Ukrainian Resilience: A Dataset for Detection of Help-Seeking Signals Amidst the Chaos of War

†**MSVPJ Sathvik**[1,2,5] **Abhilash Dowpati**[2,3,5] **Srreyansh Sethi**[4]
[1]IIIT Dharwad, India [2]Cognicore AI, India [3]University of Delhi, India
[4]Zero True, USA [5]Raickers AI, India
{msvpjsathvik,dowpati6215,srreyansh.sethi}@gmail.com

Abstract

The "Ukrainian Resilience" dataset is a novel collection of Ukrainian language social media posts from the ongoing Russia-Ukraine war, designed to facilitate the development of natural language processing and machine learning models for detecting subtle linguistic cues indicating distress or urgent requests for aid. Comprising a binary classification of posts requiring help or not during the conflict, this dataset aims to enhance humanitarian efforts by enabling more rapid identification of genuine calls for assistance amidst vast online content. Leveraging advanced language models and algorithms to discern nuanced signals of distress, the "Ukrainian Resilience" dataset holds significant potential for real-world impact, as evidenced by preliminary experiments using GPT-3.5 achieving 81.15% accuracy.

1 Introduction

War, particularly the ongoing conflict between Ukraine and Russia, has profound and devastating effects on innocent civilians. In times of war, civilians often find themselves caught in the crossfire, facing the harsh consequences of violence, displacement, and the breakdown of essential services. The Ukraine-Russia war is no exception, as it has led to a humanitarian crisis with significant repercussions for the Ukrainian population. Innocent lives are disrupted, families are torn apart, and communities are shattered as a result of the conflict.

One of the most immediate and visible impacts is the displacement of civilians. Many Ukrainians have been forced to flee their homes in search of safety, facing the challenges of finding shelter, food, and medical care. The psychological toll on individuals, especially children, is severe, with the trauma of war leaving lasting scars on the mental well-being of those affected. The destruction of infrastructure, including schools and hospitals, further exacerbates the suffering of the innocent, limiting access to education and healthcare.

Moreover, the war has economic ramifications, with livelihoods disrupted and economic instability compounding the difficulties faced by civilians. The struggle for resources, combined with the breakdown of social structures, can lead to a desperate situation for many Ukrainians, particularly vulnerable populations such as the elderly and children.

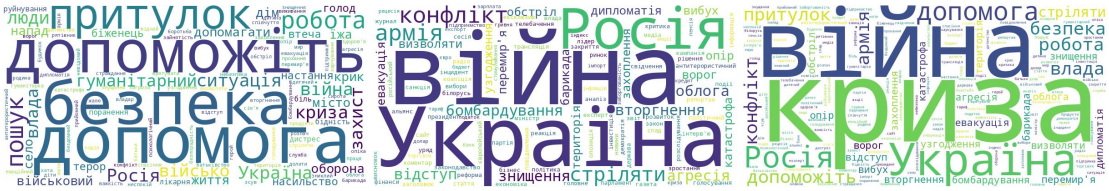
In the digital age, social media plays a crucial role in conveying the human stories behind conflicts. A dataset designed for binary classification to distinguish social media posts that signal a need for help versus those that do not can be invaluable in crisis response efforts. Such a dataset could utilize natural language processing (NLP) techniques to analyze the content of posts, identifying key indicators of distress, urgency, or specific requests for assistance.

This approach not only leverages technological advancements but also demonstrates the potential for data science to contribute to humanitarian efforts, offering a more efficient means of identifying and addressing the needs of those affected by war. It emphasizes the importance of harnessing technology to provide timely and targeted assistance to vulnerable populations in crisis situations.

Motivation: The war impacts the lives of innocents some leave their country, some are wounded, a few can become orphans, and most lose their jobs and property. In these critical situations, social media can help in addressing their problem and can request help. Detection of the help-seeking posts during the time of war could help refugees or migrants. Governments and non-governmental organizations (NGOs) can utilize and help them through social media. It might be helpful to provide jobs, shelter, food, etc.

The key contributions of this work are as follows:

Findings of the Association for Computational Linguistics: EMNLP 2024, pages 294–300



(a) Word cloud for Positive label (b) Word cloud for negative label (c)Word cloud for entire dataset

Figure 1: World clouds simulated on the annotated dataset(Ukrainian Resilience)

1. As of our knowledge, we are the first to develop a dataset that helps the innocents in the Russian-Ukraine war through social media in the Ukrainian language.

