Towards XAI for Information Extraction on Online Media Data for Disaster Risk Management

Jane Arleth dela Cruz*
Centre for Language Studies,
Centre for Language and Speech Technology,
Radboud University, Netherlands
janearleth.delacruz@ru.nl

Iris Hendrickx
Centre for Language Studies,
Centre for Language and Speech Technology,
Radboud University, Netherlands
iris.hendrickx@ru.nl

Martha Larson
Centre for Language Studies,
Inst. for Computing and Information Sciences,
Radboud University, Netherlands
martha.larson@ru.nl

ABSTRACT

Disaster risk management practitioners have the responsibility to make decisions at every phase of the disaster risk management cycle: mitigation, preparedness, response and recovery. The decisions they make affect human life. In this paper, we consider the current state of the use of AI in information extraction (IE) for disaster risk management (DRM), which makes it possible to leverage disaster information in social media. We consolidate the challenges and concerns of using AI for DRM into three main areas: limitations of DRM data, limitations of AI modeling and DRM domain-specific concerns, i.e., bias, privacy and security, transparency and accountability, and hype and inflated expectations. Then, we present a systematic discussion of how explainable AI (XAI) can address the challenges and concerns of using AI for IE in DRM.

Keywords

Disaster Risk Management, Information Extraction, Explainable AI (XAI), Explainability

INTRODUCTION

Disaster risk management (DRM) is a cycle whose aim is to help affected communities mitigate, prepare for, respond to and recover from disasters. DRM research and practice are always relevant with the ever-present threat of natural hazards around the globe, a recent major one being the Turkey-Syria earthquakes on the 6th of February 2023 – which the World Health Organization's director for Europe described as the "worst natural disaster" in the EU region in the century (Rasheed and Pietromarchi 2023).

DRM is implemented by DRM practitioners, individuals whose professional function includes contributing to the management of disaster risk within a sector, geographic area or organization (Davis et al. 2013). The emergence of big data has potential in being used for AI solutions in DRM in diverse areas ranging from creating more accurate risk models to optimizing relief logistics (Sun et al. 2020; Moitra et al. 2022). In many fields, such as medicine, finance and retail, AI, and specifically Deep Learning Models, has gained traction due to advances in algorithms, a growth in computational power, and the availability of large data sets (Kuglitsch et al. 2022). Within DRM, it is hoped that such technologies can also be utilized, capitalizing on big data, to strengthen our understanding of natural hazards, the timeliness of detection, and the effectiveness of emergency communications.

*corresponding author

To facilitate informed DRM in practice, many AI-based decision support tools have been developed by research institutes and industrial companies in the past years with the focus on tools in the disaster response phase (Sun et al. 2020). A small portion of tools use sensor measurements, remote sensing data, or mobile phone data as input, but many tools use social media (Sun et al. 2020).

Information Extraction (IE) is the process of transforming unstructured text into information that is more readily digested and analyzed (Cowie and Lehnert 1996). The IE task is to extract named entities such as names of people and places, relations of entities such as who lives in what location, or other “knowledge nuggets” from text (Zhai and Massung 2016). IE is a crucial task in DRM with the need for relevant information in a timely manner. IE has been applied for casualty information extraction from news (Chaulagain et al. 2019) and automated disaster-related information extraction from tweets (Zahra et al. 2022; St Denis et al. 2020). Most AI-based IE applications for social media data are based on supervised machine learning (ML) methods which can automatically extract named entities by learning from manually labeled examples (Li et al. 2017) and when we refer to AI-based tools we refer to this type of supervised ML methods.

In this paper, we will consolidate the challenges and concerns of introducing AI into DRM and discuss how they can be addressed. Our focus is on explainable AI (XAI) for IE in DRM. Our work is inspired by the recognition that understanding and addressing AI’s limitations is required to fully realize its benefits (Kuglitsch et al. 2022). We identified the challenges and concerns as belonging to three areas: (1) limitations of DRM data, (2) limitations of AI modeling, and (3) DRM domain-specific concerns. We briefly explain the DRM domain-specific concerns in more detail here.

