accurately identify expressions of needs for and availability of resources in various crisis events, they often require significant domain adaptation to yield reliable results (Diesner 2015; Van Holt et al. 2013; Rudra et al. 2015). Also, prior methods have been tested on single crisis events and types, often on natural crises such as earthquakes (Basu et al. 2017; Sarol, Dinh, and Diesner 2021) and floods (Verma et al. 2011; Sarkar, Roy, and Basu 2019). Furthermore, prior studies have focused on detecting needs for a specific crisis, but not comparing findings across different crisis types. Consequently, there is a need for more extensive and comparative research in needs detection that encompasses a broader spectrum of crisis events, including man-made crises, to develop adaptable approaches for identifying and responding to needs during crises.

This paper proposes a human-in-the-loop needs detection pipeline to reliably detect crisis-related needs across four crisis events for the purpose of identifying unique characteristics and resource demands associated with each crisis type . First, we extract concepts that represent "needs" and "supplies" in a word embedding space. Second, we extract who-needs-what triples to determine individual or organizational entities and the specific resources /services they need. Third, we code indicator terms of needs into domain-adjusted categories and, as part of that, domain-adapt two existing needs category schemas for natural crisis events. We demonstrate the efficacy of our method by first applying it to tweets about the Ukraine-Russia conflict. We then compare the needs detected for this conflict to needs detected for three crises of other types, namely the COVID-19 pandemic, the 2010 Haiti earthquake, and the 2022 Hurricane Ian. Through this comparative analysis, we aim to uncover the needs and response challenges unique to each crisis event to contribute to more comprehensive crisis management.

We address three research questions:

- RQ1: What needs do Twitter users express in the context of the Ukraine-Russia conflict?
- RQ2: How do these needs change over time?
- RQ3: How do needs expressed in a warfare situation compare to other types of disasters, namely natural disasters and a biological crisis?

This study makes three contributions. **First**, we show how existing need categorization schemas align with empirical needs detected by our method and in our context, and what kind of domain-adaptation these schemas need. For example, needs related to the concept "Military" and "Financial and Money" are not included in official disaster response schemas (e.g., FEMA and OCHA). We also found that additional features, such as the language and geolocation of a tweet, are useful features for analyzing needs within different communities and regions. **Second**, our longitudinal analysis reveals that needs mentions over time relate to the disaster timeline. For example, a surge of mentions of military needs (e.g., weapons, armaments) may imply an escalation of armed conflicts. **Third**, we test our method across different types of crisis events, showing that our approach achieves high accuracy in identifying needs across all four events examined. We also show how our method can help to identify needs stated across crisis contexts, such as "medical resources" and "humanitarian assistance". Our findings have implications for adopting NLP approaches to study crises using social media data and informing policy development of domain-specific disaster response guidelines.

Related Work

Extant literature in crisis informatics has shown how NLP methods can help to detect crisis-related information, e.g., about affected individuals and their locations (Olteanu, Vieweg, and Castillo 2015), infrastructure damages (Verma et al. 2011), and immediate needs expressed by vulnerable populations (Sarol, Dinh, and Diesner 2021; Purohit et al. 2014). Such information can help crisis responders in their work and to review response effectiveness. A common theme across these studies is the focus on detecting reliable information about affected individuals and/ or communities so that responders can prioritize relief efforts with respect to the urgency of a need (Sarol, Dinh, and Diesner 2021; Imran, Mitra, and Castillo 2016; Purohit et al. 2014; Basu et al. 2018). While these studies have reported insights from natural crises, limited research has examined the efficacy of these method for other types of crises. For example, a large-scale natural disaster may require resources such as medical supplies, food, and shelter (Sarol, Dinh, and Diesner 2021; Sarol et al. 2020), while a man-made disaster require safety and security related resources (Petrescu-Prahova and Butts 2005; Palen et al. 2009). To test whether needs expressed on social media differ across crisis types, we examine how the tweets about the 2022 Ukraine-Russia conflict differ from tweets about the 2010 Haiti earthquake, the COVID-19 pandemic, and the 2022 Hurricane Ian. This selection of crisis events enables us to explore a range of crisis dynamics: from geopolitical and humanitarian complexities of the Ukraine-Russia conflict, to the immediate and long-term impacts of a natural disaster (earthquake), to the prolonged effects of a global health crisis, and the localized but severe impact of a natural crisis like (hurricane). By analyzing these distinct events, we aim to determine patterns and variations in the needs expressed. Further, we expand needs detection by tracing changes in needs over time, across locations, and languages, and determine if there are certain needs that might be missing from existing disaster response guidelines (e.g., financial).

Data

We rehydrated a dataset from Chen and Ferrara (2023). Their dataset contains 620,510,853 tweet Ids collected from February 22, 2022, through February 17, 2023. The tweets were pulled based on the existence of keywords related to the Ukraine-Russia conflict. Examples of these keywords are: *Ukraine, Russia, Zelensky*. Of these tweets, approximately 70.97% or 440 million were in English, 2.18% (13.5 million) in Russian, and 1.8% (1.12 million) in Ukrainian. To preserve the distribution of tweets, we took the number of Ids per day, multiplied that number by 10 million (former monthly academic API quota), and divided the result by the total size of the dataset. We then randomly sampled that percentage of Ids from the corresponding day. We used Twarc (Summers et al. 2023) to obtain raw tweet text and available metadata based on the tweet Ids.

After removing non-English and deleted tweets, 9,642,371 tweets remained posted by 2,520,389 accounts from 2/25/2022 (the second day after the outbreak of the conflict) to 1/8/2023. In this subset of data, 115,925 unique tweets had location codes specified by the tweet authors. Among these tweets, 14,013 (12%) were from countries surrounding Ukraine (UA), including Russia (RU), Moldova (MD), Poland (PL), Romania (RO), and Hungary (HU), and 101,912 from a total 58 other countries, including the United States (US), United Kingdom (GB), and Canada (CA) as the top three countries. However, as English is not the major language in Ukraine and its surrounding countries, focusing on interpreting English tweets might introduce biases to our analysis. Thus, with the use of machine translation (Tiedemann and Thottingal 2020), we collected and translated non-English tweets. Among all of the non-English tweets, 206,488 unique tweets were in Russian and 117,594 in Ukrainian based on their language codes added by Twitter. As translation takes time, we randomly selected 50,000 tweets per language for further comparison. We also applied Botometer API (Yang, Ferrara, and Menczer 2022) to estimate the ratio of Tweets posted by bots. We found that 82% accounts were not bots (as the distribution of scores indicates a cut-off with a threshold of 0.7).

To compare needs expressed across different disaster contexts, we ran needs detection on three existing tweet datasets: COVID-19 (Sarol et al. 2020), 2010 Haiti earthquake (Sarol, Dinh, and Diesner 2021), and an original 2022 Hurricane Ian we collected. The COVID-19 dataset contains 665,667 English tweets that were posted from February 28 to May 8, 2020. These tweets that contained at least one predetermined hashtag, such as #COVID19, #COVID-19, #coronavirusoutbreak. The 2010 Haiti earthquake dataset contains 54,660 English tweets that were posted between January 12 and June 1, 2010, which aligns with the early recovery phase of the earthquake response. The tweets contained the keyword “haiti earthquake”, or hashtags #haiti or #haitiearthquake. The 2022 Hurricane Ian dataset comprises 300,000 English tweets from September 21 to October 5, 2022, covering the response and recovery phases of the hurricane. The dataset includes tweets with “Hurricane Ian” or #HurricaneIan.