2. The proposed dataset can be translated to various other languages can be helpful for diverse languages and countries.

3. The baselines are implemented on the dataset and the highest accuracy is 81.15%.

Imran et al. (2022) proposed a method for rapid damage assessment using microblogs during disasters. The study likely focuses on the real-time analysis of social media data for assessing the extent of damage caused by disasters. Sufi and Khalil (2022) introduced an approach that combines AI-based location intelligence and sentiment analysis for automated disaster monitoring through social media posts. This work is likely to address the spatial and emotional aspects of disaster-related information.

Sufi (2022) presented an AI-based software, AI-SocialDisaster, designed for identifying and analyzing natural disasters from social media. This software might offer a comprehensive solution for extracting actionable insights from social media during disasters. Chen et al. (2022) explored the application of satellite data and AI in efficient agricultural disaster financing. This work likely focuses on using advanced technologies to optimize financial strategies in the agricultural sector during disasters. Lu et al. (2022) provided an overview of the applications of AI and machine learning in disasters and public health emergencies. This work is likely to offer insights into the broader spectrum of AI applications in disaster management.

Recent literature has made significant contributions in enhancing disaster management by leveraging social media. This progress has led to the identification of various applications that can be instrumental in aiding people during times of crisis. However, employing social media and NLP techniques to help the innocents at the war is still unexplored. The uniqueness of the work lies in developing a dataset in the Ukrainian language for helping innocents in the Russia-Ukraine war which is not done before.

Table 1: Statistics of the Dataset

Metric	label 0	label 1	Total/Overall
Data Size	5782	5895	11677
Number of Words	204859	217289	422148
Words per data point	35.43	36.86	36.15

2 Methodology

2.1 Annotation Scheme

We gathered data from two major social media platforms, Twitter and Reddit, utilizing their respective APIs to collect information in the Ukrainian language. We have extracted the posts till May 2023. The keywords used to extract the posts are Russia Ukraine war, війна, допомога, росія україна війна. To maintain a focus on textual content, we excluded images and videos from the extracted data.

The annotation process involved three master's students of the Ukrainian language. They have scored first class marks in all the subjects in their curriculum. They were provided with the CSV file which contains data point. They are asked to label in the columns. The dataset was divided among the annotators so that each data point was annotated by two annotators. The final label for each data point was determined based on their annotations. If both annotators assigned the same label, that label was accepted as the final label. If the annotations differed, the data point was flagged for discussion and subsequently labeled after a consensus was reached.

The annotators had to choose from the following

Table 2: Overview of the dataset

Text	Label[0/1]
Друзі, знаходжусь у важкій ситуації через війну. Шукаю тимчасове притулок. Якщо у вас є можливість надати допомогу або пораду, буду вдячний. Дякую за розуміння та підтримку.	1
Інформація свідчить про те, що приблизно сімдесят осіб трагічно загинули в ході триваючого конфлікту між Росією та Україною. Серед загиблих - цивільні, військовослужбовці та люди, які опинилися в поле зору вогню. Ця приголомшлива кількість свідчить про необхідність негайних дипломатичних рішень та міжнародного втручання для припинення насильства. Наші думки разом із сім'ями та спільнотами, які постраждали від цієї серцеїдної ситуації. У міру подальшого розвитку конфлікту наголошуємо на важливості об'єктивного інформування та міжнародної співпраці для вирішення кризи та досягнення тривалого миру.	0
Звертаюсь з проханням про допомогу. Останні два дні не маю доступу до їжі через складні обставини. Якщо хтось може надати допомогу з продуктами харчування, буду вдячний за вашу доброту. Дякую всім, хто розуміє та підтримує.	1
За останніми звітами, конфлікт між Росією та Україною призвів до загострення гуманітарної ситуації. Цивільні населення в зоні конфлікту стикаються з важкими викликами. Міжнародні гуманітарні організації звертаються до світу з проханням надати допомогу тим, хто постраждав. Спільно працюємо над знайомством з потребами та допомогою.	0

during annotation:

1. "Help-seeking" - If the provided text indicates seeking help during the war.

2. "Non help seeking" - If the provided text does not indicate any need neither seeking help.

In order to diversify the dataset beyond war-related content, we have added a subset of data labeled as 0. This included non-relevant information such as social media posts about actors, movies, games, etc. This strategy aims to make the dataset versatile and applicable to real-time social media scenarios, not exclusively limited to war environments.