Although social media platforms have created opportunities for people to share relevant information about disasters that can be useful for DRM, we face *limitations of DRM data* from these sources (Imran, Ofli, et al. 2020; Li et al. 2017). There has been a trend of people reporting critical information on social media platforms during disasters. Such information is about injured people and casualties, survivors, evacuations, early warning, cautions, damage to infrastructure, calls for volunteer and donations, and calls for sympathy and emotional support (Imran, Ofli, et al. 2020; Imran, Mitra, et al. 2016). However, Adrot et al.’s (2022) findings suggest that research on social media data usage primarily focuses on technological opportunities and affordances and, hence, lacks practical implementation aspects in organizations. A limitation with social media datasets is that they capture a specific time period, typically when there is an influx of tweets or the use of trending hashtags, making it difficult to understand both the causes of the disaster and the entire period of aftermath where the disaster impact is felt (Crawford and Finn 2015). Another challenge with these social media datasets is that they can have different hazard-specific entities (Li et al. 2017). These limitations pose a challenge for natural language processing (NLP) techniques in the parsing, analysis, and extraction of information useful for disaster response (Wagenaar et al. 2020).

AI-based tools can be used to detect or forecast events by combining multiple data sources or modeling techniques, but there are still *limitations in AI modeling*. For instance, it is important to know how to evaluate AI-based tools and what level of explainability is required (Kuglitsch et al. 2022). The ‘black-box’ nature of currently popular Deep Learning models makes it harder to explain the outcomes of the model, which in turn will make it harder to attest decisions made on these AI models outcomes. With reduced transparency and explainability of Deep Learning models comes increased challenges in ensuring that DR experts and decision-makers are accountable for their communities (Gevaert et al. 2021).

DRM-specific concerns about introducing AI span a broad range. The studies of Gevaert et al. (2021) and Moitra et al. (2022) highlight the following: (1) bias, (2) privacy and security, (3) transparency and accountability, and (4) hype and inflated expectations. Bias is a significant risk, not only due to the data attributes used to train the AI models but also because the models generalize in such manner that minorities are often not well represented by the model. An example are DRM training datasets which when they are available, they are often for wealthier countries and cities; when models are trained on such data they may fail to be transferable in the context of rural and less wealthy areas (Moitra et al. 2022). Although most DRM data is often not directly personal in nature (e.g. high-resolution drone images, buildings or other spatial units), people’s homes and communities are still represented in these data and therefore privacy and security concerns are still relevant (Moitra et al. 2022). Moitra et al.’s study revealed concerns among DRM practitioners over the lack of accountability in the current practices in DRM and felt that the introduction of AI may further exacerbate this situation. With the associated complexity introduced by AI models in DRM modeling practices, concerns for transparency and accountability arise, which lead to a reduced sense of ownership of the AI models and difficulty in motivating decisions from these models (Soden, Wagenaar, and Tijssen 2021). Hype surrounding AI also created major challenges in DRM projects. Interviewed practitioners in Moitra et al.’s study referred to pressure placed on them by their organization who seek to establish themselves as leading experts and publicize innovativeness of their AI capabilities while often overlooking limitations and risk.

The paper is structured as follows. First, we discuss the current state of the art of AI for DRM and provide more details about the three areas of challenges and concerns. Second, we introduce XAI for decision support and discuss other domains in which it is being developed and used. Then, we bridge the gap between XAI and current state of AI solutions for IE for DRM. We discuss which concerns and challenges are most likely to be addressed by XAI. Finally, we provide an outlook on the way forward.

AI FOR INFORMATION EXTRACTION FOR DRM

Current State of AI for DRM

AI models support informed DRM with their capacity for quicker identification of optimal responses (Yu et al. 2018) and cost reductions in implementation (Simões-Marques and Figueira 2019). A study by Imran, Ofli, et al. (2020) showed that AI tools, specifically, tools for text analysis being employed in DRM, are used for multiple purposes: crisis event detection, understanding public reaction, eyewitness identification, situational awareness, crisis communication, actionable information gathering, and information veracity. These AI applications are not used to substitute human expertise, but rather used to complement or support in the existing practices within DRM.

AI applications have been utilized across the four phases of the DRM cycle. For the first phase of the cycle, disaster mitigation, there have been studies applying AI methods to develop susceptibility maps for different types of hazards, to estimate possible impacts and assess community vulnerability, and comparing mitigation strategies (Sun et al. 2020). For disaster preparedness, AI methods can serve as alternative cost-effective early-warning solutions using sensor data from the field (including social media data), not only for pre-positioning resources needed but also for evacuation support systems (Sun et al. 2020). McCreadie et al. (2016) developed Emergency Analysis Identification and Management System (EAIMS), a prototype service that provides real-time detection of emergency events, related information finding and credibility analysis tools for use over social media during emergencies leveraging on ML over data gathered from past emergencies and disasters for the purposes of enhancing the decision-making processes of emergency response agencies. The disaster response and recovery phases are where AI techniques for information extraction from the field is most useful (Sun et al. 2020). For disaster response, event maps and damage information generated from different AI methods, can provide vital information from planning search and rescue operations, to distributing resources and understanding survivors' needs (Sun et al. 2020). Ramchurn et al. (2016) developed a novel disaster response system called HAC-ER, which utilizes crowd-sourcing combined with ML to obtain most important situational awareness from large streams of reports posted by members of the public and trusted organizations. For disaster recovery, AI methods can help eliminate human efforts on assessment of disaster-induced impact to infrastructure and economy, and to track disaster recovery (Sun et al. 2020).

