Writing out the Storm: Designing and Evaluating Tools for Weather Risk Messaging

Sophia Jit
University of Toronto
Toronto, Ontario, Canada
sophia.jit@mail.utoronto.ca

Jennifer Spinney
York University
Toronto, Ontario, Canada
jspinney@yorku.ca

Priyank Chandra
University of Toronto
Toronto, Ontario, Canada
priyank.chandra@utoronto.ca

Lydia Chilton
Columbia University
New York, New York, USA
lc3251@columbia.edu

Robert Soden
University of Toronto
Toronto, Ontario, Canada
soden@cs.toronto.edu

ABSTRACT

Communicating risk to the public in the lead-up to and during severe weather events has the potential to reduce the impacts of these events on lives and property. Globally, these events are anticipated to increase due to climate change, rendering effective risk communication an integral component of climate adaptation policies. Research in risk communications literature has developed substantial knowledge and best practices for the design of risk messaging. This study considers the potential for quantifying the compliance of severe weather risk messages with these best practices, individually and at scale, and developing tools to improve risk communication messaging. The current work makes two contributions. First, we develop a string-matching approach to evaluate whether messaging complies with best practices and suggest areas for improvement. Second, we conduct an interview study with risk communication professionals to inform the design space of authoring tools and other technologies to support severe weather risk communicators.

CCS CONCEPTS

• **Human-Computer Interaction**; • **Empirical Studies in HCI**; • **Information Systems**;

KEYWORDS

Creativity Support; Crisis/Disaster; Empirical study that tells us about how people use a system

ACM Reference Format:
Sophia Jit, Jennifer Spinney, Priyank Chandra, Lydia Chilton, and Robert Soden. 2024. Writing out the Storm: Designing and Evaluating Tools for Weather Risk Messaging. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3613904.3641926

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3641926

1 INTRODUCTION

Weather risk communication is an interdisciplinary field of research and practice that concerns the communication of accurate and actionable information during weather-related hazards that facilitates effective disaster response [96, 112]. Weather risk messages may be issued before, during, and after such events occur and may thus span both the short- and long-term. Short-term messaging concerns the communication of immediate threats while long-term messaging is intended to encourage preparation and planning in advance. Regardless of the time frame, both are integral to disaster response because of their ability motivate the public to take protective actions [17]. Indeed, effective risk messaging is characterized by the ability to cause a cognitive shift in the audience that may result in life-saving behavioural changes [59, 71, 79]. To achieve this, prior work has found that risk messages must contain certain thematic and stylistic elements which determine which actions the audience takes and their confidence to take them [117]. However, the creation of effective risk messaging remains a challenging task as risk communication professionals must often contend with varying degrees of certainty and competing priorities in a high-stress situation [58, 84]. Such challenges may result in hazard-related gaps in the flow of information and risk perception in the population, which may benefit from technological and digital innovation [46].

The evaluation of risk communication at scale and the development of writing support tools for risk communication professionals offers opportunities for HCI and crisis informatics researchers to contribute to risk communication and practice. To date, research in these disciplines has largely focused on the exchange of risk information between the public and official agencies [72, 122, 130], or online risk communication during public health emergencies [52, 53, 96]. The development of quantitative approaches to risk communication remains a nascent area of research. Meanwhile, much of the literature that has looked to develop writing support tools in non-fiction settings has focused on sentence completion [21] and prediction in online communication [21, 63]. Recent work in idea generation [49, 93] offers particular insights into the development of a writing support tool for risk communication professionals as the inclusion of particular ideas or themes is integral to preventing losses.

To develop an approach to evaluate the degree to which alert messages issued by Environment and Climate Change Canada (ECCC) comply with best practices, and to understand the design space for an authoring support tool for risk communication professionals, this paper undertook two related studies. First, we used string-matching [20, 110], a methodology that has been used in fields such as computational biology [6, 22, 119] and text retrieval [4, 62], to semi-automatically evaluate structured risk communication at scale. Canada was chosen as the subject of the analysis as it is currently warming at twice the global rate and is expected to experience significant losses, damages, and disruptions due to climate change over the next 20-years [86]. Thematic analysis – a methodology employed in much of the HCI literature to analyze qualitative data [25] – acts as a starting prompt for much of this work. We initially used this methodology to inductively and deductively identify key themes in a subsample of risk messages issued by ECCC. To evaluate the entire corpus, we calculated the readability score of the messages and used our string-matching algorithm to semi-automatically label the entire sample. In doing so, we are able to overcome the small sample sizes that often characterize data in this literature and evaluate the corpus' compliance with best practices in risk communication.

Second, we conducted semi-structured interviews with 10 emergency managers (EMs) and risk communication professionals from the study region. Using the results of the first study, we presented participants with a potential writing support tool which identified omitted themes, and four variations of a risk message to explore possible features of this technology. Participants confirmed the importance of the themes identified through our string-matching method, and noted the potential utility of a writing support tool for both experienced and inexperienced risk communicators. In addition to noting practical considerations regarding the design and adoption of emerging technologies in this domain, participants also highlighted the potential for participatory approaches to the design of risk communications. This research thus informs HCI research into risk communications policy and strategy, and contributes design guidance towards the creation of technologies to improve severe weather risk communication.

2 RELATED WORK

2.1 HCI and Crisis-Informatics

Risk communication has been a growing subject of interest to both HCI and crisis informatics researchers. Much of the work in the latter discipline has examined the exchange of risk information between authoritative bodies and the public to understand the role of ICTs in aiding contextual awareness between these groups [72, 122, 130]. For example, Bica et al. [13] examine the diffusion, prevalence, and public reception of hurricane risk images from the 2017 Atlantic hurricane season shared by authorities on social media. Recent work in this field has similarly turned its attention to informal risk communication, including how volunteer technology communities organize to support such messaging [107, 115]. Early work at the intersection of HCI and crisis informatics has similarly examined how emergency managers and volunteers curate and manipulate social media at various stages of a disaster [26, 116], and how digital humanitarians construct real-time narratives during crises [85]. Researchers in this domain have paid particular attention to how the public uses social media to communicate and understand localized risk [89], and to spread, curate, and seek information in response to disasters [69, 88, 98, 108] and public health emergencies [52, 53, 96]. Many authors have noted that social media allows for more nuanced crisis communication, with an emphasis on collective sensemaking [64, 67, 68, 114], collective action to understand crises [64], and two-way dialogue between the public and official sources [57, 69]. Severe weather risk communication specifically has been the focus of recent design-related HCI scholarship. Researchers have outlined a series of design opportunities for HCI practitioners to support digital risk communication and used the public's behavior during storms to help inform the design of tools to better meet their needs [112].

2.2 Risk Communication

Creating effective risk communication has been the focus of much of the emergency management literature. This field of research largely employs qualitative and quantitative methods to understand how individuals make risk-related decisions during hazardous events, and how to communicate complex information to the public. This literature has found that trust in organizations, governmental agencies, or spokespeople is a major determinant of risk reduction and protective actions during weather-related hazards [81, 123]. These trusted authorities should work with a diverse group of institutions, scientists, and public officials to communicate accurate information with a unified message or voice [35]. Although most risk communication during hazards is often unidirectional – going from authorities to the public – a growing body of work has found that two-way dialogue between creators and recipients of risk communication is more effective during crisis situations and disaster mitigation efforts [33, 131]. Furthermore, educational materials and risk messages should be targeted to a specific audience and use multiple communication channels [12, 105, 111]. Successful weather risk communication should thus be designed to meet the needs, vulnerabilities, and cultural beliefs of the intended audience [65, 80, 83, 111].

Identifying the elements of effective risk communication has been of particular interest to emergency management researchers. Research in this domain has found that effective risk messaging is often consistent, clear, and concise and does not rely on fear-based language [31, 78, 82]. Specificity is a key determinant of protective actions as the audience is more likely to both believe a threat is credible and personalize the hazard risk [36]. As a general guideline, risk communication should thus include discussion of hazards, location, timing, vulnerability, and recommended actions [28, 29, 44, 77, 78, 95]. The inclusion of hazards and its associated impacts may help inform the logic of protective actions, while hazard-related risk is determined by its location and the time that the event will strike – information which allows individuals to respond accordingly [78]. Similarly, an understanding of one's vulnerability and recommended actions are integral to the adoption of protective behaviors [28, 29, 95]. Stylistically, discussion of the hazard and safety-related actions should be delivered with a high degree of certainty in plain language that is understandable to the public [94].

The first part of this study expands upon previous literature by using string-matching to quantify the corpus' compliance with best practices in risk communication. The proposed string-matching algorithm is predicated on inductive and deductive thematic analysis, where the previously identified themes related to specificity – hazards, location, timing, vulnerability, and recommended actions – informed the latter. These themes were chosen by virtue of their role in informing risk perception and motivating protective actions during severe weather events [15, 16, 29, 78]. To evaluate the plainness and comprehensibility of the language, reading level metrics for the corpus were also computed. Additionally, the second part of this study confirms much of the findings of previous risk communication literature, with many interview participants referring to these best practices, and their role in personalizing risk, in their assessment of effective risk messaging.

2.3 Writing Support Tools

With the development of improved and advanced language models, the academic literature on writing support tools has experienced a resurgence in recent years, with tools to support both creative and scientific writing. Much of the work in the former focused on narrow support for writing tasks, such as storytelling [102] and metaphor writing [19]. More varied and diverse assistance to writers, such as assistance with descriptions, plot points, or providing questions [27, 38] has been the subject of recent work in the field. Scholarship in the domain of scientific writing has thus far been more limited, with much of this work focusing on sentence completion [21] or smart replies in emails [63]. As opposed to providing complete phrases or sentences, Peng et al. [93] develop a tool to assess and provide suggestions to posts created for online mental health communities. Similarly, Gero et al. [49] follow up on their nascent research on online scientific communication on Twitter [50] and use pre-trained language models to generate suggestions for science-related text. The present study builds upon this literature on idea generation to suggest a writing support tool for risk communication professionals. In lieu of the machine learning (ML) methods employed in previous studies, this research employs string-matching to identify the absence of themes related to best practices in risk communication, which can be used to both analyze risk communication at scale and suggest the inclusion of omitted themes. To further explore the design space for a writing support tool, we conducted semi-structured interviews with EMs and risk communication professionals to evaluate the utility of the proposed technology and identify potential features that could assist this group in practice.