Every data point is scanned manually twice. First by the diploma student and verified by the masters students. We assessed inter-annotator agreement using kappa(Krippendorff, 2011) as the metric. The three master's students (A, B, C) with the diploma students with the scores of $\alpha_A = 0.81$, $\alpha_B = 0.84$, and $\alpha_C = 0.83$. The overall average score is $\alpha = 0.827$.

2.2 Analysis

Table I reports the statistics of the dataset. The dataset comprises 5,782 instances labeled as 0 and 5,895 instances labeled as 1, resulting in a total of 11,677 instances. In terms of the number of words, label 0 has 204,859 words, label 1 has 217,289 words, and the overall dataset contains 422,148 words. On average, each instance in label 0 has 35.43 words, each instance in label 1 has 36.86 words, and the mean words per data point for the entire dataset is 36.15.

Table 3: Test results: Detection of help-seeking posts

Model	Precision	Recall	Accuracy
XLM RoBERTa	68.81	71.43	69.63
mBERT	67.91	69.40	68.52
XLM-ProphetNet	69.46	70.10	70.91
LLAMA 2 7B	72.58	73.84	73.81
LLAMA 2 13B	73.86	72.84	75.38
GPT 3 Ada	72.68	73.51	72.81
GPT 3 Babbage	73.41	74.58	74.68
GPT 3 Curie	76.71	74.62	76.86
GPT 3.5	80.13	81.12	**81.15**

Figure 1 illustrates three sets of word clouds: one for the positive class, one for the negative class, and one representing the overall category. Each cloud visually displays the most prominent words in its respective group. Moreover, table 2 presents the overview of dataset 1 indicating the posts where users are requesting help and 0 indicates no help seeking posts.

2.3 Baselines

We have used various Large Language models for performing the experiments on the proposed dataset. They are: GPT 3.5 (Chen et al., 2023), GPT 3 (Brown et al., 2020), Large Language Model Meta AI 2 (LLaMA 2) (Touvron et al., 2023), mBERT(Pires et al., 2019),

XLM-RoBERTa(Conneau et al., 2020) and XLM-ProphNet(Qi et al., 2020).

The GPT-3 and LLAMA models come in various versions, each tailored to specific parameters. GPT-3, for instance, offers a range of variants, including GPT-3 Ada, GPT-3 Babbage, GPT-3 Curie, and GPT-3 Davinci. Meanwhile, in LLAMA 2, we have implemented models with 7 billion parameters (7B) and 13 billion parameters (13B). All the models implemented are by finetuning.

In the course of our experiments, we divided the dataset into two distinct sets: the training set and the testing set. The training set comprises 80% of the entire dataset, while the testing set makes up the remaining 20%. It is important to note that this division was executed randomly to ensure a representative distribution of data in both sets.

As for the hyperparameters, we maintained uniformity across all models by setting the number of epochs to 5. Additionally, the remaining parameters were set as their default values.

3 Experimental Results and Discussion

In the realm of predictive modeling for identifying help-seeking posts, a comprehensive evaluation of different algorithms is implemented for classifying text.

The table 3 presents a detailed analysis of various language models' performance in the task of detecting help-seeking posts. Notably, models like RoBERTa, BERT, and DistilBERT exhibit moderate accuracy, precision, and recall values. However, the LLAMA 2 and GPT-3 series showcase superior capabilities, with GPT 3.5 emerging as the standout performer. The GPT 3.5 model demonstrates remarkable precision (80.13%), recall (81.12%), and accuracy (81.15%), indicating its effectiveness in accurately identifying help-seeking posts. The trend of improvement across the GPT-3 series suggests that advancements in language models contribute to enhanced performance in this specific application.

Error analysis: Upon closer inspection, a recurring theme among false positives emerges. Many of these misclassifications seem to stem from posts that contain emotionally charged language or expressions of distress, but are, in fact, providing updates on non-emergency situations. The model might struggle to distinguish between urgent pleas for help and general expressions of concern. Additionally, false positives seem to be more prevalent in posts that contain ambiguous or sarcastic language, as the model may struggle with nuanced linguistic cues.

Conversely, false negatives present a different set of challenges. A notable observation is that the model tends to overlook subtleties in posts where the urgency is not explicitly stated. For instance, posts that discuss longer-term issues or structural problems related to the conflict may be incorrectly classified as not needing immediate help. Furthermore, false negatives are more common in posts that use local dialects or slang, which the model might not have adequately captured during training.