Methods

Extraction of Need Terms

Word-embedding Approach We applied a word embedding approach to extract a list of needs using the seed words ”needs” and ”supplies”, following (Sarol et al. 2020; Sarol, Dinh, and Diesner 2021). This approach has four steps: 1). Detecting phrase viaAutoPhrase (Shang et al. 2019) with a threshold of 0.8. 2). Removing @RT_username, replacing full url links with ”URL”, splitting tweets into sentences, and tokenizing sentences. 3). Applying the word2vec model on the sentences from Step 2. 4). Retrieving and list-ranking the top nouns closest to the word embeddings of ”needs” and ”supplies” based on cosine similarity. We manually checked the top 500 nouns and noticed after the top 100 nouns, the accuracy of nouns decreased significantly. Thus, we chose the top 100 nouns for follow-up analysis.

For steps 3. and 4., we also tested recent state of the art pre-trained language models for creating word embeddings and calculating cosine similarities. We used three models: BERT from Google (Kenton and Toutanova 2019), TwHIN-BERT (Zhang et al. 2023), and CrisisTransformers (Lamsal, Read, and Karunasekera 2023). BERT is the first Transformers-based (Kenton and Toutanova 2019) bidirectional encoder representation model widely adopted in NLP. It is trained on the BooksCorpus (800M words) and English Wikipedia text (2,500M words) and has achieved state of the art results on various NLP tasks. TwHIN-Bert uses the same Transformer architecture as BERT (Kenton and Toutanova 2019), but with different training data from Twitter. CrisisTransformers is trained on more than 30 annotated crisis-related datasets collected from Twitter. Since our dataset is also from Twitter, using this model may improve the accuracy of need detection.

However, after extracting the top nouns, we found that BERT models do not show advantages over the Word2Vec model that we trained for this study. Table 1 compares model performance (in terms of accuracy rates) for the top (50, 100, 150, 200) nouns that we assume to represent needs. Overall, Word2Vec achieved the best average accuracy of 81.08% across the four models. CrisisTransformers achieved the best performance (94%) on the top 50 phrases indicating the needs, and BERT achieved the best performance (81%) for the top 100 phrases. This observation indicates that a smaller model that is tailored for our task can outperform general large language models.

We further applied our approach to the other three datasets and manually validated the accuracy of the top 100 need terms per dataset. For COVID-19, our approach to detecting need-related terms achieved an accuracy of 0.77; for the Haiti earthquake, the accuracy was 0.82; and for Hurricane Ian, it was 0.85. These rates suggest that our approach can still achieve decent performance on other datasets.

Need Categorization We further applied open and axial coding to categorize expressions of needs within their context of use. First, two coders independently labeled our top 100 need words identified in Ukraine-Russia dataset into categories defined by each coder, along with notes on their rationale for each (category and) assignment. Both

	word2vec	BERT	Crisis Trans-formers	Twitter-BERT
Top50	92.00%	86.00%	94.00%	28.00%
Top100	80.00%	81.00%	78.00%	22.00%
Top150	75.33%	75.33%	70.00%	26.67%
Top200	77.00%	70.00%	63.50%	25.50%
Average	81.08%	78.08%	76.38%	25.54%

Table 1: Comparison of Model Accuracy for Detecting Need Terms

coders leveraged the same two existing categorizations of needs (namely, OCHA (2022)'s seven categories of needs and FEMA (2019)'s seven community lifelines) as reference points to develop their own self-generated categories. The OCHA categories are: *Education, Protection, Food Security and Livelihoods, Shelter and Non-Food Items, Health, Water,* and *Sanitation and Hygiene (WASH)*. The FEMA categories are: *Safety and Security, Food and Water and Shelter, Health and Medical, Energy, Communications, Transportation,* and *Hazardous Materials*. These two schemas were chosen as reference points as FEMA's community lifelines have been the standard for U.S. governmental agencies to prioritize response efforts, and OCHA's needs schema was created based on events that happened (and are happening) in the Ukraine as a result of this conflict. The two coders then cross-validated their annotations and reconciled their disagreements until all annotations aligned. After independent annotations of needs categories, two coders achieved a Cohen's Kappa of 0.82 for inter-rater reliability, which indicates a near perfect agreement (0.8 - 1.0) of annotations (McHugh 2012) with an agreement of 0.91.

Extraction of Who-Needs-What

We also applied a rule-based approach to extract {who, need, what} triples (Sarol et al. 2020), where *who* represents requesters and *what* represents resources and/ or targets. We used the syntactic structure of sentences to extract these triples. Our method consists of four steps: 1). Apply spaCy's dependency parser (Honnibal and Montani 2017) to tokenize sentences. 2). Find "need" terms in sentences. 3). Extract the subject (who) and direct object (what) from sentences where the "need" term is a verb. 4). If the "need" term is a noun, extract the descendant of the need term (what), which links through a preposition and the copular verb of the need term. We use the left child of the copular verb term as the *who* term.

We saw that different from other types of disasters, i.e., natural disasters and the pandemic, warfare includes more stakeholders with different needs. We used the extracted triples to study who the stakeholders are and what needs they express. We explored other terms that also indicate needs, such as "demand" and "request".

Additionally, since generative AI solutions such as ChatGPT have recently achieved impressive performance in information extraction (Han et al. 2023), we also tested the who-needs-what extraction task in the most recent version of ChatGPT-4. Results were pulled on 11 September 2023 on a paid ChatGPT account. The prompts designed in this study consist of elements including task instruction, demonstration examples, and input text, following the prompt design in (Han et al. 2023). *Prompt: Do not show code; extract who-need-what triple information from the following text data. Each row represents a sentence. For example, for the text "We need UNESCO's help!", the triple should be "we-need-help".*

The ChatGPT's output of who-needs-what triples achieved a comparative accuracy (94%) with our rule-based approach. After inspecting 50 randomly sampled who-needs-what triples, we identified three cases that ChatGPT did not extract correctly. For example, the snippet "Hungary, please we need a total #EmbargoRussianOilandGas!" was identified as one object by ChatGPT, while our approach correctly identified only "EmbargoRussianOilandGas" as an object. This error suggests a lack in ChatGPT's ability to extract non-standard English content from Twitter. Therefore, we used our rule-based approach for the following analysis.

Results

RQ1: Need Terms and Needs Categorization

Need Categorization Based on our open-coding results, we identified the categories *Financial and Money*, *Military*, and *Generic* as relevant group labels in our data, even though these categories are not yet included in the considered predefined guidelines. Categories that we identified and that also occur in these guidelines are *Transportation* (FEMA), *Health and Medical* (FEMA), *Energy* (FEMA). We also combined FEMA's *Food, Water, Shelter* category with OCHA's *Shelter & Non-Food Items* category to create a new category called *Food, Water, Shelter, and Non-Food Items*. We conceptualize generic needs as true positives that are abstract concepts or under-specified expressions, such as *assistance*, as well as phrases that need a reference object, for example: *plans* (for what?) or *purchases* (of what?). Additionally, in this category, we included needs that did not belong to any of the categories above, such as *efforts, dependence*, and *incentives*.