3 STUDY 1: ANALYSIS OF ECCC RISK MESSAGES

In order to understand current practices in risk communication in Canada, we undertook a quantitative and qualitative analysis of historical risk messages issued by ECCC – the branch of the Canadian federal government responsible for coordinating environmental policies and programs. We first computed the reading level metrics for the sample, engaged in inductive and deductive thematic analysis on a subset of the data to identify the presence of themes associated with best practices, and then used these themes to develop a string-matching algorithm to semi-automatically label the corpus.

3.1 Methodology

3.1.1 Data. The current work employs 9,181 public weather alerts issued by ECCC between June 2020 and December 2021. In the data, weather alerts span four categories – special weather statements, advisories, watches, and warnings – and 23 hazards of varying intensity (see Figure 4). The structure of weather alerts is comprised of three sections as indicated in Figure 1: (1) summary, (2) location information, and (3) message body. The summary provides an overview of the message, including the type of weather event, the affected locations, and the date and time the message was issued. The location section summarizes all the affected locations, and the message body contains the main risk message. Multiple alerts may have been issued for a single weather event, where the maximum number of messages associated with an event in the sample is five. The current work focuses on an analysis of the message bodies. Although the data set is composed of weather alerts issued by the federal government to the public, the corpus was compiled by and purchased from the issuing body by the research team and thus cannot be shared publicly.

3.1.2 Reading Level Metrics. To understand the reading level of the messages, the Flesch Reading Ease (FRE) score – one of the most common methods to assess text readability [41, 45] – was computed for the message bodies of the sample. This score ranges from 0 to 100, where higher scores indicate that the text is relatively easy to read for the average adult (see Table 3). To calculate the FRE score, basic text pre-processing was undertaken such as replacing all website names and social media accounts with generic word identifiers such as "website URL" and "email account", before applying the TextStat library to the data. Like most readability formulas, the FRE score does not evaluate features of text such as content, organization, word order, format, mood, or tone [99]. An assessment of these linguistic features, and attributes such as consistency, clarity, and conciseness, may thus require ML-based methods for automatic text evaluation [9, 40, 70, 97] which are beyond the methodological scope of the present work.

3.1.3 Manual Thematic Analysis. To assess the messages' compliance with best practices in risk communication, the research team first engaged in deductive and inductive thematic analysis [25]. Given the size of the data set (N=9,181), each member of the research team reviewed 10 random message bodies for each of the five most prevalent weather events to inductively find repeated patterns and themes. This process helped identify a preliminary list of qualitative codes, which the team then updated and deductively refined through discussion of similarities and differences between the patterns in the message bodies and best practices in risk communication. This deductive and inductive process was repeated iteratively until a final list of themes, was agreed upon: hazards, location, timing, affected populations, impacts, and recommended actions. The inductive method generated two additional themes – "impacts" and "additional information" – relative to the deductive approach although the latter is omitted from the remainder of the analysis as it is not a theme associated with best practices. Table 4 provides a summary of the identified themes and their corresponding interpretation.

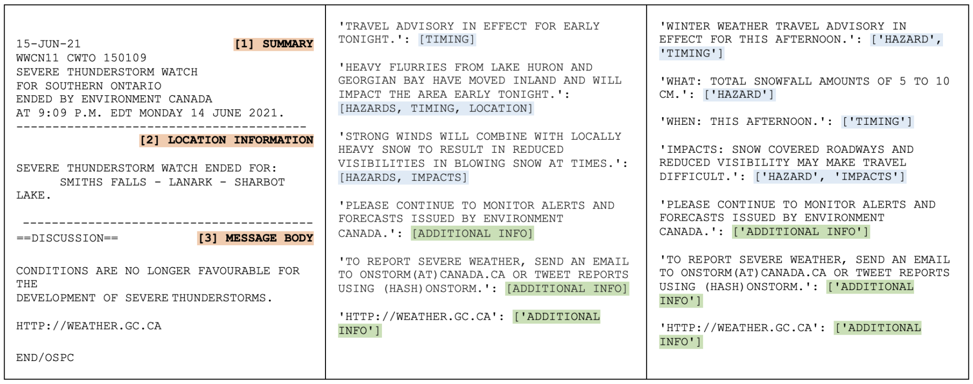

Figure 1: ECCC Severe Weather Risk Messages. From left to right, an example of: (a) an ECCC risk message, (b) a message labelled using thematic analysis, and (c) a message labelled using semi-automatic message classification. Text in green denotes a theme – "Additional Information" – that was omitted from the primary analysis.

3.1.4 Semi-Automatic Message Classification. To systematically analyze the sample's compliance with best practices in risk communication, the research team used the labels identified in 3.1.3 and keywords to semi-automatically label the data set. Although ML models may be employed with this data, the research team chose to use string-matching because such an algorithm is highly interpretable and computationally efficient relative to these models and they may still learn the necessary representations while avoiding extensive training. For each of the five most prevalent weather events, a preliminary set of keywords and phrases associated with the identified themes were selected to label the data [58, 84]. Given the general uniformity of the messages within weather events and the frequent recycling of text, keywords and phrases were chosen by virtue of their repeated appearance in weather alerts and their ability to uniquely identify a theme. The presence of a given keyword or phrase within a sentence was used to label a sentence with that theme, where sentences could have multiple labels. Keywords or phrases that uniquely identified a given theme in one context, but whose meaning varied given the context of another theme, was used to create a list of exclusion criteria. For example, the words "north" and "northerly" were used as inclusion and exclusion criteria respectively to identify discussion of locations.

To develop a full list of keywords and phrases, a similar process as in 3.1.3 was followed where the preliminary list was used to label 10 random messages for each of the five events. The authors then evaluated the accuracy of the keywords on these messages. If the accuracy was below 90%, the list of keywords was updated with additional keywords and phrases, and further evaluated on 10 new, random messages; this process was repeated until the desired accuracy was achieved. Once an accuracy of 90% or greater was achieved, the authors then tested this set of keywords and phrases, which ranged in size from 105 - 150 across the events, on 50 randomly selected messages per event and confirmed that the model continued to achieve this accuracy. Figure 1 includes an example of results from both the manual thematic analysis and the semi-automatic message classification.

3.2 Results

3.2.1 Summary Statistics. Table 1 provides a summary of the data set, where the hazards are organized by the event category they belong to. The percent column notes the total percent of the entire sample that comprises each category and the total percent of the category that comprises a given hazard. As indicated in the table, five events – severe thunderstorm warnings (23.3%), special weather statements (17.04%), weather advisories (11.53%), heat warnings (7.87%), and snow squall warnings (6.42%) – make up 66.18% of the total sample (see Figure 4). Across the messages, the word count ranged from one word to a maximum of 372 words (Tornado Warning). Messages that only included a single word reflected a message body that simply included a URL for additional information. Of the total sample, 2.97% contain only URLs of which 2.86% are the first message to be issued in the sample, and the remaining 0.11% are the second message to be issued. For the analysis, this data is dropped from the sample. Although the content and structure of the message bodies varied across weather events, message bodies within a given weather event demonstrated significant similarity such that many messages repeated the same text with event-specific details. The uniformity of messages within weather events is thus conducive to the use of string-matching [18, 20, 110] to analyze the data.

3.2.2 Reading Level Metrics. The final columns of Table 1 provide summary statistics for the FRE score for each event category and hazard. For the five most prevalent weather events, the range of average FRE scores ranged from 58.23 (Severe Thunderstorm Warning) to 67.1 (Special Weather Statement). Although there was significant variation within the unit of analysis, these results indicate that the reading level of the risk messages ranged from fairly difficult to standard difficulty. As noted by [74], for risk communication to be accessible to the public, it should ideally be written at a fifth or sixth grade reading level, which corresponds to an FRE score between 80 (easy) and 100 (very easy). As noted in Table 1, the average reading level across all of the weather events, including the five most prevalent ones, is thus higher than the recommended value which may undermine their effectiveness in practice.

Table 1: Summary Statistics for Risk Messages Issued by ECCC

	Word Count				Flesch Reading Ease Score[a,b]		
Weather Event	Percent	Mean (Std)	Min	Max	Mean (Std)	Min	Max
Statement	17.04	93.7 (52.2)	1	273	67.1 (8.44)	31.55	103.63
Special Weather	100	93.7 (52.2)	1	273	67.1 (8.44)	31.55	103.63
Advisory	19.26	88.65 (49.15)	1	231	59.72 (8.95)	37.47	95.17
Blowing Snow	1.98	90.11 (42.89)	12	166	56.96 (9.00)	41.36	77.33
Freezing drizzle	3.9	88.45 (42.61)	1	167	65.11 (5.45)	54.83	77.91
Frost	16.18	71.08 (31.38)	7	161	62.19 (6.75)	47.79	81.29
Fog	18.04	69.33 (34.23)	1	154	54.7 (9.52)	37.47	89.75
Weather	59.9	99.18 (54.17)	1	231	60.31 (8.82)	43.9	95.17
Watch	9.51	131.06 (63.23)	1	253	51.89 (9.05)	36.79	81.63
Winter storm	0.46	171 (8.12)	166	183	67.13 (7.58)	57.06	72.97
Tornado	4.12	183.69 (72.14)	11	253	50.58 (5.97)	41.26	59.3
Snow Squall	29.55	130.45 (50.86)	1	247	62.02 (6.15)	44.64	81.63
Severe Thunderstorm	65.86	127.76 (66.35)	1	242	47.32 (6.02)	36.79	77.91
Warning	54.2	143.69 (68.39)	1	372	60.86 (10.5)	36.96	100.24
Blizzard	0.24	152.58 (22.77)	129	191	68.31 (6.22)	57.67	76.62
Extreme Cold	0.68	81.76 (41.93)	12	132	67.32 (10.7)	49.82	77.53
Winter Storm	1.91	117.91 (52.76)	8	204	60.63 (5.02)	49.31	72.97
Wind	3.64	86.77 (39.9)	11	150	73.9 (4.75)	58.79	80.28
Tornado	4.18	213.49 (76.71)	5	372	56.69 (7.66)	43.19	99.23
Freezing Rain	5.47	101.91 (48.83)	1	193	64.04 (6.61)	45.93	77.91
Snowfall	5.83	115.87 (49.23)	1	239	61.39 (9.62)	37.98	81.33
Rainfall	7.6	125.33 (63.05)	7	270	60.71 (10.56)	37.98	88.74
Snow Squall	11.84	124.33 (54.71)	8	239	62.29 (5.16)	41.87	83.96
Heat	14.53	166.63 (57.02)	11	347	63.89 (6.08)	47.79	77.13
Severe Thunderstorm	44.09	153.46 (71)	1	287	58.23 (12.47)	36.96	100.24