Although the machine learning models make errors the model can be deployed for real-time monitoring and detection for the Russia-Ukraine war. Once deployed for real-time monitoring of social media posts the machine learning models can detect help-seeking posts in Ukrainian language during the war. The NGOs and governments can provide the shelter for people who require shelter, can provide jobs for jobless, an orphan can join an orphanage, etc.

4 Conclusion and Future Work

We present a novel dataset for the detection of help seeking posts at the time of war. Based on the conducted experiments, it has been established that the proposed dataset has significant potential to assist governments and non-governmental organizations (NGOs) in aiding innocent individuals affected by war who are in dire need of support. This assistance spans various necessities, including but not limited to employment opportunities, food, shelter, and the search for missing family members.

To extend the scope of this work, future efforts will involve the creation of datasets specifically tailored for the Israel-Palestine conflict and civil wars. Additionally, there is a recognition of the importance of developing a universal dataset in the English language. This broader dataset aims to streamline response efforts by collecting and curating social media posts in English. This approach ensures that in the event of a sudden declaration of war, organizations can promptly leverage the dataset from the onset, creating a versatile resource applicable across languages and countries.

Limitations

The proposed dataset is exclusively composed of posts in the Ukrainian language. It is essential

to note that models trained on this dataset might not perform effectively when applied to social media posts in languages other than Ukrainian, such as English. Furthermore, it is important to highlight that the dataset is specifically focused on the Russia-Ukraine war. This specialization means that utilizing the same dataset for different conflicts may not yield satisfactory results.

Another significant aspect is that the dataset exclusively comprises text, with all images removed during the collection process. Consequently, it might not be well-suited for tasks or applications that heavily rely on visual content. Additionally, upon analysis, it was observed that some posts lack textual content, yet users are clearly expressing urgent requests for assistance. In these instances, the proposed dataset proves inadequate due to its text-only nature, as it fails to capture and represent the vital information conveyed through non-textual elements.

In summary, while the proposed dataset serves its intended purpose for Ukrainian text-based content related to the Russia-Ukraine war, it may not be versatile enough for applications involving diverse languages, different conflicts, or situations where visual information is crucial.

Ethics Statement

We acknowledge the potential societal impact of our research and will strive to ensure that the application of our findings is responsible and aligns with the welfare of individuals and communities. Efforts will be made to avoid any harmful consequences arising from the use of the dataset or the developed algorithms. We commit to continuous ethical oversight throughout the project. Regular reviews will be conducted to assess and address any emerging ethical concerns. Modifications to our approach will be made promptly in response to new information or changing circumstances. By adhering to ethical principles, we aim to conduct research that is not only scientifically rigorous but also socially responsible and respectful of the rights and well-being of individuals participating in or affected by our work.

Data availability: The dataset has risk of being used for unethical purposes for making business, cold mailing etc. The dataset will be shared only for NLP researchers and developers or organisations who are working for social good.

References

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.

Shijun Chen, Huabin Lin, and Guang Yang. 2022. Efficient agricultural disaster financing using satellite data and artificial intelligence. *Computers and Electrical Engineering*, 103:108394.

Xuanting Chen, Junjie Ye, Can Zu, Nuo Xu, Rui Zheng, Minlong Peng, Jie Zhou, Tao Gui, Qi Zhang, and Xuanjing Huang. 2023. How robust is gpt-3.5 to predecessors? a comprehensive study on language understanding tasks. *arXiv preprint arXiv:2303.00293*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Muhammad Imran, Umair Qazi, Ferda Ofli, Steve Peterson, and Firoj Alam. 2022. Ai for disaster rapid damage assessment from microblogs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12517–12523.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability.

Sally Lu, Gordon A Christie, Thanh T Nguyen, Jeffrey D Freeman, and Edbert B Hsu. 2022. Applications of artificial intelligence and machine learning in disasters and public health emergencies. *Disaster medicine and public health preparedness*, 16(4):1674–1681.

Hirak Mazumdar, Chinmay Chakraborty, MSVPJ Sathvik, sabyasachi Mukhopadhyay, and Prasanta K Panigrahi. 2023. Gptfx: A novel gpt-3 based framework for mental health detection and explanations. *IEEE Journal of Biomedical and Health Informatics*, pages 1–8.