We consider Military needs as words that directly relate to military purposes, e.g., weapons and ammunition. Additionally, needs that are more general but pertain to responsibilities or goals of the military, such as *defense*, were included in this group. Terms that are *air* references, such as anti-air or air defense, were also included here.

Who-Needs-What The extracted who-needs-what triples revealed that different stakeholders involved in the conflict under consideration are associated with different need terms, and highlighted the prominent role of military-related need terms. From the extracted 128,784 who-needs-what triples, we excluded sentences that missed either the Who or What part in a tweet. We validated the accuracy of the extracted triples by sampling 1000 of the extracted triples, then asked two annotators to independently annotate and verify the triples. Independent annotation achieved a Cohen's Kappa of 0.78 with an agreement of 0.94, which indicates a substantial agreement (McHugh 2012). We further randomly sampled

100 cases and analyzed the errors. The accuracy of the automated Who-Needs-What approach was 0.88. Among the twelve erroneous triples in our samples, we saw four types of errors: *negation*, mislabeling of *what*, mislabeling of *who*, and incomplete labeling. Six triples were instances of the *negation* error type, which misidentified **don't need** as **need**, for example: "Russia...don't need anything from the west". Four triples had incorrect/incomplete *what* parts, for example: mis-annotating quantity as *what* in a sentence"...Turkey need huge quantity of wheat...". One triple had an incorrect *who* part, mis-annotating people as who in the clause "...funds will go for people who need medicines in hospital...". One triple was extracted from an incomplete sentence, which led to incorrect syntactic parsing results.

Further, we analyzed the need expressions per stakeholder groups (i.e., the *who* in the extracted triple). Table 2 shows the 30 most frequent {who, need, what} triples with the proportion of that expression for the who subject. We observed different needs (if applicable) across different major stakeholder groups for the Ukraine-Russia conflict. We filtered out needs triples that included a pronoun (such as I, we, and they) in the 'who' role in order to ensure that extracted triples contain the specific needs of individuals and groups. We found that the majority of *who* entities are non-named entities or common nouns, such as *people*, *refugees*, *animals*, who also expressed a need for *help* - a generic category as well, which aligns with prior findings on the prominence of detecting unnamed entities (Diesner 2012). Non-generic *who* entities such as *Ukraine* and *Putin* were mentioned in relationship to specific resources related to warfare and situational awareness. For example, the needs for *German weapons or artillery and rockets* accounted for 7% of all the requests from *Ukraine* and 13% of all the requests from *Ukrainians*, which indicates an military weapons shortage on the Ukrainian side. Expressions of a need for land accounted for 25% of needs for *Russians* as a subject, which may imply a narrative around the motivation of Russia to escalate the conflict. *Europe* as a subject uttered a need for *gas* (25%), while *America* as a subject asked for *answers* (25%). The differences between expressions of needs for material supplies, such as *weapons*, *money*, *gas* and abstract or generic needs such as *answers*, *response*, and *support* suggests a distinct nature of this conflict where the need for justification of the conflict is increasing along with other resource needs.

Difference in Needs depending on Place and Language Tweets from neighboring countries of Russia and the Ukraine stated needs related to human necessities, while tweets from non-neighboring countries uttered needs for shelter and protection. Table 3 shows the highest cumulative number of need expressions posted in tweets from different geographical locations. Terms shared across all regions are *assistance*, *shelter*, and *service*. To analyze need expressions depending on location, we split the data into two sets based on geographical tags associated with tweets: 1. Ukraine and its neighboring countries (14,013 tweets); 2. The remaining 58 countries (101,912 tweets). We first compared the identified expressions of needs across these two sets. As the number of need expressions differs across these two sets, we only selected the top 70 need terms for a fair comparison. Among these 70 tweets, 48 were shared terms and 22 unique terms per region.

We found significant variations in the expressed needs between neighboring and non-neighboring countries of Ukraine, as evidenced by the extracted who-needs-what triples. Among the 123 who-needs-what triples from neighboring countries, **Ukraine-needs-weapons and humanitarian assistance** triples occurred most frequently (22 times out of 123 triples). These triples accounted for 61.1% of all need expressions with Ukraine as the subject. The second most mentioned need was **we-need-funds**, which accounted for 16 out of 123 triples. For the 698 who-needs-what triples from non-neighboring countries, top three who-needs-what triples detected include **we-need-help**, **Putin-needs-money**, and **I-need-ammunition**, which occured seven, five, and four times, respectively.

Analyzing need terms based on the language of the tweet, we observed that tweets in Ukrainian and Russian demonstrated a significantly higher concordance in terms of shared needs, with a correspondence rate of 90%. In contrast, the overlap between Ukrainian and English tweets, as well as between Russian and English tweets, was lower, each registering at 7.8%. Among the top 90 need expressions, only seven terms-**assistance, goods, items, kits, medicines, plans, resources**-appeared across English, Ukrainian, and Russian languages. Additionally, these seven need terms were the only ones that Russian or Ukrainian tweets shared with English tweets individually. However, tweets in Russian and Ukrainian shared 81 need expressions, i.e., a notable overlap. This observation indicates a potential disparity in the needs expressed within English and non-English tweets, suggesting that such differences may be rooted in the varying degrees of awareness about the conflict among different linguistic groups.

We identified a total of 478 instances of 'who-needs-what' triples in Ukrainian language, and 362 such instances in Russian language. **Ukraine-needs-weapon (5 out of 478)** is the most frequent expression across the Ukrainian and Russian languages. In addition to that, the most often occurring who-needs-what triple in Ukrainian is **Children-need-peace (3 out of 478)**. We also identified different need subjects from different languages: e.g., Russian tweets mentioned **USA-needs-Russia/drones**, while there was an absence of tweets pertaining to the needs of **the USA** within the Ukrainian language data.

RQ2: Change in Need Expressions over Time

We further analyzed how need expressions changed as the war progressed. We found that needs related to military, financial, and energy resources were mentioned throughout the conflict. Need mentions related to human necessities occurred most frequently at the beginning of the conflict and then decreased over time. Figure 1 illustrates the cumulative change in need expressions over time, with a notable right-skewed distribution indicating a higher number of need expressions in the early months. This could be attributed to the sudden onset of the conflict and the extensive media cov-

Who	What	Count (%)	Who	What	Count (%)
people	help	646 (20%)	Iranians	help	102 (98%)
Putin	ramp	533 (40%)	Scholz	courageous decisions	102 (61%)
Refugees	help	501 (53%)	Armies	logistics	80 (100%)
Pentagon	46 bio facilities	447 (98%)	animals	help	79 (61%)
soldiers	help	280 (57%)	photos	words	79 (100%)
Tigrayans	peace	237 (62%)	citizens	special certificates	75 (41%)
Russians	more land	220 (25%)	flowers	view	70 (100%)
pets	care	201 (93%)	Hospitals	support	69 (93%)
children	peace	182 (36%)	women	emergency contraception abortion	69 (51%)
Defenders	three things	173 (85%)	times	tough leaders	65 (93%)
Ukraine	German weapons	171 (7%)	Children	peace	61 (41%)
Americans	answers	169 (44%)	Families	shelter and emergency aid	55 (74%)
Ukrainians	artillery and rockets	149 (13%)	Europe	gas	54 (25%)
civilians	urgent help	142 (61%)	kids	special medical treatment	48 (31%)
refugees	proof	114 (39%)	family	urgent funding	45 (34%)