[a] *The interpretation of the Flesch Reading Ease score is as follows: 90-100: Very Easy; 80-89: Easy; 70-79: Fairly Easy; 60-69: Standard; 50-59: Fairly Difficult; 30-49: Difficult; 0-29: Very Confusing.*

[b] *A feature of the TextStat library is that the computed Flesch Reading Ease scores have an upper bound of 121.22 and no lower bound where values above 100 and below zero are interpreted as "Very Easy" and "Very Confusing" respectively.*

3.2.3 Semi-Automatic Message Classification. For the five most prevalent weather events, Table 2 shows the percent of messages that include discussion of the themes identified in 3.1.3. For this analysis, weather alerts that simply noted an event had come to an end were dropped from the sample. This decision was motivated by the assumption that these messages no longer required the audience to engage in protective actions and thus did not necessitate the corresponding elements of effective risk communication. As noted in Table 2, there is significant variation in the presence of themes related to risk communication best practices in the remaining sample. For example, across all five weather events, the majority of the sample included discussion of timing (83.33% – 100%) and hazards (86.67% – 99.77%). However, the discussion of affected populations was included in a minority of Special Weather Statements (14.93%), Snow Squall Warnings (23.62%) and Weather Advisories (39.78%). Ideally, for weather alerts to comply with best practices, the relevant themes would be present in all the messages across all weather events. These results are hardly born out in the data as no weather event includes all the themes and only a handful of select themes top 99%.

4 STUDY 2: INTERVIEW STUDY

To validate the results of the first study, and to consider the design of a writing support tool for risk communicators, we undertook an interview study with EMs and risk communication professionals. The focus of these interviews was the evaluation of a risk communication authoring tool and four risk messages, with the intention of identifying desirable features and the utility of the proposed technology, and challenges to its adoption.

4.1 Methodology

4.1.1 Participant Overview. In total, 10 EMs and risk communication professionals were interviewed for this study. To recruit participants, the research team contacted 15 current or retired EMs and risk communicators in their professional network who had work experience in Ontario. Participants were recruited in June 2023 and all of the interviews were conducted in June and July of the same year. All participants had formal education in disaster management and risk communication and were either currently employed or had retired from the field. Cumulatively, participants had experience working for government agencies, hospitals, and universities in Ontario.

Table 2: Prevalence of Themes Related to Best Practices in Risk Communication in ECCC Risk Messages

Event	Timing Information	Affected Population	Recommended Actions	Hazards	Locations	Impacts
Snow Squall Warning	96.46%	23.62%	79.72%	99.41%	56.69%	98.82%
Special Weather Statement	94.88%	14.93%	27.82%	96.49%	52.05%	55.86%
Weather Advisory	85.47%	39.78%	59.42%	86.67%	41.48%	71.74%
Heat Warning	100.0%	93.19%	100.0%	99.11%	51.7%	88.74%
Severe Thunderstorm Warning	83.33%	62.19%	83.15%	99.77%	76.12%	82.42%

4.1.2 Data Collection and Analysis. Interviews primarily concerned participant feedback on four risk messages and a potential design of a risk communication authoring support tool. For each of the five weather events that was the focus of study 1, the research team chose a risk message from that category in the ECCC data to serve as the baseline (Example 1 in Figure 2). For each baseline risk message, a potential writing support tool (Figure 3; Example 2(a)) and three variations of that message (Example 2(b) - Example 4 in Figure 2) were created and shown to participants. Using the algorithm developed in the first study, Example 2(a) displayed a static image of a text editor and a popup window identifying omitted themes in the baseline risk message and suggested their inclusion. The interviewer explained that users could write their risk message in the text editor and submit it as an input to the algorithm via the "Check" button to identify omitted themes. Example 2(b) amended the baseline risk message by including the omitted themes using text from historical risk messages. As an example of specialized risk messages that could address the needs of diverse groups, Example 3 tailored the risk message for members of the visually impaired community. This example focused on the visually impaired community because prior research has shown that the information needs of this group are often unmet by current risk communication approaches [113]. Furthermore, as a part of a larger research project investigating the intersection of HCI and risk communication, the research team engaged in an interview study about severe weather events and risk communication with members of this group. Finally, Example 4 consisted of a variation of the baseline risk message whose language had been simplified using ChatGPT to lower the reading level of the text, and thus improve its accessibility. For this example, the research team queried the chatbot to "Rewrite the following text to be at a fifth-grade reading level. Preserve the meaning of the text. Keep the format and tone of the text the same." The research team chose to use ChatGPT to simplify the messages because the size and diversity of its training data rendered it well-equipped to handle a wide range of contexts and topics across different domains. Although the research team initially considered employing popular pre-trained models for text simplification, we determined that these may not generalize well to the domain of severe weather risk communication given the particular data they were trained on.

Interviews were conducted online and asynchronously via Zoom from June 2023 to July 2023 by one member of the research team. Participants were sent an example interview guide in advance, which included reference to the risk messages to be evaluated although the actual examples were omitted. Interviews were conducted in English. Verbal consent was obtained at the start of the interview, which were both audio and video recorded. Following the completion of five interviews, the research team discussed the interview questions based on preliminary findings with the intention of reevaluating the interview guide. However, no changes were made to the guide at this time. Interviews were scheduled to last between 45 - 60 minutes, although in practice they ranged from 55 - 93 minutes. The interview data was transcribed using Otter.ai. To finalize the transcripts for analysis, the member of the research team who conducted the interviews verified and corrected them. During this process, the author consulted the audio and corrected the transcripts when they came across any unclear or ambiguous text that impeded its comprehension. Transcripts were edited to correct for unclear or ambiguous text, and not for speech disfluencies or punctuation.

To analyze the data, two members of the research team engaged in inductive thematic analysis – a form of reflexive thematic analysis where codes and themes are primarily derived from the data set [25]. As a starting point, the first author reviewed the 10 transcripts, and proposed a set of qualitative codes they thought represented their reading of the data. The first and last authors then brainstormed possible themes these codes could support and further refined them. These authors engaged in iterative rounds of discussion where both the themes and codes were further developed while consulting the transcripts. To generate the final list of codes, the first author then re-examined the coded data that corresponded to each suggested theme.

4.2 Results

4.2.1 Omitted Themes, Language, and the Efficacy of Risk Communication. For the most part, participants confirmed the analysis in Section 3.2 (as displayed in Example 2(a)), and noted the importance of including the omitted themes in risk communication. In our participants' telling, the omitted themes played varied roles in the risk message – they allowed the audience to personalize the risk and to understand how to act – such that it was important to include as many of them as possible. Doing so would thus allow a degree of specificity in the communication that enabled the audience to interpret the risk message as being relevant to them and encourage them to take actions. As one EM noted:

> *"I think there would need a little bit more detail in order to actually persuade recipients of the message to take action right now. Like, just from my point of view, it seems a little general. So especially with my EM lens on, I wouldn't I, I would be kind of questioning, is this impacting me? Is this impacting people up north? South?*

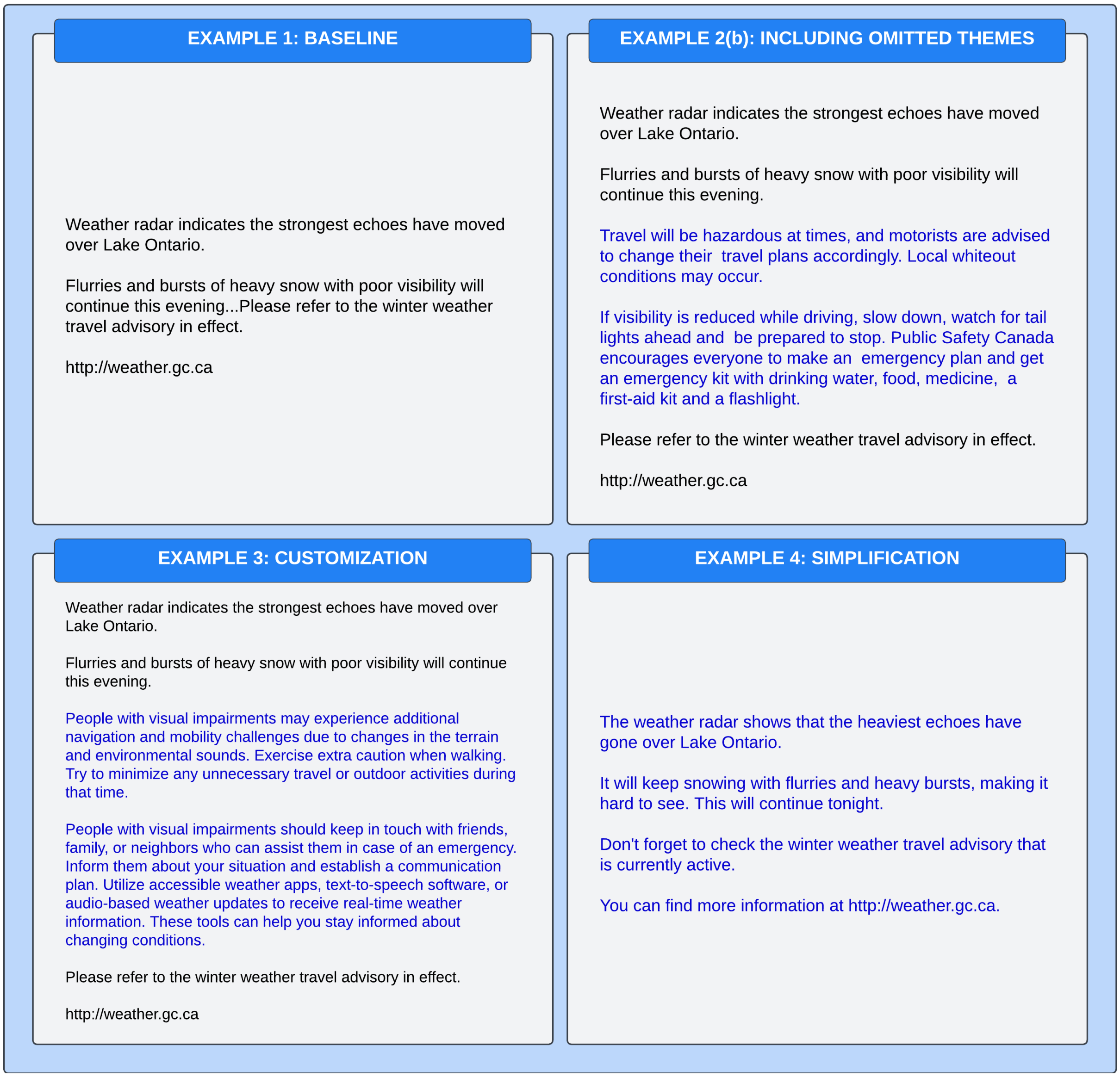

Figure 2: Example Risk Messages Shown to Participants. From top to bottom, an example of a: (1) baseline risk message taken from the ECCC data, (2b) baseline risk message with the omitted themes included, (3) baseline risk message customized for people with visual impairments, and (4) baseline risk message with simplified language. The blue text indicates the changes made to the baseline message.