Next, we turn to discuss social media. A study by Sun et al. (2020) presented 27 AI-based tools for DRM of which nine made use of social media data as input. Most of the tools are used for the disaster response and preparedness phases. There are various issues that pose challenges for NLP techniques to parse, analyze, and extract useful information for disaster response and preparedness from the less formal language structure to the sheer amount of posts often delivered in a short time span during these disaster events (Imran, Ofli, et al. 2020). Burel et al. (2017)'s results show that deep learning methods with semantics are able to successfully identify the existence of crisis events, and crisis event types (hurricane, floods, etc.) accurately from tweets. However, the performance of the model drops significantly when identifying fine-grained event-related information (affected individuals, damaged infrastructures, etc.) from tweets. Kruspe et al. (2019)'s study presented an approach that facilitates the detection of tweets pertaining to a specific crisis event on the basis of as few as ten example tweets against which new example tweets are compared. In Kejriwal and Zhou (2020)'s study they present a robust, low-supervision social media urgency system that adapts to arbitrary crises by utilizing a simple and effective transfer learning methodology.

Concerns about AI in DRM

In this subsection, we elaborate on the four identified DRM-specific concerns about AI namely bias, privacy and security, transparency and accountability, and hype and inflated expectations. As with many domains in which AI-based tools have been applied, there have been concerns about unintended negative consequences. These are well documented in various domains such as healthcare, criminal justice and finance (Mayer et al. 2020; Soden, Wagenaar, Luo, et al. 2019). Hence, there is also a need for critical and thorough evaluations of risk in using these tools for disaster (Moitra et al. 2022; Wagenaar et al. 2020). Moitra et al. (2022) organized a working group of 35 experts and practitioners who were involved in projects incorporating AI within the field of DRM where they discussed ethical concerns with the development and implementation of these AI applications. In the context of DRM, the more vulnerable to natural hazards are those who are socioeconomically disadvantaged who tend to live in hazard-prone areas (Abeygunawardena et al. 2009). Therefore, it is critical to make sure these AI applications supporting DRM should not amplify the risk of the already vulnerable population (Moitra et al. 2022).

Interviewed DRM Practitioners from Moitra et al. (2022) noted that where training datasets do exist for AI models, they are often for wealthier countries and cities. These biases may extend beyond the training sets but also in the use of property values in cost-benefit analyses where costs may be directed to dam and flood gate construction – still favoring wealthier areas and increasing social inequities, even before AI models are introduced (Moitra et al. 2022). The DRM practitioners also raised the concern of aggregation bias, which can stem from using population data within which information about minority groups e.g. women, differently abled persons are missing (Moitra et al. 2022). Aggregation bias occurs if aggregated regional-scale data (level of city or province) are used in place of small group-scale data (level of family or individual). Bias also exists in the amount of attention given to disaster-stricken areas where posts made by emerging influential contributors (EICs) on social media platforms such as meteorologists, news reporters, celebrities journalists etc. influence how much attention are given to such areas (Zhang et al. 2021).

DRM projects would mostly rely on proxy permission from local authorities, thus, reducing agency for the people impacted by these activities as they are rarely consulted with (Moitra et al. 2022). In addition, privacy and security concerns are highly dependent on context or culture. An example was having some residents considering data about their houses as not private but data about the rubbish in their backyard as private (Moitra et al. 2022).

The Moitra et al. (2022) study has revealed concerns on lack of accountability current DRM practices and raised that introduction of AI would intensify this. In the study of Soden, Wagenaar, and Tijssen (2021), one of the concerns for AI in DRM is the lack of transparency and explainability. One problem raised in Soden, Wagenaar, and Tijssen (2021) is the reduced feeling of ownership of ML systems that are 'black-box' in nature, as they argued that it is difficult to feed domain expert judgment into these ML systems, instead, expert judgments can only be used for validation of the model outcomes and improving the training data of the model. Many AI systems do not adequately communicate their methods or degrees of uncertainty, which increases the chance of misuse (Soden, Wagenaar, Luo, et al. 2019). This restriction leads to both expert and non-expert stakeholders feeling further removed from the actual modeling process.