Hirak Mazumdar, MSVPJ Sathvik, Chinmay Chakraborty, Bhuvan Unhelkar, and Saïd Mahmoudi. 2024. Real-time mental health monitoring for metaverse consumers to ameliorate the negative impacts of escapism and post trauma stress disorder. *IEEE Transactions on Consumer Electronics*, 70(1):2129–2136.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, and Ming Zhou. 2020. ProphetNet: Predicting future n-gram for sequence-to-SequencePre-training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2401–2410, Online. Association for Computational Linguistics.

Sathvik. 2023. Enhancing machine learning algorithms using gpt embeddings for binary classification. *TechRxiv*.

Msvpj Sathvik, Abhilash Dowpati, and Revanth Narra. 2024. French GossipPrompts: Dataset for prevention of generating French gossip stories by LLMs. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1–7, St. Julian's, Malta. Association for Computational Linguistics.

Zuhair Shaik, Sai Kartheek Reddy Kasu, Sunil Saumya, and Shankar Biradar. 2024a. IIITDWD-zk@DravidianLangTech-2024: Leveraging the power of language models for hate speech detection in Telugu-English code-mixed text. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 134–139, St. Julian's, Malta. Association for Computational Linguistics.

Zuhair Hasan Shaik, Dhivya Prasanna, Enduri Jahnavi, Rishi Thippireddy, Vamsi Madhav, Sunil Saumya, and Shankar Biradar. 2024b. FeedForward at SemEval-2024 task 10: Trigger and sentext-height enriched emotion analysis in multi-party conversations. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 745–756, Mexico City, Mexico. Association for Computational Linguistics.

Fahim K Sufi. 2022. Ai-socialdisaster: An ai-based software for identifying and analyzing natural disasters from social media. *Software Impacts*, 13:100319.

Fahim K Sufi and Ibrahim Khalil. 2022. Automated disaster monitoring from social media posts using ai-based location intelligence and sentiment analysis. *IEEE Transactions on Computational Social Systems*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models.

A Real time scenarios

Here we list few real time scenarios and to whom this dataset can be useful for.

1. As a volunteer for the local UN organization, I can leverage social media and an algorithm trained on a specific dataset. By simply typing 'war' on Twitter, the algorithm can automatically identify and highlight individuals in my city who have been affected by war and are in need of assistance. This targeted approach allows for more effective and efficient help to be provided based on the filtered search results.

2. In my role as an HR professional, I recognize that individuals impacted by war may be willing to work for lower wages. Using the algorithm, I can filter through job seekers to identify those who have been affected by war and are actively seeking employment opportunities.

3. As the CEO of a company or NGO, if I wish to contribute through donations of clothing, money, or food, the algorithm becomes a valuable tool. By utilizing the algorithm, I can obtain detailed statistics about the needs in my locality or any specific area where I intend to make donations, ensuring that the assistance is targeted and impactful.

4. As an individual with surplus food or clothing available for donation, the proposed algorithm provides a user-friendly solution. By using the algorithm, I can efficiently filter and identify those in need of clothing or food, facilitating a more direct and purposeful contribution to those who require assistance.

B Novel algorithms: Scope for improving accuracy

There is significant scope for developing novel algorithms to enhance the accuracy of models as done in our previous works (Mazumdar et al., 2023; Sathvik, 2023), especially in low-resource language settings through rigorous experimentation(Shaik et al., 2024a,b). As computational models continue to evolve, traditional approaches often face challenges in capturing the complexities of tasks and domains with limited data(Sathvik et al., 2024; Mazumdar et al., 2024). Therefore, innovative algorithmic solutions are essential to address these gaps.

In our previous work, we proposed a new algorithm designed to improve the accuracy of the detection(Mazumdar et al., 2023). Specifically, we demonstrated its efficacy by applying it to a dataset we developed to track help-seeking signals related to the ongoing conflict in Ukraine. This dataset was built to aid in the identification of individuals seeking help during wartime, a particularly challenging task given the limited availability of high-quality training data for such a specialized context.

C Real time deployments and applications of the classifier

Here we list few real time deployments and applications that are possible.

1. **AI-driven Social Media Crawlers:** Utilizing AI web crawlers enables the continuous monitoring and filtration of help-seeking posts, presenting opportunities for diverse applications outlined in Section C.

2. **Raw Data-based Detection:** Initially, data is gathered from social media platforms using API keys with specific keywords, followed by the application of a classifier to filter out help-seeking posts.

3. **Real-time Web Application:** A web application is designed to monitor real-time posts, showcasing help-seeking content from both Twitter and Reddit. This valuable information can be employed by NGOs, government entities, and individuals to provide assistance to those affected by conflicts.