> *Do I need to worry about this?...So I think adding more details where readers or recipients could personalize it would be very helpful because as soon as you personalize it, that's when people are more apt to take action. As soon as they see that it relates to them." (P8)*

While most participants agreed that the second example (e.g., Example 2(b) in Figure 2) included the omitted themes, they similarly critiqued the lack of specificity of the additional text. Respondents observed that the new text did not include the appropriate level of detail to instruct actions and enable an evaluation of personal risk. One participant likened their inclusion to a legal formality, stating that *"It's almost like...legally, we got to put in something...that says...exercise caution, but really, it's not giving any, like not really instructing anyone to do anything." (P7)*. Furthermore, despite modest gains in the efficacy of the message, interviewees highlighted that features such as jargon, technical language, and the reading level continued to pose challenges to protective decision-making.

4.2.2 Reading Level, Jargon, and Tone. Most participants critiqued the language used in the baseline message, which they hypothesized was not simple enough for the layperson to understand. Of

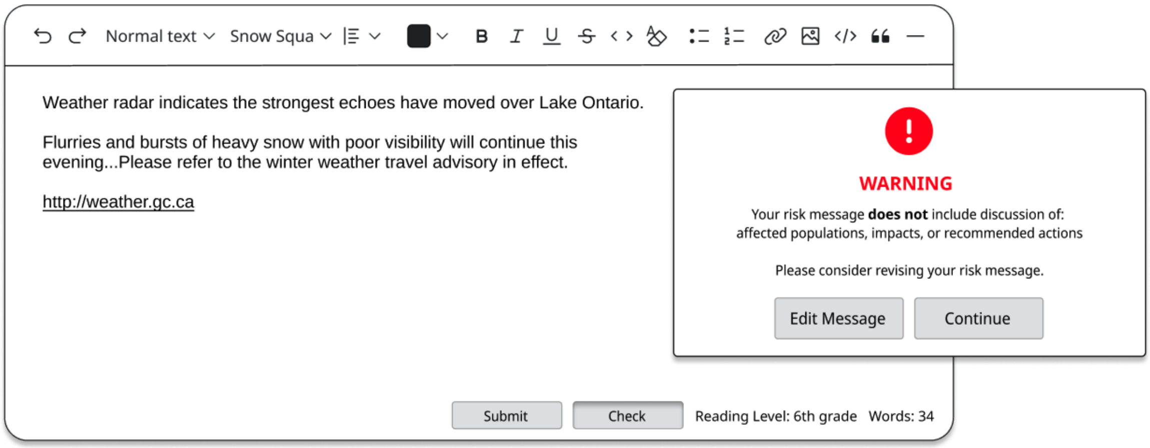

Figure 3: Example 2(a): A Potential Writing Support Tool.

particular concern was the use of scientific concepts and jargon in the message. Some interviewees indicated that such language may be confusing to the audience or fail to capture their attention – both deterrents to appropriate risk evaluations – or, worse yet, would be a disincentive to read the messages entirely. Similarly, some subjects highlighted the potential confusion around the use of metrics in risk messaging. Chief amongst the participants' concerns was the inability of the layperson to conceptualize metrics and thus accurately assess their personal risk. In addition to simply being unable to grasp, conceptually, how much a given quantity is, as one participant noted, the interpretation of a metric might itself vary with the season:

> *"So scrolling through kind of linearly and systematically, heavy rainfall, up to 60 millimeters, that can mean a variety of different things, depending on the season, and depending on the type of year so we could have 25 millimeters of rain not being an issue if it's taking place at this time of the year. But if it's in the winter, spring, it's a bit more. And so that's one of the challenges we find with this, is that just interpreting what is 60 millimeters? And what does that look like? And depending on the conditions, what does that mean?" (P2)*

Although the use of the AI chatbot was intended to simplify the language and thus improve the efficacy of the risk messages, many participants noted the counter productivity of this endeavor. First, since the chatbot was instructed to simplify the risk messages to a fifth grade reading level, the scientific concepts and jargon were still present in the communication. Participants thus acknowledged that the revised messages employed simpler language, which was potentially more intelligible, although they highlighted persistent issues with comprehension. Second, despite specifying that the message retain the tone and content of the baseline message, most interviewees observed that the simplified message was too casual, and that much of the urgency was therefore lacking. This casualness, our interviewees worried, made the amended message sound less authoritative and credible, which would potentially cause the audience to underestimate the urgency of the situation and fail to both capture their attention and motivate them to act.

> *"I think a lot of individuals also take comfort from the fact it's like, psychologically, when you're reading a message from Environment Canada, you appreciate the detail it provides and you appreciate that you're reading sounds like it's formal. And it's come from a body of experts, rather than something that you could draft yourself in your own plain language. I think removing those details and replacing it with very light, you know, casual language would take away what was meant and also might people might take it a little less seriously, less likely to follow the recommended protective actions. If it sounds like something, you know, your friend would tell them as opposed to Environment Canada." (P8)*

To address potential trade-offs between including the omitted themes and overwhelming the audience with information, participants highlighted the importance of strategically organizing and displaying the message. In addition to suggestions such as including specific information in lieu of general details, and writing more succinctly, many interviewees suggested reorganizing the message and themes so that the audience received the most important information first. For example, some proposed that explaining the causes of the weather event was perhaps less important, and thus should be placed below information such as impacts and recommended actions. Many participants also noted the potential for images and visuals to enhance audience comprehension, particularly when dealing with technical and quantitative information. For instance, some participants proposed the use of images to help interpret metrics or geography, by displaying millimeters or centimeters of weather accumulation (e.g., snow or rain), or mapped photos.

4.2.3 Utility and Integration of Technology into Pre-Existing Workflows. Many participants noted the utility of a proposed writing support tool, such as those shown in Example 2(a) and 2(b), in aiding experienced and inexperienced risk communicators alike. Many of these participants, who had formal training in the field, acknowledged that their own work could benefit from a resource that *"checks tone or text to make sure that you're hitting all the key elements of crisis communication, or risk communication" (P9)* given the fast-paced nature of emergency management. Perhaps more importantly, interviewees frequently highlighted the benefits of a writing support tool for other risk communicators who may lack the relevant experience or training. Participants explained that, in the current emergency management landscape in Canada, community leaders, such as fire and police chiefs, are often "double hatted" in their role. On top of the responsibilities of their primary job, they are also tasked with communicating with the public during emergencies. For this group, a pop-up window (e.g., Example 2(a) in Figure 2) or suggested text to include (e.g., Example 2(b) in Figure 2), would potentially compensate for gaps in their knowledge and enable the creation of standardized messages. As one participant put it, when this group is working,

> *"...you want to give to those people the best tools possible to the people who are less experienced, you don't want to leave it to chance that they're gonna get the right the right message. So having a certain amount of standardized, but then the, your system will come in and will compensate for some of those issues that may or may not have..." (P1)*

Regardless of the technologies currently employed by their organizations, most participants proposed design recommendations for a risk communication tool that could be integrated into pre-existing workflows. To communicate risk, both with the public and within their organizations, participants relied on traditional mediums and platforms (e.g., texts, emails, intranet), new media (e.g., X, Instagram, Facebook) and government approved, organization-specific apps and software. These messages were often drafted using templates, on text editing software such as Microsoft Word or email. Given the mediums used to craft these messages, many participants thus suggested a software plug-in with the functionality of Examples 2(a) and 2(b), similar to typing assistants such as Grammarly or spellcheck. Some participants expanded on Example 2(b), and highlighted the benefits of having a program that acted as a repository or database of risk messages to choose relevant statements from. Consistent with work being done by Sutton and Kuligowski [118], others noted the value of a tool that would automatically generate a message that organized the content by the order of importance.

Despite the potential role of technology in improving risk communication, interviews surfaced challenges to their adoption in practice. In addition to concerns about cost, data privacy, and transparency in programming, most participants noted that, above all else, the tools should be user-friendly. For one participant, implementing technology was challenging due to capacity constraints and budgetary pressures, so that simple technology would be easier to implement and provide training for. Drawing on past experience, another interviewee noted that deploying technology with "double-hatted" communicators was historically challenging as *"they don't have time to use it, they're only going to log in when they need to" (P1)*. While emphasizing the importance of creating user-friendly technology, the same participant noted that Example 2(b) and the database/repository idea noted by another participant, were potentially well-suited for this population due to their simplicity.

4.2.4 Vulnerable Populations and Dialogic Risk Communication.

Many participants were initially critical of the risk messages that were customized for people with visual impairments (e.g., Example 3 in Figure 2). These interviewees disagreed with the content of these messages because, in their telling, the purpose of risk communication was to reach a wide audience in a timely fashion. Participants also noted that the creation of risk messages for one population would also require the creation of such communication for all diverse groups. However, many participants' evaluation of these messages was limited by current practices and technological capabilities. When provided with a hypothesized technology that could reach this population directly, in lieu of a general message, many agreed that there was a benefit to this type of specialized communication.