Hype is an inappropriate amount of publicity and/or disproportionate amount of high expectations for the benefits of emerging technologies (Soden, Wagenaar, and Tijssen 2021). Hype surrounding AI created major challenges for ensuring these technologies were being introduced in DRM projects in ways that were safe and responsible (Moitra et al. 2022). Unrealistic expectations of AI in DRM may not only direct funding to projects that over promise, but may also lead to harm when AI systems are prematurely deployed to undertake tasks they were not fully tested for (Soden, Wagenaar, and Tijssen 2021).

EXPLAINABLE AI FOR DECISION SUPPORT

The goal of XAI is to help researchers, developers, domain experts, and users to better understand the inner operation of ML models, while preserving their high performance and accuracy, making these models more accountable and transparent to humans (Adadi and Berrada 2018; Cambria et al. 2023). XAI emerged as a research field as a response to issues of AI that produces useful output, but whose inner-workings cannot be understood by the people who develop and use it. According to a study on XAI by Adadi and Berrada (2018), there are at least four overarching motivations for explainability: explain to justify, explain to control, explain to improve, and explain to discover. XAI is essential if users and decision-makers are to understand, appropriate trust and reliance, and effectively manage AI model's results (Cambria et al. 2023; Adadi and Berrada 2018).

For real-world use cases, it is important to determine when explanations are useful and needed. Interestingly, the study by Adadi and Berrada (2018) argues that explainability, although essential, is not always a necessity. Requiring all AI systems to explain every decision could result to having compute-intensive, less efficient systems and undoubtedly expensive. Explanations can serve end users, but can also help developers evaluate the model itself. This intuition is supported by the fact that jointly learning to both predict and explain, results in more robust classifiers that are less susceptible to weaknesses such as reliance on spurious correlations (Rajagopal et al. 2020; Atanasova et al. 2020; Lei et al. 2016).

The need for explainability depends on two factors: the degree of functional opacity caused by the complexity of the AI model and the degree of resistance of the application domain to errors (Adadi and Berrada 2018). An AI system is considered opaque to the user to the extent that the user is not in a position to grasp the causal explanations of its outcomes (Vaassen 2022). Any domain where the cost of making wrong prediction is very high present a potential application domain of XAI approaches (Adadi and Berrada 2018).

XAI has a potential to improve decision support in a wide range of domains, including transportation, healthcare, legal, finance and the military, mentioned by Adadi and Berrada (2018). Here, we discuss the medical domain in more depth because the use of XAI is relatively more advanced, and it has parallels with DRM because of the

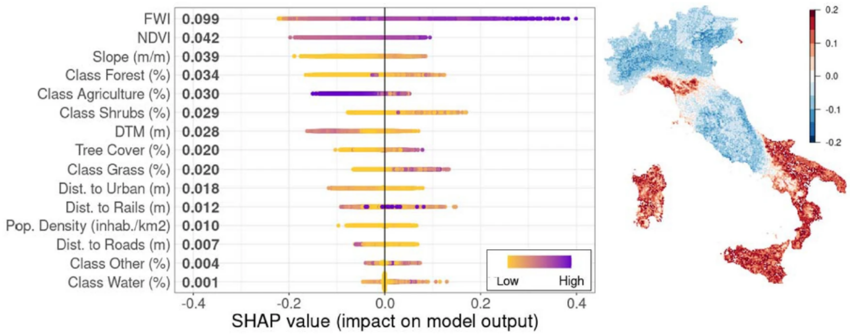

Figure 1. Cilli et al. 2022's Shapley values for wildfire occurrence prediction

life-critical nature of the decisions involved. Previous survey studies on XAI in healthcare can be found such as Tjoa and Guan (2021)'s survey for medical XAI and Payrovnaziri et al. (2020)'s survey in XAI for electronic health care data. The parallels are related to the field's factors of risk and responsibilities, where when medical responses are made lives may be at stake (Tjoa and Guan 2021). In Tonekaboni et al. (2019), the researchers were able to identify classes of explanations that clinicians identified as most relevant and crucial for effective translation to clinical practice. The medical field and DRM have an overlap in this case, where leaving important decisions to machines could turn catastrophic when exploited with malicious intent. However, in the medical field, there are a few more works dedicated to exploring explainability (Tjoa and Guan 2021; Tonekaboni et al. 2019; Payrovnaziri et al. 2020) and this ought to be translated in the DRM field where risk is also very high.