Table 2: Top need expressions per different stakeholder

Shared		**Neighbouring**	**Non-neighbouring**
agencies	long-term	join	offer
assistance	medical-care	ongoing	tiger
billions	mercy	direct	clothing
bn	millions	rehabilitation	woods
canada	mothers	resources	catholic
cash	options	buy	pledge
catholics	organization	food	meals
charities	organizations	medicines	businesses
checkout	package	drive	harley
click	part	goal	corps
communities	partners	lgbt	clean
community	plans	dollars	learn
contact	reconstruction	cards	scrambles
contribute	recovery	products	surplus
debt	services	customers	passes
disaster-relief	shelter	visit	aids
effort	step	campaign	call
emergency	support	goods	umcor
emergency-relief	team	operation	tents
gears	teams	kits	donor
gm	unicef	governments	multitude
group	united	match	wells
groups	weekend		
items	workers		

Table 3: Highest cumulative number of need expressions for the Ukraine versus other countries based on tweets with geolocation specified by account holders.

erage. Additionally, the rapid movement of Russian forces and the lack of resources and structure for effective response may have contributed to the high rates of need expressions during this period.

As the conflict progressed, there was a decline in the number of need expressions in July and August of 2022, followed by a resurgence in September. Notable spikes in the number of need terms in early April, early May, and late June of 2022 may correspond to major military events such as the Battle of Donbas, the Battle of Siverskyi Donets, and missile attacks in Kremenchuk. The increase in need expressions at the end of August coincides with the start of massive Ukrainian counteroffensives on August 29th. This change in trends of need expressions can serve as a potential information source for analyzing the overall development and identifying intense conflicts within a long-term, large-scale conflict.

The prevalence of need categories also changed over time, as shown in Figure 2. The trends for non-generic need expressions, such as *Transportation*, *Food Security & Livelihoods*, *Health & Medical*, and *Energy*, followed a similar pattern to the overall change, with more frequent need mentions at the beginning and during escalations of the conflict. In contrast, *Military* and *Financial and Money* related need expressions showed a more consistent and stable pattern, indicating continuously unmet or reemerging needs, possibly due to supply shortages and weapon destruction.

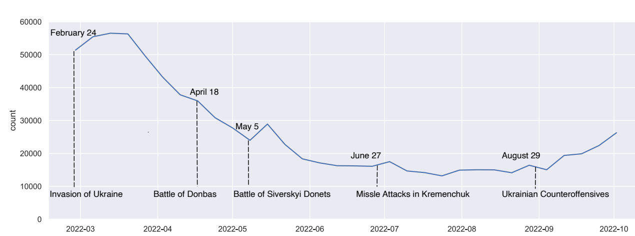

Figure 1: Change in the number of need expressions over time (3/2022 - 10/2022).

RQ3: Comparison of Need Expressions across Different Crises Contexts

Across the three crisis types (warfare, natural disasters, and biological disaster), generic need terms were dominant, followed by needs terms specific to each crisis situation. From the 300 occurrences of unique need terms across three dis-

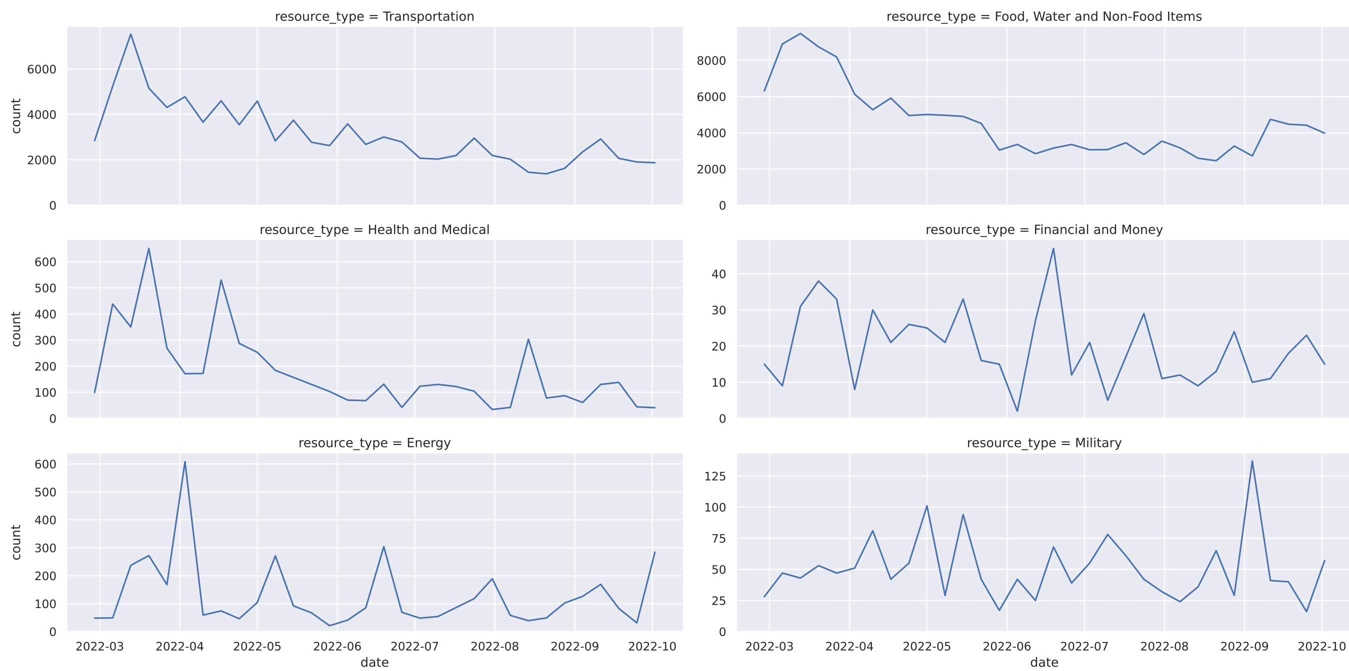

Figure 2: Change in the number of need expressions per category over time (3/2022 - 10/2022).

aster events (100 need words per category), Figure 3 shows minimal overlaps in the needs expressed per crisis. There are six terms in common across the Ukraine-Russia conflict, COVID-19 pandemic, and Haiti earthquake: *assistance*, *goods*, *kits*, *medicines*, *plans*, and *resources*. The Ukraine-Russia conflict data and the Haiti earthquake data share only one term, namely *items*. However, between the Ukraine-Russian conflict and the COVID-19 pandemic, there are 18 overlapping terms out of 100 need words per category: *ability*, *aid*, *capacity*, *deliveries*, *demands*, *efforts*, *equipment*, *essentials*, *manufacturers*, *medical-equipment*, *necessities*, *packages*, *production*, *purchases*, *shortages*, *supply*, *systems*, and *utilities*. Between the COVID-19 pandemic and the Haiti earthquake, there are two overlapping terms, namely *services* and *food*. This suggests that a conflict and a pandemic feature more similar expressions of needs compared to a conflict and a natural disaster, which has implications for the use of policies and incidence management plans in a conflict. In particular, certain logistical processes developed for rapid distribution of aid during the pandemic may be applied to conflict zones. Relatedly, similar expressions of medical needs, as well as immediate attention to vulnerable populations, indicate that preparedness plans for both crisis types should focus on medical facilities and resources for at-risk individuals and communities.