> *"So if you have the app, an app that does that, that's great. And you can have an app for each different kind of vulnerability that people have. And they can turn to it. And that's great... then yes, if you have an app that will give them a little bit more? Absolutely. I think that's a good idea… Yeah, if it goes directly to them, and they're the only ones that are getting that message, then yeah, all the information that they need is there. Definitely the tools. I think it's a good message." (P1)*

In contrast, some participants were much more enthusiastic about the customized risk messages at the outset. These participants, who noted personal connections to the visually impaired community, often critiqued current risk communication for being exclusionary and unable to meet the needs of more diverse groups. For these participants, the customized risk messages addressed the needs of an underserved population by providing more specific information for people with visual impairments, which could potentially improve their ability to undertake protective actions. In noting the merits of this example, one participant argued that all groups are deserving of better risk messaging.

> *"I like how specific it is. There's someone in my life with a visual impairment. And honestly, I was a bit taken aback by this because they are infrequently considered in these important communications. And so I think that's fantastic. And I'm appreciative of whomever led you down this path. I think it doesn't really address the challenge of it being like wordy, but it's nothing. It's adding value in a different way. It's not only communicating an affected population, it's communicating potential impacts, as well as things they can do to minimize those impacts. I think it's better to be transparent with people about what those things are, instead of them having to interpret them. I think I think the same could be true for a lot of vulnerable populations, and, you know, pretending to be I'm not one. And so I, I don't want to overstep, but in my experience, we we deal with people...nephrology patients...people who have who are hard of hearing… And I think the reality is that those people deserve really great risk communication as well. I'm not sure if you're consistently getting that." (P9)*

Among participants who consistently noted an appreciation for the customized risk messages, many highlighted the importance of dialogic approaches to risk communication. These participants acknowledged the unique vulnerabilities experienced by marginalized groups and people within their circle of care, and the inability for outside risk communication professionals to create messages that adequately addressed their needs. Interview subjects thus suggested that risk communication should *"come from these populations" (P9)* and noted the potential for co-creation workshops or focus groups to achieve this.

5 DISCUSSION

The results of the first and second study surfaced several important design considerations for the development of an authoring support tool for risk communication professionals, the role of artificial intelligence (AI) in achieving this, and the many ways that HCI methods can help inform the creation of more inclusive risk communication. Specifically, we highlight notable aspects of the design space and desirable features of an authoring support tool suggested by this research, which may help develop and tailor this tool to the needs of risk communicators. Within this context, we then consider the potential role of AI in risk communication and discuss some implications for its safe introduction into current practices. Finally, we highlight the challenges associated with crafting messages for vulnerable communities and underscore the instrumental role of HCI research and methods in fostering a more inclusive approach to risk communication.

5.1 Designing Tools for Risk Communicators

Our study suggests that the design of an authoring support tool for risk communication professionals should account for the complexities of this design space, including the presence of varying degrees of expertise amongst risk communicators themselves. Many of our participants noted that part of the potential utility of a risk communication writing support tool is its ability to help less experienced risk communicators working in the field. Seeing as previous work has found that authoring support tools may ease the speed of writing by allowing the quick adaptation of language [49], improve user confidence [93, 126] and potentially promote learning [126], it is possible that a writing assistant in this domain may reduce the mental workload of risk communicators while improving confidence in the messages they create. For "double hatted" risk communicators who lack formal training in the field, such a tool could also improve their understanding of the elements of persuasive risk messages by teaching through examples [126]. However, this divergence in expertise may render inexperienced risk communicators prone to automation bias – the undue deference given to automated systems by human actors [5] – as experience and lack of confidence in one's own abilities have been shown to increase reliance on these systems in field such as healthcare [37, 75, 104]. Given the flaws inherent to automated systems built from data, it may thus be preferable to have systems with humans in the loop in this safety critical domain. To help mitigate against automation bias, designers might consider features such as the level of on-screen detail [128], the type of display [104], and display prominence [11] when creating an authoring support tool for risk communicators, in addition to the comprehensibility of the underlying algorithm.

Furthermore, our findings indicate several additional features of an authoring support tool that can help risk communicators write messages that are consistent with message design theories – frameworks that guide the creation and structuring of effective risk communication – that emphasize structure, content and style [106, 117]. As it relates to content, many participants noted the benefits of a tool that functioned similar to Example 2(a) and determined if the relevant themes – hazards, impacts, recommended actions, location, timing, message source, and affected populations [15, 16, 43, 44] – were present in the message and suggested their inclusion. Additionally, an authoring tool could suggest sentences or phrases to include based on text from historical risk messages similar to Example 2(b). To evaluate message style – how designers use linguistic style to present information [106] – a feature of this writing assistant could be to highlight jargon and suggest alternative, simplified text. To enhance the structure of the message and its presentation, this tool could also assess the sequence in which information is presented and propose a reordering based on both importance and readability.

5.2 AI and Risk Communication

Despite the emergence of ML-based writing support tools and powerful large language models (LLMs) for natural language generation, our participants frequently expressed skepticism about the current utility of these tools to support risk messaging. As noted by Ogie et al. [87] in their review of uses of AI in disaster risk communication, the application of AI in this field has largely focused on: (1) predicting, monitoring and early warnings of disasters [1, 47, 54, 55, 76, 127] and (2) information extraction and classification for situational awareness [3, 8, 39, 90, 103, 129]. Experimental uses of AI to generate risk messages are still quite limited, and a number of our interviewees expressed dissatisfaction with the message simplified by the LLM, and aversion to any potential use of AI to create risk messaging. They explained this aversion, in part, by the fact that many of the risk messages in our sample were already flawed, so any risk messaging produced by AI trained on these messages would simply reproduce, or even exacerbate, their shortcomings. Moreover, the use of the LLM to simplify the language level did not address concerns about jargon in the risk messages further limiting the utility of this tool. Instead, this study suggests that more traditional applications of AI for tasks such as spelling and grammar support [24, 48, 60], tone detection [2, 56], and language translation [7, 23, 32, 51, 61] offer more immediate utility for risk messaging.

In addition to hesitance on the part of risk communicators we spoke with, audience trust in risk messaging is key, and a very well documented challenge in the risk communication literature [73]. This is in part because many forms of protective action, including evacuation, closing a small business, or staying home from work, can be costly to the individual who takes them [120]. Many people who live in hazard prone-areas have also received warnings in the past where the storms did not in fact occur, or were less damaging than models predicted [101, 109]. This is to some degree unavoidable due to the significant uncertainties involved in severe weather prediction. But they speak to a real challenge of building trust in the audience for risk messages that the successful introduction of AI into this setting will have to consider, and contend with. Prior research in risk communication and explainable AI may inform ways to introduce AI in risk communication practices. In both of these domains, transparency – either information about an AI system's decision-making process [34, 100] or the source of risk messages [79] – are thought to contribute to trust and confidence in the system. Specifying the ways that AI are used in creating risk messages and the source of risk information may thus be an avenue for establishing confidence in these systems, but should be considered alongside other relevant factors such as message length and complexity [117]. This suggests that further work to evaluate the potential impacts of the use of AI on public trust in risk messaging may help ensure successful adoption of writing support tools in this complex and high-stakes domain.

5.3 Co-creation and Participatory Design as Strategies for Supporting Diverse Audiences' Needs

Disasters do not impact all people equally, and many of our participants noted that vulnerable groups experience disproportionate effects of weather hazards. The study thus surfaced a central challenge in designing for vulnerable communities: the tension between making technologies and services accessible to a wide audience, while ensuring that vulnerable groups are not excluded. In the context of risk communication, this is complicated by the need to disseminate information quickly and effectively to audiences who have diverse needs, capacities, and relationships to risk. In most settings, the ability to send customized messages is limited, in part, by limitations in the design of underlying alert broadcasting systems, but also by a lack of knowledge of the relevant content to include in this type of messaging. For example, prior research in disaster risk communication has found that publicly available sources of risk information generally do not cater to people with disabilities [92], which can result in a lack of relevant assistance to these users [121]. In addition to relying on different technologies for seeking and receiving risk information [66], disabled and non-disabled users also diverge in the import given to particular details of the hazard, have different risk thresholds, and different temporalities for decision-making and therefore information needs [113].

When discussing the possible benefits of specialized risk messages that meets the needs of diverse groups, participants frequently appealed to participatory approaches to the design of risk communication products and strategies. As Soden et al. [112] note, HCI has long contributed to research in accessibility and is thus well positioned to consider how risk communication can better meet the needs of people with disabilities. In addition to extended collaborations with this user group [10], participatory approaches to risk communication, such as the co-creation workshops and focus groups suggested by interview participants, are a potential avenue to do so. Indeed, this study suggests that participatory design of risk information products may help ensure that risk messages are informed by more diverse forms of expertise, knowledge, and lived experiences. Participatory approaches are also consistent with deliberative or dialogic models of risk communication, which emphasize the value of two-way or multi-party dialogue between experts and the public [30, 131]. Such a dialogue would allow the public to express their concerns and contextualized knowledge about hazardous events, which authorities could account for and thus potentially improve the quality and implementation of decisions [30].

In addition, design research methods that focus less on the deficiencies of individuals or communities, and instead place their skills, abilities, capacities and resources at the center of design processes may be most productive for the creation of risk communication messaging systems. For example, in the design of interactive systems, Wobbrock et al. [124] argue for an ability-based design approach. This approach foregrounds what individuals *can* do, rather than what they *cannot* do, thereby transferring the responsibility of access and use from the users to the systems. Similarly, an assets-based design approach seeks to support the agency and autonomy of vulnerable groups by leveraging their assets, relationships, and capacities [125]. Given that efficacy is a well-known predictor of effective protective decision-making [73], such approaches should be considered when engaging community members in participatory design processes. Furthermore, assets or capacity-based approaches to risk communication messaging systems that cater to specific individuals and communities may potentially introduce a "curb cut effect" [14]. Namely, designing messages for vulnerable populations might also prove beneficial for the general public and other communities not explicitly targeted in the process, thereby improving the accessibility and efficacy of these messages.

5.4 Limitations

The present study may be subject to several limitations. First, after updating the string-matching algorithm based on repeated evaluations of 10 random subsamples to achieve an accuracy of 90% or greater, the algorithm was then tested on a sample of 50 previously unseen messages. This sample size was based on the observation that the performance of the algorithm plateaued between five and six iterations during the development phase. Although the repeated evaluations of random subsamples was intended to capture as much variation in the data as possible, this objective may have been tempered by the limited size of the test set. It is therefore possible that the test data lacked diversity and outliers. To enhance the robustness of the findings, future studies may consider testing the model on a larger test set to account for more variations in the data. Given the evolving nature of climate change and severe weather events, special attention should also be paid to updating the algorithm over time to reflect updated guidance and information.