BRIDGING THE GAP XAI FOR IE FOR DISASTER RISK MANAGEMENT

We have discussed the current state of AI for DRM with emphasis on information extraction from social data media and have introduced XAI for decision support. In this section, we turn to consider what needs to be done to bridge the gap between how DRM uses AI and what XAI has to offer. In other words, we take a look at how XAI for IE can be made useful to DRM practitioners. The literature on AI for DRM has already begun to call for the development of XAI techniques. Various studies have raised the need for models to be transparent and explainable (Sun et al. 2020; Wagenaar et al. 2020; Moitra et al. 2022).

Concerns about fairness, accountability and transparency in AI systems have been thoroughly investigated by social and data scientists in domains such as medicine and criminal sciences but not yet in the field of DRM (Wagenaar et al. 2020). AI systems for DRM make key decisions such as relief distribution and evacuation strategies, which are critical to people's lives. DRM Practitioner interviewees from the study of Moitra et al. (2022) felt it was important for justification or explanation to be provided for the predictions made by a model.

A study by Cilli et al. (2022) showcases a first attempt to provide an XAI framework for estimating wildfire occurrence using a Random Forest model with SHAPley values for interpretation by Italian forest managers. This study made use of a different input dataset that included twelve biophysical exploratory variables: seven land cover classes such as agriculture, forest, grass, shrubs, wetland, water, other lands, one tree cover density, the Normalized Difference Vegetation Index (NDVI), slope and elevation (DTM), and four human-related variables: distance from a gridpoint of a human settlement, from a road and from rails, and population density and one climate Fire Weather Index (FWI) (Cilli et al. 2022). An example explanation from the paper is found in Figure 1 where on the right side a heat map of the wild fire prediction threat level in Italy is depicted and the left side lists the impact of various biophysical features such as NDVI and human-related features like distance to roads or urbanization level. The features are sorted by decreasing importance while the x-axis shows the SHAP value distribution and denotes whether a variable contributes to reduce or to increase the fire probability (Cilli et al. 2022). The SHAPley values indicate that the feature FWI (Forest Fire Weather Index) has the highest impact on predicting wild fires compared to other factors, which is helpful knowledge for forest managers. The possibility to identify wildfire driving factors at local scale can support the adoption of environmental prevention and conservation policies (Cilli et al. 2022). The authors argued that in contexts where the human and natural spheres constantly intermingle and interact, the XAI framework, suitably integrated with decision support systems, could assist forest managers to prevent and mitigate future disasters and develop strategies for effective fire management, response, recovery, and resilience (Cilli et al. 2022). Although this example is not about IE, we included it because it makes more tangible how explanations can be integrated into real-world DRM decision-making systems.

Table 1. Bridging the gap: concerns and challenges and how likely XAI can address them

Concerns and Challenges		Can XAI address them?
Limitations on DRM Data		Potentially Likely
Limitations on AI Modelling		Very Likely
DRM Domain-Specific Concerns:		
	Transparency and Accountability	Very Likely
	Hype and Inflated Expectations	Very Likely
	Bias	Potentially Likely
	Privacy and Security	Potentially Likely

To understand what makes a model explainable for DRM practitioners particularly decision-makers, we take inspiration in studies about decision-makers in the medical field. Implemented models ought to provide decision-makers context within which the model operates and to promote awareness of when the model may fall short (Tonekaboni et al. 2019). Unlike the medical field, there has been, to our knowledge, no study that has taken into account DRM practitioners' insights on what makes DRM models explainable. Healthcare clinicians viewed explainability as a means of justifying their clinical decision-making in the context of the model's prediction (Tonekaboni et al. 2019). According to Tonekaboni et al.'s research, clinicians think that models that fall short in accuracy were deemed acceptable so long as there is clarity around why the model underperforms.

We draw some general insights on where XAI can bridge the gap and address the concerns and challenges of introducing AI to the DRM practice. We distinguish the concerns from the very likely to be addressed vs. the least likely or the ones that need more future work summarized in Table 1.