Figure 4 shows the categories of needs per event dataset (schema is discussed in *Needs Categorization* section). We include three additional categories (i.e. *other*, *generic*, and *not a resource*) for need terms that do not fit into any of the predefined categories. Terms related to "other" are valid need terms that do not fit in any of our predefined categories but are important resources for a particular crisis event. For instance, assistance programs such as *paid-sick-leave* and administrative bodies in charge of pandemic response (i.e. *local-governments* are important indicators of needs and/ or responses. "Generic" are legitimate needs-related terms, but they would need further specification, such as the addition of an object, to be able to clearly understand what resources are mentioned (though a tweet might be lacking this specification). For instance, the terms *resources*, *requests*, and *donor* signal that some support is needed or available, but they do not indicate the specific kind of need. "Not a resource" are terms that are taken out of the context of the original tweet and do not indicate a need, for example, *tastes* and *senses*. For the Ukraine-Russia conflict, we found six categories of needs mentioned that are consistent with our needs categorization schema, with the most frequently mentioned one being *transportation* (14%), and the two least mentioned ones being *Food, Water, Shelter, & Non-Food Items* (3%) and *Financial and Money* (3%). For the COVID-19 pandemic, we found seven categories of needs that fit into the predefined schema, with the most salient category being *health & medical* (25%). Least frequently-mentioned categories are *safety & security* (1%) and *Energy* (1%). For the Haiti earthquake, we observed four categories of needs, namely *health & medical* (3%), *food, water, shelter, & non-food items* (5%), *safety & security* (3%), and *financial and money* (4%). For Hurricane Ian, we identified six needs categories, including *health & medical* (2%), *food, water, shelter, & non-food items* (14%), *transportation* (7%), *safety & security* (5%), *financial and money* (6%), and *energy* (3%). Across all four events, terms are most frequently labeled as *generic* (59% for Ukraine-Russia conflict, 50% for COVID-19, 68% for Haiti earthquake, and 25% for Hurricane Ian).

Needs Categories	Ukraine-Russia Conflict		COVID-19		Haiti Earthquake		Hurricane Ian	
	Ratio	**Examples**	**Ratio**	**Examples**	**Ratio**	**Examples**	**Ratio**	**Examples**
Health & Medical	3%	medicines, medical-equipment	25%	medical-supplies, protective-gear	3%	medicines, medical-care	2%	medications
Food, Water, Shelter, & Non-Food Items	3%	grains, rations	5%	food-bankds, groceries	5%	clothing, tents	14%	meals, tenants
Transportation	14%	deliveries, shipments	2%	deliveries, distribution	0	0	7%	vehicles
Military	3%	ammunitions, armaments	2%	defense-production-act, dpa	0	0	0	0
Safety & Security	0	0	1%	childcare	3%	mothers, woods	5%	safety
Financial and Money	3%	expenditure, cheques	11%	expenses, grants	4%	charities, debt	6%	loans, funds
Energy	7%	generators, oil	1%	utilities	0%	0	3%	fuel
Other	0%	0	3%	local-governments, paid-sick-leave	10%	worldvision, partners	23%	communities, services
Too Generic	59%	requests, resources	50%	handouts, assistance	68%	mercy, donor	25%	aid, efforts
Not a Resource	8%	taster, mobilizes	0%	0	7%	weekend, glance	15%	customers

Table 4: Categorization of needs (with examples) across four different crisis events: Ukraine-Russia Conflict, COVID-19 Pandemic, Haiti Earthquake, and Hurricane Ian

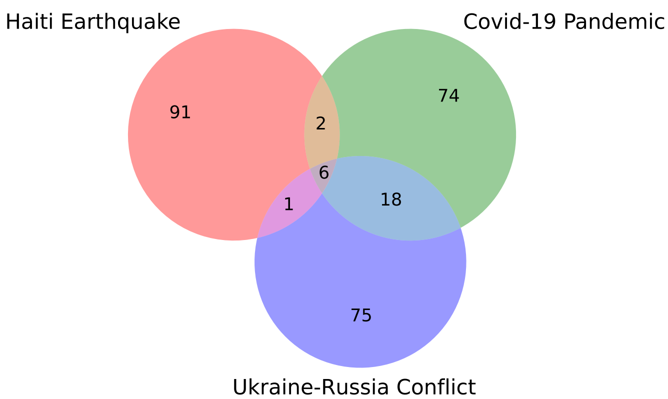

Figure 3: Venn diagram showing the overlap of need terms for the Haiti Earthquake, Covid-19 Pandemic, and Ukraine-Russia Conflict

Table 5 presents the top twenty need terms for each disaster event. These terms were selected because they are the closest to the term *need* in the word embedding space. For the Ukraine-Russia conflict, top need terms relate to supply-chain issues, e.g., *deliveries*, *shipment*, and *exports*. There are also need terms related to essential items for the vulnerable population such as *medicines*, *humanitarian-aid*, and *essentials*, though some of these terms are underspecified. For COVID-19, a majority of need terms referred to to health and medical equipment, e.g., *personal-protective-equipment* and *medicines*. There are three generic terms that may allude to the needs of essential goods for day-to-day livelihood during lockdown periods (i.e., *essentials*, *essential-items*, *stockpile*. There is also a notable proportion of terms (25%) relating to supply chain topics, such as *supply*, *distribution*, *goods*, *manufacturers*, and *manufacturing*. For the Haiti earthquake, needs were primarily related to essential items for livelihood, e.g., *clothing*, *medical-care*, *americares* (healthcare resource), *shelter*, and *woods* (for heat). Other terms related to finances, such as *debt*, *services*, *cash*, and *pledge* (which may relate to donation campaigns). There was a number of generic terms (e.g., *assistance*, *team*, *agencies*), which may suggest an involvement of response organizations, but further examination is needed to contextualize these terms. For Hurricane Ian, needs were closely related to those found in the Haiti earthquake dataset, where human necessities such as *shelter*, *food*, and *kit* were frequently mentioned. Similarly, the focus on *businesses*, *residents*, and *communities* might indicate the broader economic and social impact of the hurricane. These need terms were also frequently observed for other events, showing the importance of material support in crisis response in general. The mentioning of *efforts* and *equipment* in this table (and also evidenced in Table 4, where this need category has a ratio of 25% occurrence in the dataset), likely relates to the logistics and execution of relief operations, which is also notably present in the Ukraine-Russia conflict dataset, where operations relating to *shipments*, *supply*, and *ammunitions* are frequently occurring. A notable theme in Hurricane Ian's top needs, which is less evident other crisis events, is the emphasis on community-based activities, evidenced by the frequent use of terms like *communities*, *services*, and *programs*. This theme is also evident in Table 4, where the "communities, services" category of needs occurs most frequently (ratio = 23%).