Additionally, the simplified messages generated by ChatGPT, which were poorly received by a majority of the study participants, may have been a product of the prompt, and not the underlying technology, used to generate them. The research team attempted to address this limitation by testing multiple prompts, where the final prompt was chosen based on discussion and agreement on the part of the research team, although it is possible that a different prompt would have generated improved results. To address this, future research may consider prompt engineering in this domain and working with risk communicators to identify how best to use AI-based tools for this messaging.

6 CONCLUSION

Severe weather events are anticipated to increase in duration and frequency due to climate change [42, 91], rendering effective risk communication an integral component of climate adaptation policies. The present study employed string-matching and semi-structured interviews with EMs and risk communicators to evaluate the adoption of best practices in risk messages issued in Canada and to examine the design space of an authoring support tool for this group. This research thus offers insights into risk communication policy and practice, and provides guidance for the design of technologies to facilitate the creation of risk messages. Future work will consider steps, such as engaging in additional text pre-processing and extending the analysis to the remaining 18 weather events, to begin to operationalize the suggested tool. Furthermore, an evaluation of linguistic features, such as consistency, clarity, and conciseness, of the corpus would greatly supplement the current analysis and provide additional guidance on improving severe weather risk communication.

REFERENCES

[1] P Abhigna, S Jerritta, R Srinivasan, and V Rajendran. 2017. Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in Bay of Bengal. In *2017 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, Chennai, India, 1856–1860.

[2] Fazeel Abid, Muhammad Alam, Muhammad Yasir, and Chen Li. 2019. Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. *Future Generation Computer Systems* 95 (2019), 292–308.

[3] Flavia Sofia Acerbo and Claudio Rossi. 2017. Filtering informative tweets during emergencies: a machine learning approach. In *Proceedings of the First CoNEXT Workshop on ICT Tools for Emergency Networks and DisastEr Relief*. ACM, Incheon,Republic of Korea, 1–6.

[4] Alfred V Aho and Margaret J Corasick. 1975. Efficient string matching: an aid to bibliographic search. *Commun. ACM* 18, 6 (1975), 333–340.

[5] Saar Alon-Barkat and Madalina Busuioc. 2023. Human–AI interactions in public sector decision making:"automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory* 33, 1 (2023), 153–169.

[6] Izzat Alsmadi and Maryam Nuser. 2012. String matching evaluation methods for DNA comparison. *International Journal of Advanced Science and Technology* 47, 1 (2012), 13–32.

[7] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2016. Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473 [cs.CL]

[8] A Balamurugan and R Jeberson Retna Raj. 2006. Effective Analysis of Social Networks for Emergency Alert and Disaster Management Using Android. *Effective Analysis of Social Networks for Emergency Alert and Disaster Management Using Android* 11, 17 (2006), 10291–10296.

[9] Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics* 34, 1 (2008), 1–34.

[10] Cynthia L Bennett and Daniela K Rosner. 2019. The promise of empathy: Design, disability, and knowing the" other". In *Proceedings of the 2019 CHI conference on human factors in computing systems*. ACM, Glasgow, Scotland UK, 1–13.

[11] Eta S Berner, Richard S Maisiak, Gustavo R Heudebert, and K Randall Young Jr. 2003. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. In *Amia annual symposium proceedings*, Vol. 2003. American Medical Informatics Association, USA, 76.

[12] Melissa Bica, Julie L Demuth, James E Dykes, and Leysia Palen. 2019. Communicating hurricane risks: Multi-method examination of risk imagery diffusion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow,Scotland UK, 1–13.

[13] Melissa Bica, Joy Weinberg, and Leysia Palen. 2020. Achieving accuracy through ambiguity: The interactivity of risk communication in severe weather events. *Computer Supported Cooperative Work (CSCW)* 29, 5 (2020), 587–623.

[14] Angela Glover Blackwell. 2017. The curb-cut effect. *Stanford Social Innovation Review* 15, 1 (2017), 28–33.

[15] Noel T Brewer, Gretchen B Chapman, Frederick X Gibbons, Meg Gerrard, Kevin D McCaul, and Neil D Weinstein. 2007. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health psychology* 26, 2 (2007), 136.

[16] Philip Bubeck, Wouter JW Botzen, Heidi Kreibich, and Jeroen CJH Aerts. 2013. Detailed insights into the influence of flood-coping appraisals on mitigation behaviour. *Global environmental change* 23, 5 (2013), 1327–1338.

[17] Kate Burningham, Jane Fielding, and Diana Thrush. 2008. 'It'll never happen to me': understanding public awareness of local flood risk. *Disasters* 32, 2 (2008), 216–238.

[18] Nicoletta Calzolari. 1984. Detecting patterns in a lexical data base. In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*. ACL, Stanford, California, USA, 170–173.

[19] Tuhin Chakrabarty, Xurui Zhang, Smaranda Muresan, and Nanyun Peng. 2021. MERMAID: Metaphor Generation with Symbolism and Discriminative Decoding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 4250–4261. https://doi.org/10.18653/v1/2021.naacl-main.336

[20] Christian Charras and Thierry Lecroq. 2004. *Handbook of exact string matching algorithms*. King's College, London, UK.

[21] Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, and Zhifeng Chen. 2019. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery I& Data Mining*. ACM, Anchorage,AK,USA, 2287–2295.

[22] Lok-Lam Cheng, David W Cheung, and Siu-Ming Yiu. 2003. Approximate string matching in DNA sequences. In *Eighth International Conference on Database Systems for Advanced Applications, 2003.(DASFAA 2003). Proceedings*. IEEE, Kyoto, Japan, 303–310.

[23] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, Dekai Wu, Marine Carpuat, Xavier Carreras, and Eva Maria Vecchi (Eds.). Association for Computational Linguistics, Doha, Qatar, 103–111. https://doi.org/10.3115/v1/W14-4012

[24] Shamil Chollampatt, Kaveh Taghipour, and Hwee Tou Ng. 2016. Neural Network Translation Models for Grammatical Error Correction. arXiv:1606.00189 [cs.CL]

[25] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 222, 2015 (2015), 248.

[26] Camille Cobb, Ted McCarthy, Annuska Perkins, Ankitha Bharadwaj, Jared Comis, Brian Do, and Kate Starbird. 2014. Designing for the deluge: understanding & supporting the distributed, collaborative work of crisis volunteers. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, Baltimore, MD, USA, 888–899.

[27] Andy Coenen, Luke Davis, Daphne Ippolito, Emily Reif, and Ann Yuan. 2021. Wordcraft: a Human-AI Collaborative Editor for Story Writing. arXiv:2107.07430 [cs.CL]

[28] Joop De Boer, WJ Wouter Botzen, and Teun Terpstra. 2015. More than fear induction: Toward an understanding of people's motivation to be well-prepared for emergencies in flood-prone areas. *Risk analysis* 35, 3 (2015), 518–535.

[29] Natascha De Hoog, Wolfgang Stroebe, and John BF De Wit. 2007. The impact of vulnerability to and severity of a health risk on processing and acceptance of fear-arousing communications: A meta-analysis. *Review of General Psychology* 11, 3 (2007), 258–285.

[30] David Demeritt and Sebastien Nobert. 2014. Models of best practice in flood risk communication and management. *Environmental Hazards* 13, 4 (2014), 313–328.

[31] Julie L Demuth, Rebecca E Morss, Betty Hearn Morrow, and Jeffrey K Lazo. 2012. Creation and communication of hurricane risk information. *Bulletin of the American Meteorological Society* 93, 8 (2012), 1133–1145.

[32] Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul. 2014. Fast and robust neural network joint models for statistical machine translation. In *proceedings of the 52nd annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL, Baltimore, Maryland, USA, 1370–1380.

[33] Amy Rosamund Donovan, Maud Anais Heloise Borie, and Sophie Elizabeth Blackburn. 2019. Changing the paradigm for risk communication: integrating sciences to understand cultures.

[34] Finale Doshi-Velez and Been Kim. 2017. Towards A Rigorous Science of Interpretable Machine Learning. arXiv:1702.08608 [stat.ML]

[35] Kirstin Dow and Susan L Cutter. 1998. Crying wolf: Repeat responses to hurricane evacuation orders. *Coastal Management* 26 (1998), 237–252.

[36] Thomas E. Drabek and John S. Stephenson Iii. 1971. When Disaster Strikes. *Journal of applied social psychology* 1, 2 (1971), 187–203. https://doi.org/10.1111/j.1559-1816.1971.tb00362.x ArticleID:JASP187.

[37] Stephan Dreiseitl and Michael Binder. 2005. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artificial intelligence in medicine* 33, 1 (2005), 25–30.

[38] Xiaodan Du, Nicholas Kolkin, Greg Shakhnarovich, and Anand Bhattad. 2023. Generative Models: What do they know? Do they know things? Let's find out!

arXiv:2311.17137 [cs.CV]

[39] Lars Döhling and Ulf Leser. 2011. EquatorNLP: Pattern-based Information Extraction for Disaster Response. *CEUR Workshop Proceedings* 798 (01 2011).

[40] Samira Ellouze, Maher Jaoua, and Lamia Hadrich Belguith. 2016. Automatic evaluation of a summary's linguistic quality. In *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings 21*. Springer, Salford, UK, 392–400.

[41] James N Farr, James J Jenkins, and Donald G Paterson. 1951. Simplification of Flesch reading ease formula. *Journal of applied psychology* 35, 5 (1951), 333.

[42] Christopher B Field, Vicente Barros, Thomas F Stocker, and Qin Dahe. 2012. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, England, Book 5, 291–338.

[43] Laura Fischer, David Huntsman, Ginger Orton, and Jeannette Sutton. 2023. You have to send the right message: examining the influence of protective action guidance on message perception outcomes across prior hazard warning experience to three hazards. *Weather, climate, and society* 15, 2 (2023), 307–326.

[44] Baruch Fischhoff. 1989. Helping the public make health risk decisions. In *Effective risk communication: The role and responsibility of government and nongovernment organizations*. Springer, Boston, MA, USA, 111–116.

[45] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[46] United Nations Regional Information Centre for Western Europe. 2022. 2021 floods: UN researchers aim to better prepare for climate risks. https://unric.org/en/2021-floods-un-researchers-aim-to-better-prepare-for-climate-risks/

[47] Gustavo Furquim, Geraldo PR Filho, Roozbeh Jalali, Gustavo Pessin, Richard W Pazzi, and Jó Ueyama. 2018. How to improve fault tolerance in disaster predictions: a case study about flash floods using IoT, ML and real data. *Sensors* 18, 3 (2018), 907.