We argue that XAI is very likely to address the limitations of AI Modelling, transparency and accountability, and AI hype and inflated expectations (the second to fourth rows in the table). On AI Modelling, XAI can be a powerful tool for justifying decisions, i.e., for helping to verify predictions, for improving models, and for gaining new insights into the problem at hand (Adadi and Berrada 2018). For transparency and accountability, decision-makers may prefer to use models that can be easily explained by to avoid public distrust and concerns (Soden, Wagenaar, and Tijssen 2021) where XAI can motivate decisions not only for decision-makers but also in turn for the public. With regard to the AI hype and inflated expectations, XAI can demystify the models and adjust expectations to more realistic ones.

From our perspective, the limitations of the DRM data, and the concerns of bias, and privacy and security and cannot be fully addressed by XAI. These concerns need the help of beyond XAI strategies such as understanding local context and advocating for interdisciplinary and multi-stakeholder teams. Participants from the study of Moitra et al. (2022) stressed the importance of understanding the local context in which AI tools were to be applied to help determine what data to collect to avoid bias and help identify privacy problems. Furthermore, it is important to involve relevant stakeholders at different stages of AI in DRM project and ensuring diverse team with adequate representation of people (Moitra et al. 2022). Certain aspects of privacy and security issues although not being directly solved, are being addressed via XAI's justifications' compliance with the General Data Protection Regulation (GDPR) "right to explanation" (European Commission 2016). For the DRM data, XAI can address the possibility of misused spurious relationships found in the data where XAI reveals which data the decisions are based on so the decision-makers can judge for themselves if the data deserve trust.

THE WAY FORWARD

In the previous sections, we presented the current state of AI for DRM, its concerns and challenges, then we presented our arguments for XAI to address these challenges. In this section, we present some priorities moving forward for the adoption of XAI in real-world DRM applications.

As we look towards XAI for IE for DRM, we identify that explanations ought to be evaluated against specific evaluation procedures based on the needs of DRM practioners. There are still no clear guidelines or standards that exist to support researchers/developers and those evaluating or implementing AI-based tools for DRM (e.g., policy-makers/governments, individuals/consumers, and humanitarian organizations) (Kuglitsch et al. 2022).

We argue for the development and standardization of evaluation procedures for explanations of XAI models for DRM. Inspiration can be mined from Tonekaboni et al.'s study, where they identified three different evaluation metrics: domain appropriate representation, potential actionability and consistency. Domain appropriate representation refers to how the explanation is evaluated in terms of whether the representation is coherent with respect to the application task – evaluation metrics for explanations in the disaster informativeness extraction model would be different from that of a disaster location extraction model. This also translates to evaluation metrics of explanations across different phases of the DRM cycle where evaluation metrics for disaster preparation are distinct from that of disaster recovery.

Potential actionability refers to how the explanation is evaluated based on how it informs a follow-up on the workflow or essentially the model's usability. In an applied field, explanations that are informative but have no impact on the workflow are of less importance (Tonekaboni et al. 2019). Similarly, explanations should be timely with the nature of the domain.

Consistency refers to how the set of explanations should be injective, i.e., changes in model predictions should lead to discernible changes in the explanation and that these changes should be invariant to the underlying model design variations (Tonekaboni et al. 2019). Explanations that are inconsistent violate their reliable actionability and negatively impact trust of the users (Tonekaboni et al. 2019).

Tonekaboni et al. provided metrics that may be applied in XAI for IE in DRM but ultimately the perspectives of DRM practitioners should be gathered through a future survey. We agree with Moitra et al.'s study, which presented opportunities for mitigation of AI adoption concerns in DRM including: (1) understanding of local context and (2) advocating for interdisciplinary and multi-stakeholder collaboration. Future work on XAI should be grounded in these two strategies to address concerns about AI in DRM.

The way forward for XAI for DRM is to survey DRM practitioners who use AI technology on the effectiveness of XAI from the user perspective, similar to what was done in the medical domain (Tonekaboni et al. 2019). We will survey DRM practitioners with possible XAI scenarios for IE on social media data. We want to collect meaningful insights from the practitioners about what explainability means to them and the usefulness of XAI in practice. These insights will be used to design prototype DRM systems with XAI components that are robust against the previously mentioned AI concerns and to create evaluation metrics for these systems.

ACKNOWLEDGMENTS

This work is partly financed by the Dutch Research Council (NWO) with project number NWA.1292.19.399.

REFERENCES

Abeygunawardena, P., Vyas, Y., Knill, P., Foy, T., Harrold, M., Steele, P., Tanner, T., Hirsch, D., Oosterman, M., Rooimans, J., et al. (2009). *Poverty and Climate Change: Reducing the Vulnerability of the Poor through Adaptation*. Washington, D.C. : World Bank Group.