Discussion

Our goals were to identify expressions of need related to the Ukraine-Russia conflict and to compare these expressions to those found in the context of other types of disasters, including natural and biological disasters. To this end, we applied a word-embedding approach to a corpus of tweets, and we were able to extract terms representing both generic and specific needs, such as resources (generic) vs. medical-equipment (specific). In addition, we applied a simple dependency parser-based approach to detect {Who-needs-what} triples with an accuracy of 0.88. Our results demonstrate how the text-based detection/extraction, contextualization, and classification of terms and phrases related to needs enable us to capture expressions—such as those pertaining to 'Financial and Money' and 'Military'—that are not yet included in widely used disaster response and humanitarian crisis guidelines, such as those by OCHA and FEMA. We also showed how reliably extracting 'who-need-what' triples identifies needs expressed by or referring to a broad set of stakeholders involved in a crisis, including geopolitical actors (e.g., Europe), individuals (e.g., Putin), and unnamed entities (e.g., people).

Analyzing the top need terms (Table 5), we found that need terms go beyond material items across the identified needs categories (Table 4). First, we found evidence that tangible goods are often the primary indicators of needs as they relate to human necessities such as *medicines*, *shelter*, and *food*. Second, our who-needs-what analysis showed that intangible goods are also salient, such as community rebuilding (e.g., *services*, *programs*, *residents*), justifications for certain response decisions (e.g. "Russians need more land", "Ukraine needs German weapons"), and prescriptions of crisis-related actions (e.g., "Children need peace"). These results align with Coombs (2004)'s framework of crisis communication strategies, which highlights how organizations frame needs in ways that extend beyond immediate material needs. In fact, Coombs asserted that crisis communication often involves strategic framing of needs as justifications for actions, and this theme is salient in our Ukraine-Russia conflict results. All in all, our findings underscore that the nature of needs is complex; immediate needs to help vulnerable populations are often intertwined with broader political and strategic goals.

Suggestions for Improving Needs Detection

While our needs detection framework effectivly identifies need terms and {who-needs-what} triples, there is space for improvement of extracting accurate expressions of needs. More than half of the need terms we found are generic terms that lack further specification such as a reference point or object. This finding reveals that expressions of needs on social media are primarily generic, and could benefit from further exploration of the context surrounding these terms to accurately infer their corresponding reference objects. We plan to detect both general and specific needs concurrently to provide users with the opportunity to examine both the "big picture" of a crisis situation and tangible resource needs of an affected population. For example, authorities might utilize the generic needs detection model to quickly identify and prioritize areas of need. On the other hand, if an NGO wants to raise funding and needs to decide what kinds of supplies they should provide, using a specific needs detection tool can be helpful in giving the necessary level of detail.

Furthermore, our qualitative error analysis uncovered potential challenges for developing need detection tools: we identified errors including negation, mislabeling of what and who, and incomplete labeling. Negation errors could be addressed by crafting more sophisticated rules and in-domain annotations (Wu et al. 2014). To address incomplete labeling errors, distantly-supervised NER tools could be a solution to improve the accuracy of *Who* and *What* entities extracted. Alternatively, a taxonomy with a comprehensive list of potential resource mentions might help to make needs detection more transparent and robust. For example, a taxonomy that integrates instances of need expressions from the guidelines from FEMA, OCHA, CDC, and other authorities could allow people or jurisdictions to directly match their needs with supplies from corresponding disaster response organizations. Such an approach would need to be coupled with a computational solution for term disambiguation (Kim, Kim, and Diesner 2014; Diesner, Evans, and Kim 2015) and rigorous validation.

Our application of large language models to detect need terms and who-needs-what triple showed a need for further improvement of LLMS for this specific task. BERT, Crisis-Transformers, and TwHIN-BERT all performed worse than the word2vec approach. This finding also shows that models fine-tuned on specific datasets may lack generalizability to unseen data, even when the domain of the dataset is similar. For the *who-needs-what* triple extraction task, ChatGPT achieved competitive performance to the rule-based approach. ChatGPT rarely extracted invalid triples, but may be limited in its ability to identify triples that involve platform-specific content such as Twitter hashtags (e.g., "#EmbargoRussianOilandGas"). Further work could expand the application of generative models to crisis informatics tasks (Goecks and Waytowich 2023).

Need Expressions as Indicators of Conflict Evolution?

Social media data can be an effective (supplemental) information source for monitoring and analyzing the development of armed conflicts. Though previous research has focused on using Twitter to track the evolution of natural (Gupta, Joshi, and Kumaraguru 2012) and man-made disasters (Azam et al. 2015), there is a sustained need to understand how social media data can be used to monitor massive, long-term, and international armed conflicts. Our analysis of different aspects of needs expressed related to the Ukraine-Russia conflict adds to the existing literature on social media data-mining, crisis informatics, and using NLP to model conflicts (Diesner 2015; Althaus et al. 2014). Our extracted {who, need, what} triples reflect the needs of different geo-political entities (e.g., *Ukraine*), regions (e.g., *Europe*), frontline roles (e.g., *soldiers*) affected populations (e.g., *refugees*), and specific individuals (e.g., *Putin*) and organizations (e.g., *UN*). We observed a number of need expressions that suggest Twitter users' hope for explanations and perceptions about the conflict, for example, {Americans, need, answers}. Generic need terms such as *assistance*, *help*, and *supplies* indicate that certain places or communities need multiple kinds of help. Specific need mentions such as *urgent funding* and *shelter and emergency aid* indicate the tangible resource needs of affected populations. These observation aligns with previous work (Suh et al. 2021) where social media data were analyzed to understand the public's hierarchy of needs as a crisis progresses, where the needs moved from physiological/safety to cognitive/self-actualization needs.

The change in need mentions over time might relate to the timeline of the conflict. Figure 1 shows the change in needs along with the trajectory of the conflict. For non-directly involved people (e.g., the general public from countries unaffected by the conflict), to be able to gain situational awareness of the conflict, including connecting military actions to changes in needs, might enable them to better understand opinions, supplies and demands, and the position or perspective of various stakeholders. Moreover, if NGOs or disaster response authorities want to prepare resources and raise funding, knowing about expressions of needs, of what type they are, and how those change over time can support decision-making.

Needs, as the counterpart of supplies, may imply potential supply-chain issues on a local to global scale. The expression of high demand of energy from Europe reflects the consequence of the conflict on gas/oil supplies to European countries. The increase in statements implying Military needs may suggest an increased supply-chain pressure on national defense industries (Sciutto et al. 2023). In a next step, one could integrate supply-chain data, e.g., the Purchasing Managers' Index (PMI), to estimate the discrepancy between need identified from open-source data and market supplies.

Comparing Need Mentions across Disaster Types

Our third research question asked about the differences in need expressions across three different disaster types, namely warfare (Ukraine-Russia conflict), natural disasters (Haiti earthquake, Hurrican Ian), and a biological disaster (COVID-19 pandemic). We find the highest overlap in need terms between warfare and COVID (24%), and the lowest overlap between the natural disasters and warfare (7%). The intersection of all three event types is 6%, and those need mentions are medical resources and humanitarian assistance. The top 20 needs we identified for each disaster event (in Table 5) shows a minimal overlap in top need terms across events. The prominent themes across top need terms were also different: In the Ukraine-Russia conflict dataset, they were related to logistics and possibly supply chain issues (e.g., exports, supply). In the COVID-19 dataset, they were health and medical supplies, specifically personal protective equipment and medicines. For the Haiti earthquake, they were financial-related terms, maybe indicating the potential urgency of financial aid to support response activities. In fact, Sarol, Dinh, and Diesner (2021) in their examination of the same dataset also found notable mentions of donations and charity events, possibly to help victims of the Haiti earthquake. Our categorization of needs (Table 4 revealed that a number of these need terms were valid yet *generic* or broad terms. This could be an artifact of our needs detection algorithm, which we can further refine to extract more specific needs, or reflect actual communication or needs, or be due to the fact that language use on Twitter is short and may contain noise(Sarol, Dinh, and Diesner 2021), which would make more specific needs detection more difficult.