[48] Michael Gamon, Jianfeng Gao, Chris Brockett, Alexandre Klementiev, William B Dolan, Dmitriy Belenko, and Lucy Vanderwende. 2008. Using contextual speller techniques and language modeling for ESL error correction. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*. ACL, Hyderabad, India.

[49] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Designing Interactive Systems Conference*. ACM, Australia, 1002–1019.

[50] Katy Ilonka Gero, Vivian Liu, Sarah Huang, Jennifer Lee, and Lydia B Chilton. 2021. What Makes Tweetorials Tick: How Experts Communicate Complex Topics on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.

[51] Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III. 2014. Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. In *Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP)*. ACL, Doha,Qatar, 1342–1352.

[52] Xinning Gui, Yubo Kou, Kathleen Pine, Elisa Ladaw, Harold Kim, Eli Suzuki-Gill, and Yunan Chen. 2018. Multidimensional risk communication: public discourse on risks during an emerging epidemic. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal, QC, Canada, 1–14.

[53] Xinning Gui, Yubo Kou, Kathleen H Pine, and Yunan Chen. 2017. Managing uncertainty: using social media for risk assessment during a public health crisis. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM, Denver, CO, USA, 4520–4533.

[54] Iwan K Hadihardaja, Dian Indrawati, Yadi Suryadi, and Neill S Griggs. 2012. Decision support system for predicting flood characteristics based on database modelling development (case study: Upper Citarum, West Java, Indonesia). *WIT Transactions on Ecology and the Environment* 167 (2012), 375–385.

[55] Iwan K Hadihardaja, Hamzah Latief, and Iyan E Mulia. 2011. Decision support system for predicting tsunami characteristics along coastline areas based on database modelling development. *Journal of Hydroinformatics* 13, 1 (2011), 96–109.

[56] Abdalraouf Hassan and Ausif Mahmood. 2017. Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on control, automation and robotics (ICCAR)*. IEEE, Nagoya, Japan, 705–710.

[57] Changyang He, Yue Deng, Wenjie Yang, and Bo Li. 2022. " Help! Can You Hear Me?": Understanding How Help-Seeking Posts are Overwhelmed on Social Media during a Natural Disaster. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–25.

[58] Jen Henderson, Erik R Nielsen, Gregory R Herman, and Russ S Schumacher. 2020. A hazard multiple: Overlapping Tornado and flash flood warnings in a National Weather Service Forecast Office in the Southeastern United States. *Weather and Forecasting* 35, 4 (2020), 1459–1481.

[59] Daniel Henstra. 2012. Toward the climate-resilient city: extreme weather and urban climate adaptation policies in two Canadian provinces. *Journal of Comparative Policy Analysis: Research and Practice* 14, 2 (2012), 175–194.

[60] Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and Kenneth Heafield. 2018. Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 595–606. https://doi.org/10.18653/v1/N18-1055

[61] Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. 2017. Neural Machine Translation in Linear Time. arXiv:1610.10099 [cs.CL]

[62] Krishna Prakash Kalyanathaya, D Akila, and G Suseendren. 2019. A fuzzy approach to approximate string matching for text retrieval in NLP. *J. Comput. Inf. Syst. USA* 15, 3 (2019), 26–32.

[63] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, and Peter Young. 2016. Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, SanFrancisco,CA,USA, 955–964.

[64] Yubo Kou, Xinning Gui, Yunan Chen, and Kathleen Pine. 2017. Conspiracy talk on social media: collective sensemaking during a public health crisis. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–21.

[65] Thomas Kox. 2015. Criteria affecting people's decision to take protective measures during winter storm XAVER on 5 December 2013. In *The 12th International Conference on Information Systems for Crisis Response and Management*. ISCRAM, Krystiansand, Norway.

[66] Heather Lazrus, Betty H Morrow, Rebecca E Morss, and Jeffrey K Lazo. 2012. Vulnerability beyond stereotypes: Context and agency in hurricane risk communication. *Weather, Climate, and Society* 4, 2 (2012), 103–109.

[67] Alex Leavitt and Joshua A Clark. 2014. Upvoting hurricane Sandy: event-based news production processes on a social news site. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, Toronto, ON, Canada, 1495–1504.

[68] Alex Leavitt and John J Robinson. 2017. Portland, OR, USA. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. ACM, Portland, OR, USA, 1246–1261.

[69] Xuyang Li, Antara Bahursettiwar, and Marina Kogan. 2021. Hello? Is There Anybody in There? Analysis of Factors Promoting Response From Authoritative Sources in Crisis. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[70] Ziheng Lin, Chang Liu, Hwee Tou Ng, and Min-Yen Kan. 2012. Combining coherence models and machine translation evaluation metrics for summarization evaluation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL, Jeju,Republic of Korea, 1006–1014.

[71] Michael K Lindell and Ronald W Perry. 2012. The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis: An International Journal* 32, 4 (2012), 616–632.

[72] Thomas Ludwig, Christian Reuter, Tim Siebigteroth, and Volkmar Pipek. 2015. CrowdMonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul, Republic of Korea., 4083–4092.

[73] Elaina MacIntyre, Sanjay Khanna, Anthea Darychuk, Ray Copes, and Brian Schwartz. 2019. Evidence synthesis Evaluating risk communication during extreme weather and climate change: a scoping review. *Health promotion and chronic disease prevention in Canada: research, policy and practice* 39, 4 (2019), 142.

[74] Nandita S Mani, Terri Ottosen, Megan Fratta, and Fei Yu. 2021. A health literacy analysis of the consumer-oriented COVID-19 information produced by ten state health departments. *Journal of the Medical Library Association: JMLA* 109, 3 (2021), 422.

[75] Katharina Marten, Tobias Seyfarth, Florian Auer, Edzard Wiener, Andreas Grillhösl, Silvia Obenauer, Ernst J Rummeny, and Christoph Engelke. 2004. Computer-assisted detection of pulmonary nodules: performance evaluation of an expert knowledge-based detection system in consensus reading with experienced and inexperienced chest radiologists. *European radiology* 14 (2004), 1930–1938.

[76] Hajime Mase, Tomohiro Yasuda, and Nobuhito Mori. 2011. Real-time prediction of tsunami magnitudes in Osaka Bay, Japan, using an artificial neural network. *Journal of waterway, port, coastal, and ocean engineering* 137, 5 (2011), 263–268.

[77] Christopher B Mayhorn and Anne Collins McLaughlin. 2014. Warning the world of extreme events: A global perspective on risk communication for natural and technological disaster. *Safety science* 61 (2014), 43–50.

[78] Dennis S Mileti and Lori Peek. 2000. The social psychology of public response to warnings of a nuclear power plant accident. *Journal of hazardous materials* 75, 2-3 (2000), 181–194.

[79] Dennis S Mileti and John H Sorensen. 1990. *Communication of emergency public warnings: A social science perspective and state-of-the-art assessment*. Technical Report. Oak Ridge National Lab., TN (USA).

[80] BH Morrow and JK Lazo. 2015. Effective tropical cyclone forecast and warning communication: Recent social science contributions. *Tropical Cyclone Research and Review* 4, 1 (2015), 38–48.

[81] Betty H Morrow, Jeffrey K Lazo, Jamie Rhome, and Jesse Feyen. 2015. Improving storm surge risk communication: Stakeholder perspectives. *Bulletin of the American Meteorological Society* 96, 1 (2015), 35–48.

[82] Rebecca E Morss, Cara L Cuite, Julie L Demuth, William K Hallman, and Rachael L Shwom. 2018. Is storm surge scary? The influence of hazard, impact, and fear-based messages and individual differences on responses to hurricane risks in the USA. *International journal of disaster risk reduction* 30 (2018), 44–58.

[83] Laura Myers. 2016. NOAA Weather Information and Dissemination All Hazards Stakeholder Project: Phase 1.

[84] Erik R Nielsen, Gregory R Herman, Robert C Tournay, John M Peters, and Russ S Schumacher. 2015. Double impact: When both tornadoes and flash floods threaten the same place at the same time. *Weather and Forecasting* 30, 6 (2015), 1673–1693.

[85] Wendy Norris, Amy Voida, and Stephen Voida. 2022. People Talk in Stories. Responders Talk in Data: A Framework for Temporal Sensemaking in Time-and Safety-critical Work. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–23.

[86] Government of Canada. 2025. Climate change adaptation in Canada. https://www.nrcan.gc.ca/climate-change/what-adaptation/10025

[87] Robert Ighodaro Ogie, Juan Castilla Rho, and Rodney J Clarke. 2018. Artificial intelligence in disaster risk communication: A systematic literature review. In *2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. IEEE, Sendai, Japan, 1–8.

[88] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 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*. ACM, Vancouver, BC, Canada, 994–1009.

[89] Leysia Palen and Kenneth M Anderson. 2016. Crisis informatics—New data for extraordinary times. *Science* 353, 6296 (2016), 224–225.

[90] Neha Pandey and S Natarajan. 2016. How social media can contribute during disaster events? Case study of Chennai floods 2015. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, Jaipur, India, 1352–1356.

[91] Mathaios Panteli and Pierluigi Mancarella. 2015. Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies. *Electric Power Systems Research* 127 (2015), 259–270.

[92] Eun-Seon Park, DK Yoon, and Yeon-Woo Choi. 2019. Leave no one behind: Experiences of persons with disability after the 2017 Pohang earthquake in South Korea. *International Journal of Disaster Risk Reduction* 40 (2019), 101261.

[93] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. 2020. Exploring the effects of technological writing assistance for support providers in online mental health community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, HI, USA, 1–15.

[94] Ronald W. Perry, Michael K. Lindell, and Marjorie R. Greene. 1982. Threat Perception and Public Response to Volcano Hazard. *The Journal of social psychology* 116, 2 (1982), 199–204. https://doi.org/10.1080/00224545.1982.9922771

[95] Gjalt-Jorn Ygram Peters, Robert AC Ruiter, and Gerjo Kok. 2013. Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health psychology review* 7, sup1 (2013), S8–S31.