Adadi, A. and Berrada, M. (2018). "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)". In: *IEEE access* 6, pp. 52138–52160.

Adrot, A., Auclair, S., Coche, J., Fertier, A., Gracianne, C., and Montarnal, A. (2022). "Using Social Media Data in Emergency Management: A Proposal for a Socio-Technical Framework and a Systematic Literature Review". In: *ISCRAM 2022-19th International Conference on Information Systems for Crisis Response and Management*, pp. 470–479.

Atanasova, P., Simonsen, J. G., Lioma, C., and Augenstein, I. (2020). "A Diagnostic Study of Explainability Techniques for Text Classification". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3256–3274.

Burel, G., Saif, H., Fernandez, M., and Alani, H. (2017). "On Semantics and Deep Learning for Event Detection in Crisis Situations". In: *Workshop on Semantic Deep Learning (SemDeep), at ESWC 2017*.

Cambria, E., Malandri, L., Mercorio, F., Mezzanzanica, M., and Nobani, N. (2023). "A survey on XAI and natural language explanations". In: *Information Processing Management* 60.1, p. 103111.

Chaulagain, B., Shakya, A., Bhatt, B., Newar, D. K. P., Panday, S. P., and Pandey, R. K. (2019). "Casualty Information Extraction and Analysis from News". In: *Proceedings of the 16th International Conference on Information Systems for Crisis Response And Management*. Valencia, Spain: Iscram.

Cilli, R., Elia, M., D'Este, M., Giannico, V., Amoroso, N., Lombardi, A., Pantaleo, E., Monaco, A., Sanesi, G., Tangaro, S., et al. (2022). "Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe". In: *Scientific Reports* 12.1, p. 16349.

Cowie, J. and Lehnert, W. (Jan. 1996). “Information Extraction”. In: *Communications of the ACM* 39.1, pp. 80–91.

Crawford, K. and Finn, M. (2015). “The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters”. In: *GeoJournal* 80, pp. 491–502.

Davis, I., Fraser, R., Ahmed, A., Basnayake, S., Deocariza, M., Ficcadenti, B., Iglesias, G., Pearson, L., Roy, A., and Tawhid, K. (Jan. 2013). *Integrating Disaster Risk Management into Urban Management. Disaster Risk Management Practitioner's Handbook Series.*

European Commission (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).*

Gevaert, C. M., Carman, M., Rosman, B., Georgiadou, Y., and Soden, R. (2021). “Fairness and accountability of AI in disaster risk management: Opportunities and challenges”. In: *Patterns* 2.11, p. 100363.

Imran, M., Mitra, P., and Castillo, C. (May 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'16)*. Portorož, Slovenia: European Language Resources Association (ELRA), pp. 1638–1643.

Imran, M., Ofli, F., Caragea, D., and Torralba, A. (2020). “Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions”. In: *Information Processing Management* 57.5, p. 102261.

Kejriwal, M. and Zhou, P. (2020). “On detecting urgency in short crisis messages using minimal supervision and transfer learning”. In: *Social Network Analysis and Mining* 10.1, p. 58.

Kruspe, A., Kersten, J., and Klan, F. (2019). “Detecting event-related tweets by example using few-shot models”. In: *16th International Conference on Information Systems for Crisis Response and Management*. Valencia, Spain: Iscram.

Kuglitsch, M. M., Pelivan, I., Ceola, S., Menon, M., and Xoplaki, E. (2022). “Facilitating adoption of AI in natural disaster management through collaboration”. In: *Nature Communications* 13.1, p. 1579.

Lei, T., Barzilay, R., and Jaakkola, T. (Nov. 2016). “Rationalizing Neural Predictions”. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, pp. 107–117.

Li, T., Xie, N., Zeng, C., Zhou, W., Zheng, L., Jiang, Y., Yang, Y., Ha, H.-Y., Xue, W., Huang, Y., et al. (Mar. 2017). “Data-Driven Techniques in Disaster Information Management”. In: *ACM Comput. Surv.* 50.1.

Mayer, A.-S., Strich, F., and Fiedler, M. (2020). “Unintended Consequences of Introducing AI Systems for Decision Making.” In: *MIS Quarterly Executive* 19.4.

McCreadie, R., Macdonald, C., and Ounis, I. (2016). “EAIMS: Emergency Analysis Identification and Management System”. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '16. Pisa, Italy: Association for Computing Machinery, pp. 1101–1104.