We were surprised to not capture any need terms pertaining to *safety & security* in our Ukraine-Russia conflict data. We hypothesize that this may be because the definition of safety and security in our category schemas focuses on the protection of vulnerable groups (e.g., mothers, children, elderly (OCHA 2022)) as opposed to military-related security activities. This also means that there were no mentions of protected groups in the tweets discussing the Ukraine-Russia conflict in our dataset, which could also be a sampling issue. On the other hand, terms alluding to protected groups such as *mothers* and *childcare* were mentioned in the Haiti earthquake and COVID-19 datasets and were detected by our methodology. Overall, we find that the actual need terms and the categories they belong differ across disaster contexts. Therefore, domain adjustment of needs schemas is recommended to ensure their alignment with the specific response requirements of each disaster type.

Limitations

We recognize the importance of using domain-specific categorization schemas when detecting needs from text-data on different disaster events. The instances we analyzed differ in location, event type, and socio-economic contexts. The earthquake data showed need mentions related to search and rescue, medical assistance, and temporary shelter (Sarol, Dinh, and Diesner 2021), possibly due to the sudden destruction of infrastructure. The hurricane data led to a broader range of need phrases encompassing not only immediate relief but also long-term rebuilding and flood mitigation efforts (Dinh et al. 2022, 2023). Furthermore, the geographical and socio-economic disparities between the affected locations imply that need expressions related to a developing country like Haiti could fundamentally differ from those in the U.S.-based regions affected by Hurricane Ian, where infrastructure and response capabilities may be more advanced. We plan to integrate the unique characteristics of each disaster event into our needs detection method by analyzing the specific environmental, socio-economic, and infrastructural factors associated with each event, alongside the identified needs. Additionally, we plan to improve the reliability of our results by adapting our needs categorization schemas to suit specific domains and develop comprehensive and accurate models that can better capture the specific needs and priorities expressed in various contexts. We plan to further match needs vs. supplies by integrating multi-source data and developing statistical models to predict discrepancies in supply chains.

We are cognizant of the limitations with using Twitter for needs detection, where expressions of needs are often self-reported or strategically communicated data, which may or may not align with actual, on-ground needs of affected individuals and/or communities. Hence, in future work, we aim to employ a mixed-methods approach where we systematically compare needs found in this study with needs expressed by actual individuals on-ground during the crisis. This will involve conducting interviews with different demographic segments, including those who may lack a digital presence or who are typically underrepresented on Twitter.

Conclusion

In this paper, we showed how the detection of needs from Twitter content, including extracting a) terms that indicate needs and supplies and b) who-needs-what triples, can help to study the unfolding of the Ukraine-Russia conflict. Our findings show that people express specific and generic needs on Twitter and that those needs relate to the unfolding of the armed conflict. We also demonstrated how needs detection can benefit from considering additional features, such as the language and geolocation of the tweet, to determine specific needs in different regions and communities. Our comparison of needs extracted from tweet corpora about three different disaster types revealed that needs detection is highly domain-dependent, that an armed conflict is more similar to a pandemic than to a natural disaster in terms of what is needed, and that domain adaption of needs categorization schemas is needed to ensure the reliability of results.

Acknowledgements

We thank Richard Sowers and the Illinois Mathematics Lab at UIUC for discussing this project with us. This paper was supported in part by the Digital Humanism Fellowship program at the Institute for Human Sciences (IWM) at Vienna; this Fellowship program is supported by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology.

References

Alexander, D. 2005. Towards the development of a standard in emergency planning. *Disaster Prevention and Management: An International Journal*, 14(2): 158–175.

Althaus, S. L.; Swigger, N.; Chernykh, S.; Hendry, D. J.; Wals, S. C.; and Tiwald, C. 2014. Uplifting manhood to wonderful heights? News coverage of the human costs of military conflict from World War I to Gulf War Two. *Political Communication*, 31(2): 193–217.

Azam, N.; Abulaish, M.; Haldar, N. A.-H.; et al. 2015. Twitter data mining for events classification and analysis. In *2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI)*, 79–83. IEEE.

Basu, M.; Ghosh, K.; Das, S.; Dey, R.; Bandyopadhyay, S.; and Ghosh, S. 2017. Identifying post-disaster resource needs and availabilities from microblogs. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 427–430.

Basu, M.; Shandilya, A.; Ghosh, K.; and Ghosh, S. 2018. Automatic matching of resource needs and availabilities in microblogs for post-disaster relief. In *Companion Proceedings of the The Web Conference 2018*, 25–26.

Chang, Y.; Diesner, J.; and Carley, K. M. 2012. Toward Automated Definition Acquisition From Operations Law. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(2): 223–232.

Chen, E.; and Ferrara, E. 2023. Tweets in time of conflict: A public dataset tracking the twitter discourse on the war between Ukraine and Russia. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 1006–1013.

Coombs, W. T. 2004. Impact of past crises on current crisis communication: Insights from situational crisis communication theory. *The Journal of Business Communication (1973)*, 41(3): 265–289.

Diesner, J. 2012. *Uncovering and Managing the Impact of Methodological Choices for the Computational Construction of Socio-Technical Networks from Texts.* Ph.D. thesis, Carnegie Mellon University, USA.

Diesner, J. 2015. Words and networks: How reliable are network data constructed from text data? *Roles, trust, and reputation in social media knowledge markets: theory and methods*, 81–89.

Diesner, J.; and Carley, K. M. 2010. Relationale Methoden in der Erforschung, Ermittlung und Prävention von Kriminalität. In *Handbuch Netzwerkforschung*, 725–738. Springer.

Diesner, J.; Evans, C.; and Kim, J. 2015. Impact of entity disambiguation errors on social network properties. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 9, 81–90.

Dinh, L.; Kulkarni, S.; Yang, P.; and Diesner, J. 2022. Reliability of methods for extracting collaboration networks from crisis-related situational reports and tweets. In *2nd Information Systems for Crisis Response and Management Asia Pacific Conference, ISCRAM 2022*, 181–195. Information Systems for Crisis Response and Management, ISCRAM.

Dinh, L.; Sarol, J.; Jeoung, S.; and Diesner, J. 2023. Are we projecting gender biases to ungendered things? Differences in referring to female versus male named hurricanes in 33 years of news coverage. *Computational Communication Research*, 5(1): 141.

FEMA. 2019. Community Lifelines. https://www.fema.gov/fact-sheet/community-lifelines. Accessed: 2023-01-15.

Goecks, V. G.; and Waytowich, N. R. 2023. DisasterResponseGPT: Large Language Models for Accelerated Plan of Action Development in Disaster Response Scenarios. *arXiv preprint arXiv:2306.17271*.

Gupta, A.; Joshi, A.; and Kumaraguru, P. 2012. Identifying and characterizing user communities on twitter during crisis events. In *Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media*, 23–26.

Han, R.; Peng, T.; Yang, C.; Wang, B.; Liu, L.; and Wan, X. 2023. Is Information Extraction Solved by ChatGPT? An Analysis of Performance, Evaluation Criteria, Robustness and Errors. *arXiv preprint arXiv:2305.14450*.