[96] Kathleen H Pine, Myeong Lee, Samantha A Whitman, Yunan Chen, and Kathryn Henne. 2021. Making sense of risk information amidst uncertainty: individuals' perceived risks associated with the COVID-19 pandemic. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–15.

[97] Emily Pitler, Annie Louis, and Ani Nenkova. 2010. Automatic Evaluation of Linguistic Quality in Multi-Document Summarization. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Jan Hajič, Sandra Carberry, Stephen Clark, and Joakim Nivre (Eds.). Association for Computational Linguistics, Uppsala, Sweden, 544–554. https://aclanthology.org/P10-1056

[98] Yan Qu, Chen Huang, Pengyi Zhang, and Jun Zhang. 2011. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, Hangzhou, China, 25–34.

[99] Janice Redish. 2000. Readability formulas have even more limitations than Klare discusses. *ACM Journal of Computer Documentation (JCD)* 24, 3 (2000), 132–137.

[100] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, SanFrancisco,CA,USA, 1135–1144.

[101] Joseph T Ripberger, Carol L Silva, Hank C Jenkins-Smith, Deven E Carlson, Mark James, and Kerry G Herron. 2015. False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk analysis* 35, 1 (2015), 44–56.

[102] Melissa Roemmele. 2016. Writing stories with help from recurrent neural networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI Press, Phoenix, Arizona, 4311–4312.

[103] Tatsuhiro Sakai, Keiichi Tamura, Hajime Kitakami, and Toshiyuki Takezawa. 2017. Photo image classification using pre-trained deep network for density-based spatiotemporal analysis system. In *2017 IEEE 10th international workshop on computational intelligence and applications (IWCIA)*. IEEE, Hiroshima, Japan, 207–212.

[104] Nadine B Sarter and Beth Schroeder. 2001. Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human factors* 43, 4 (2001), 573–583.

[105] Katja Schulze, Daniel F Lorenz, Bettina Wenzel, and Martin Voss. 2015. Disaster Myths and their Relevance for Warning Systems. In *ISCRAM*. ISCRAM, Krystiansand, Norway.

[106] Lijiang Shen and Elisabeth Bigsby. 2013. *The effects of message features*. Sage Los Angeles, CA, Los Angeles, CA, USA, 20–35.

[107] Irina Shklovski, Leysia Palen, and Jeannette Sutton. 2008. Finding community through information and communication technology in disaster response. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, San Diego, California, USA, 127–136.

[108] Irina Shklovski, Leysia Palen, and Jeannette Sutton. 2008. Finding community through information and communication technology in disaster response. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, San Diego, California, USA, 127–136.

[109] Kevin M Simmons and Daniel Sutter. 2009. False alarms, tornado warnings, and tornado casualties. *Weather, Climate, and Society* 1, 1 (2009), 38–53.

[110] Nimisha Singla and Deepak Garg. 2012. String matching algorithms and their applicability in various applications. *International journal of soft computing and engineering* 1, 6 (2012), 218–222.

[111] Marvin So, Jessica L Franks, Robyn A Cree, and Rebecca T Leeb. 2020. An evaluation of the literacy demands of online natural disaster preparedness materials for families. *Disaster medicine and public health preparedness* 14, 4 (2020), 449–458.

[112] Robert Soden, Lydia Chilton, Scott Miles, Rebecca Bicksler, Kaira Ray Villanueva, and Melissa Bica. 2022. Insights and Opportunities for HCI Research into Hurricane Risk Communication. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans, LA, USA, 1–13.

[113] Patric R Spence, Ken Lachlan, Jennifer M Burke, and Matthew W Seeger. 2007. Media use and information needs of the disabled during a natural disaster. *Journal of health care for the poor and underserved* 18, 2 (2007), 394–404.

[114] Kate Starbird. 2013. Delivering patients to sacré coeur: collective intelligence in digital volunteer communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris, France, 801–810.

[115] Kate Starbird and Leysia Palen. 2011. "Voluntweeters" self-organizing by digital volunteers in times of crisis. In *Proceedings of the SIGCHI conference on human factors in computing systems*. publisher, Vancouver, BC, Canada, 1071–1080.

[116] Kate Starbird and Leysia Palen. 2013. Working and sustaining the virtual" Disaster Desk". In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, San Antonio, Texas, USA, 491–502.

[117] Jeannette Sutton, Laura Fischer, and Michele M Wood. 2021. Tornado warning guidance and graphics: Implications of the inclusion of protective action information on perceptions and efficacy. *Weather, climate, and society* 13, 4 (2021), 1003–1014.

[118] Jeannette Sutton and Erica D Kuligowski. 2019. Alerts and warnings on short messaging channels: Guidance from an expert panel process. *Natural Hazards Review* 20, 2 (2019), 04019002.

[119] Jorma Tarhio and Hannu Peltola. 1997. String matching in the DNA alphabet. *Software: Practice and Experience* 27, 7 (1997), 851–861.

[120] Teun Terpstra and Michael K Lindell. 2013. Citizens' perceptions of flood hazard adjustments: an application of the protective action decision model. *Environment and Behavior* 45, 8 (2013), 993–1018.

[121] Marieke Van Willigen, Terri Edwards, Bob Edwards, and Shawn Hessee. 2002. Riding out the storm: Experiences of the physically disabled during Hurricanes Bonnie, Dennis, and Floyd. *Natural Hazards Review* 3, 3 (2002), 98–106.

[122] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, Atlanta, GA, USA, 1079–1088.

[123] Gisela Wachinger, Ortwin Renn, Chloe Begg, and Christian Kuhlicke. 2013. The Risk Perception Paradox-Implications for Governance and Communication of Natural Hazards. *Risk analysis* 33, 6 (2013), 1049–1065. https://doi.org/10.1111/j.1539-6924.2012.01942.x

[124] Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)* 3, 3 (2011), 1–27.

[125] Marisol Wong-Villacres, Carl DiSalvo, Neha Kumar, and Betsy DiSalvo. 2020. Culture in action: Unpacking capacities to inform assets-based design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, HI, USA, 1–14.

[126] Shaomei Wu, Lindsay Reynolds, Xian Li, and Francisco Guzmán. 2019. Design and evaluation of a social media writing support tool for people with dyslexia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow, Scotland UK, 1–14.
[127] Jia Xichun, Wang Ruilan, Dai Hao, Zhang Wei, Li Zhiwei, and Cong Peitong. 2017. A neural network method for risk assessment and real-time early warning of mountain flood geological disaster. In *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. IEEE, Xiamen, China, 540–544.
[128] Michelle Yeh and Christopher D Wickens. 2001. Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors* 43, 3 (2001), 355–365.
[129] Jie Yin, Andrew Lampert, Mark Cameron, Bella Robinson, and Robert Power. 2012. Using social media to enhance emergency situation awareness. *IEEE intelligent systems* 27, 06 (2012), 52–59.
[130] Himanshu Zade, Kushal Shah, Vaibhavi Rangarajan, Priyanka Kshirsagar, Muhammad Imran, and Kate Starbird. 2018. From situational awareness to actionability: Towards improving the utility of social media data for crisis response. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–18.
[131] Joseph Árvai. 2014. The end of risk communication as we know it. *Journal of Risk Research* 17 (2014), 1245–1249.

Appendix A

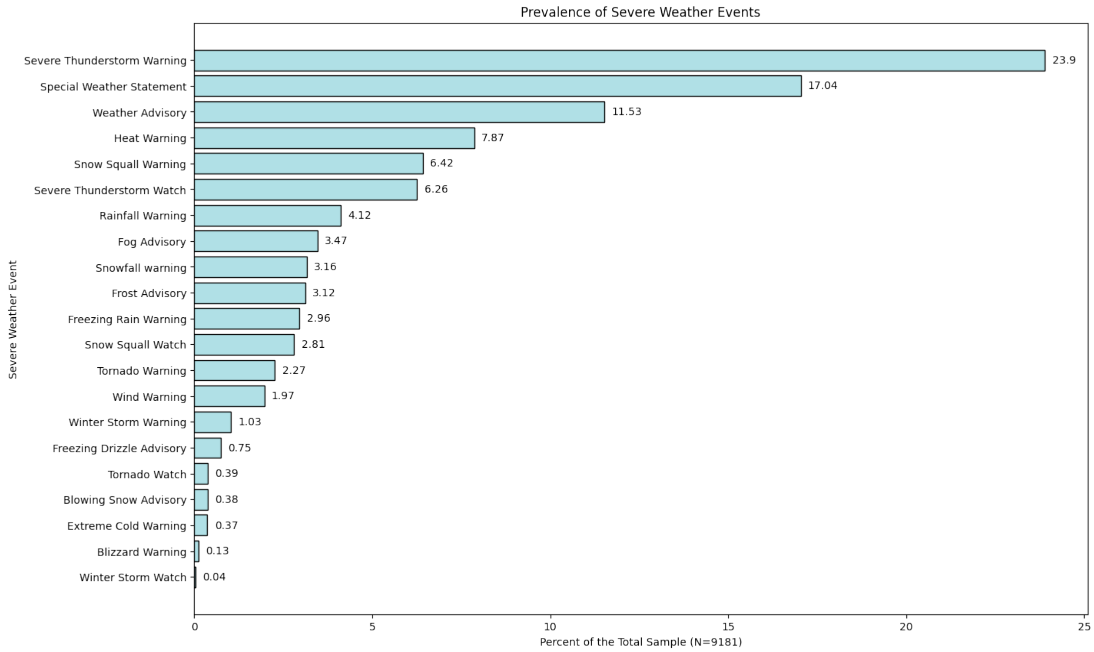

Figure 4: Prevalence of Severe Weather Events in Sample between June 2020 and December 2021

Table 3: Interpreting the Flesch Reading Ease Score[a]

Score	Difficulty
90 - 100	Very Easy
80 - 89	Easy
70 - 79	Fairly Easy
60 - 69	Standard
50 - 59	Fairly Difficult
30 - 49	Difficult
00 - 29	Very Confusing

[a]Source: TextStat 0.7.3
(https://pypi.org/project/textstat/)

Table 4: Best Practices in Risk Communication - Themes

Theme	Definition
Hazards	Health and/or safety hazards associated with the event
Location	Areas affected by the hazardous event
Timing	Information related to time such as onset, duration, and end time
Affected Populations	Groups vulnerable to hazard impacts
Impacts	Impacts of the hazardous event
Recommended Actions	Any actions or guidance to the public to mitigate the effects of the hazardous event