Moitra, A., Wagenaar, D., Kalirai, M., Ahmed, S. I., and Soden, R. (Nov. 2022). “AI and Disaster Risk: A Practitioner Perspective”. In: *Proceedings of the ACM on Human-Computer Interaction* 6.CSCW2.

Payrovnaziri, S. N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J. H., Liu, X., and He, Z. (May 2020). “Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review”. In: *Journal of the American Medical Informatics Association* 27.7, pp. 1173–1185.

Rajagopal, D., Tandon, N., Clark, P., Dalvi, B., and Hovy, E. (Nov. 2020). “What-if I ask you to explain: Explaining the effects of perturbations in procedural text”. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, pp. 3345–3355.

Ramchurn, S. D., Huynh, T. D., Wu, F., Ikuno, Y., Flann, J., Moreau, L., Fischer, J. E., Jiang, W., Rodden, T., Simpson, E., et al. (2016). “A Disaster Response System based on Human-Agent Collectives”. In: *Journal of Artificial Intelligence Research* 57, pp. 661–708.

Rasheed, Z. and Pietromarchi, V. (2023). *Turkey-Syria earthquake live: 'Worst in a century' – WHO*. URL: https://www.aljazeera.com/news/liveblog/2023/2/14/turkey-syria-earthquake-live-news-death-toll-tops-36000 (visited on 02/14/2023).

Simões-Marques, M. and Figueira, J. R. (2019). “How Can AI Help Reduce the Burden of Disaster Management Decision-Making?” In: *Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2018*

International Conference on Human Factors and Systems Interaction, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9. Springer, pp. 122–133.

Soden, R., Wagenaar, D., Luo, D., and Tijssen, A. (2019). "Taking Ethics, Fairness, and Bias Seriously in Machine Learning for Disaster Risk Management". In: *CoRR* abs/1912.05538. arXiv: 1912.05538.

Soden, R., Wagenaar, D., and Tijssen, A. (2021). *Responsible Artificial Intelligence for Disaster Management: Working Group Summary*. World Bank, the Global Facility for Disaster Reduction and Recovery, the Deltares, p. 44.

St Denis, L. A., Hughes, A. L., Diaz, J., Solvik, K., Joseph, M. B., and Balch, J. K. (2020). "'What I Need to Know is What I Don't Know!': Filtering Disaster Twitter Data for Information from Local Individuals". In: *Proceedings of the Proceedings of 17th International Conference on Information Systems for Crisis Response and Management*. Blacksburg, VA (USA): Virginia Tech, pp. 730–743.

Sun, W., Bocchini, P., and Davison, B. D. (2020). "Applications of artificial intelligence for disaster management". In: *Natural Hazards* 103.3, pp. 2631–2689.

Tjoa, E. and Guan, C. (2021). "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.11, pp. 4793–4813.

Tonekaboni, S., Joshi, S., McCradden, M. D., and Goldenberg, A. (Sept. 2019). "What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use". In: *Proceedings of the 4th Machine Learning for Healthcare Conference*. Ed. by F. Doshi-Velez, J. Fackler, K. Jung, D. Kale, R. Ranganath, B. Wallace, and J. Wiens. Vol. 106. Proceedings of Machine Learning Research. PMLR, pp. 359–380.

Vaassen, B. (2022). "AI, Opacity, and Personal Autonomy". In: *Philosophy & Technology* 35.4, p. 88.

Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., Mestav Sarica, G., Ruangpan, L., Molinario, G., and Lallemant, D. (2020). "Invited perspectives: How machine learning will change flood risk and impact assessment". In: *Natural Hazards and Earth System Sciences* 20.4, pp. 1149–1161.

Yu, M., Yang, C., and Li, Y. (2018). "Big Data in Natural Disaster Management: A Review". In: *Geosciences* 8.5.

Zahra, K., Das, R. D., Ostermann, F. O., and Purves, R. S. (2022). "Towards an Automated Information Extraction Model from Twitter Threads during Disasters". In: *Proceedings of the 19th International Conference on Information Systems for Crisis Response and Management*. Tarbes, France, pp. 637–653.

Zhai, C. and Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. Vol. 12. Association for Computing Machinery and Morgan Claypool.

Zhang, C., Yang, Y., and Mostafavi, A. (2021). "Revealing Unfairness in social media contributors' attention to vulnerable urban areas during disasters". In: *International Journal of Disaster Risk Reduction* 58, p. 102160.