Honnibal, M.; and Montani, I. 2017. spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing. Neural Machine Translation. In *Proceedings of the Association for Computational Linguistics (ACL)*, 688–697.

Imran, M.; Mitra, P.; and Castillo, C. 2016. Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, 1638–1643. Paris, France: European Language Resources Association (ELRA). ISBN 978-2-9517408-9-1.

Kenton, J. D. M.-W. C.; and Toutanova, L. K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, 2.

Kim, J.; Kim, H.; and Diesner, J. 2014. The impact of name ambiguity on properties of coauthorship networks. *Journal of Information Science Theory and Practice*, 2(2): 6–15.

Lamsal, R.; Read, M. R.; and Karunasekera, S. 2023. CrisisTransformers: Pre-trained language models and sentence encoders for crisis-related social media texts. *arXiv preprint arXiv:2309.05494*.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

OCHA. 2022. Humanitarian Needs Overview: Ukraine.

Olteanu, A.; Vieweg, S.; and Castillo, C. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 994–1009. ACM.

Palen, L.; Vieweg, S.; Liu, S. B.; and Hughes, A. L. 2009. Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event. *Social Science Computer Review*, 27(4): 467–480.

Petrescu-Prahova, M.; and Butts, C. T. 2005. Emergent coordination in the World Trade Center disaster. *Institute for mathematical behavioral sciences*, 26(3): 1–23.

Purohit, H.; Castillo, C.; Diaz, F.; Sheth, A.; and Meier, P. 2014. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday*, 19(1).

Rudra, K.; Ghosh, S.; Ganguly, N.; Goyal, P.; and Ghosh, S. 2015. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In *Proceedings of the 24th International on Conference on Information and Knowledge Management*, 583–592. 10.1145/2806416.2806485.

Sarkar, A.; Roy, S.; and Basu, M. 2019. Curating Resource Needs and Availabilities from Microblog During a Natural Disaster: A Case Study on the 2015 Chennai Floods. In *Proceedings of the India Joint International Conference on Data Science and Management of Data*, 338–341.

Sarol, M. J.; Dinh, L.; and Diesner, J. 2021. Variation in Situational Awareness Information due to Selection of Data Source, Summarization Method, and Method Implementation. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 597–608.

Sarol, M. J.; Dinh, L.; Rezapour, R.; Chin, C.-L.; Yang, P.; and Diesner, J. 2020. An Empirical Methodology for Detecting and Prioritizing Needs during Crisis Events. In *Findings of the Association for Computational Linguistics: EMNLP*, 4102–4107. 10.18653/v1/2020.findings-emnlp.366.

Sciutto, J.; Herb, J.; Lillis, K. B.; and Liebermann, O. 2023. US is running low on some weapons and ammunition to transfer to Ukraine. *CNN Politics*.

Shang, J.; Liu, J.; Jiang, M.; Ren, X.; Voss, C. R.; and Han, J. 2019. AutoPhrase: Automated Phrase Mining from Massive Text Corpora. Technical report, Retrieved 2019-08-22, from https://github. com/shangjingbo1226/AutoPhrase . . . .

Suh, J.; Horvitz, E.; White, R. W.; and Althoff, T. 2021. Population-Scale Study of Human Needs During the COVID-19 Pandemic: Analysis and Implications (WSDM'21). Association for Computing Machinery, New York, NY, USA, 4–12.

Summers, E.; Brigadir, I.; Hames, S.; van Kemenade, H.; Binkley, P.; tinafigueroa; Ruest, N.; Walmir; Chudnov, D.; Thiel, D.; Betsy; Chartier, R.; celeste; Lin, H.; Alice; Chosak, A.; Lenz, M.; McCain, R. M.; Milligan, I.; and Elwert, F. 2023. DocNow/twarc: v2.14.0.

Tiedemann, J.; and Thottingal, S. 2020. OPUS-MT — Building open translation services for the World. In *Proceedings of the 22nd Annual Conferenec of the European Association for Machine Translation (EAMT)*. Lisbon, Portugal.

Tierney, K. J.; and Trainor, J. E. 2004. Networks and resilience in the World Trade Center disaster. *Emergency Response and Recovery*.

U.S. Department of Homeland Security. 2008. National Response Framework.

Van Holt, T.; Johnson, J. C.; Carley, K. M.; Brinkley, J.; and Diesner, J. 2013. Rapid ethnographic assessment for cultural mapping. *Poetics*, 41(4): 366–383.

Verma, S.; Vieweg, S.; Corvey, W. J.; Palen, L.; Martin, J. H.; Palmer, M.; Schram, A.; and Anderson, K. M. 2011. Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In *Proceedings of the 5th International Conference on Weblogs and Social Media (ICWSM)*, 385–392. 10.1.1.371.9621.

Wu, S.; Miller, T.; Masanz, J.; Coarr, M.; Halgrim, S.; Carrell, D.; and Clark, C. 2014. Negation's not solved: generalizability versus optimizability in clinical natural language processing. *PloS one*, 9(11): e112774.

Yang, K.; Ferrara, E.; and Menczer, F. 2022. Botometer 101: Social bot practicum for computational social scientists. *CoRR*, abs/2201.01608.

Zhang, X.; Malkov, Y.; Florez, O.; Park, S.; McWilliams, B.; Han, J.; and El-Kishky, A. 2023. TwHIN-BERT: A Socially-Enriched Pre-trained Language Model for Multilingual Tweet Representations at Twitter. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5597–5607.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? Yes
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? No, instructions are already in the paper.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, computing can be completed in laptop.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, detail evaluation has been made.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No, it is not a classification task.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? No, they are public accessible assets for research community.
   (c) Did you include any new assets in the supplemental material or as a URL? No.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? No, the data was from Twitter API.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR ? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA

Ethical Statement

We used data collected from a publicly accessible online platform. While our focus is on extracting keywords from tweets and not individual users, such data may carry risks for privacy issues. To mitigate these issues and comply with the terms of service, we will not release tweet IDs for the dataset we used. We do not endorse the use of the presented needs detection methods for other purposes beyond research and experimentation. We are committed to presenting our findings objectively and without intent to support or advance any particular political viewpoints.

Appendix

Additional Table

Ukraine-Russia Conflict	COVID-19	Haiti Earthquake	Hurricane Ian
deliveries	medical-equipment	assistance	resources
shipments	equipment	groups	plans
flows	medical-supplies	long-term	donations
exports	protective-gear	plans	shelter
supply	stockpile	disaster-relief	kit
delivery	protective-equipment	catholics	benefits
aid	ppe	cash	businesses
transfers	manufacturing	agencies	assistance
suppliers	personal-protective-equipment	debt	options
alternatives	medicines	services	food
medicines	#ppe	tiger	residents
equipments	supply	clothing	efforts
humanitarian-aid	distribution	reconstruction	equipment
ammunitions	goods	woods	distribute
supplie	manufacturers	respond	communities
sales	funds	team	services
essentials	plans	pledge	programs
plans	essentials	americares	customers
provision	essential-items	medical-care	batteries
terminal	financial-relief	shelter	insecurity

Table 5: Comparisons of top 20 need terms for Ukraine-Russia Conflict, Covid-19 Pandemic, Haiti Earthquake, and Hurricane Ian