Building a Domain Agnostic Framework for Efficient and Effective Risk Communication Messages

Ross Gore, Barry Ezell, Christopher J. Lynch, Jessica O'Brien, Virginia Zamponi, Erik Jensen

Old Dominion University
Norfolk, VA

[rgore, bezell, cjlynch, jcordner, vzamponi, ejens005]@odu.edu

Ann Marie Reinhold, Clemente Izurieta, Madison Munro, Elizabeth Shanahan

Montana State University
Bozeman, MT

[reinhold, clemente.izurieta, shanahan]@montana.edu
madison.munro@student.montana.edu

ABSTRACT

Across hazard domains, traditional risk communication messages convey scientific information about risk using probability and certainty language. Such risk communication often fails to engage target audiences and is ineffective as a result. Our team is building a framework for developing risk communication messages that applies across hazard domains. The framework provides functionality to help practitioners create messages that are effective and can be created rapidly. To meet this objective, we are employing large language models to operationalize best practices in risk communication research. These practices include the creation of risk messages with highly customized framing that incorporate key tenants of the Narrative Policy Framework (NPF). We are testing messages created using this approach across three distinct hazard types: natural disasters, cyber security, and homeland security. The goal of this testing is to demonstrate how, regardless of hazard domain, a risk communication message can be shaped to engage its target audience. In this paper, we present ongoing research and future directions that build towards the goal of effectively engaging target audiences and motivating them towards protective actions.

ABOUT THE AUTHORS

Dr. Ross Gore is a Research Associate Professor at the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University. His current work focuses on data science and predictive analytics.
Dr. Ann Marie Reinhold is an Assistant Professor in the Gianforte School of Computing at Montana State University. She specializes in the development and application of computational methods to understand the mechanisms underpinning pressing environmental, societal, and cybersecurity problems.
Dr. Barry Ezell leads the Computational Modeling, Analysis and Simulation Enterprise for VMASC at Old Dominion University. His work combines decision analysis, simulation, risk management, and data analytics.
Dr. Christopher J. Lynch is a Research Assistant Professor and Certified Cloud Practitioner at VMASC at Old Dominion University. His work focuses on developing new methods for analyzing simulation results.
Dr. Clemente Izurieta is a Professor in the Gianforte School of Computing at Montana State University. His research interests include empirical software engineering, design and architecture of software systems, design patterns, the measurement of software quality and cybersecurity.
Jessica O'Brien is an Applied Research Cloud Architect at VMASC at Old Dominion University. Her research interests lie at the intersection of cloud infrastructure and cybersecurity.
Virginia Zamponi is a Simulation Developer at VMASC at Old Dominion University. Her research is focused on understanding interactions between humans and the built environment through agent-based modeling.
Madison Munro is a first year MS in Computer Science student at Montana State. She is passionate about using data science as a tool to solve human-based problems in cybersecurity and beyond.
Erik Jensen is a graduate student in Electrical and Computer Engineering at Old Dominion University. His research interests include validating, evaluating, and understanding the limits of generative artificial intelligence.
Dr. Elizabeth A. Shanahan is a full Professor in Political Science and Associate Vice President for Research Development at Montana State University. She is an expert in narrative risk communication and co-architect of the Narrative Policy Framework. In addition, she is a fellow of the National Academy of Public Administration.

Building a Domain Agnostic Framework for Efficient and Effective Risk Communication Messages

Ross Gore, Barry Ezell, Christopher Lynch, Jessica O'Brien, Virginia Zamponi, Erik Jensen

Old Dominion University
Norfolk, VA

[rgore, bezell, cjlynch, jcordner, vzamponi, ejens005]@odu.edu

Ann Marie Reinhold, Clemente Izurieta, Madison Munro, Elizabeth Shanahan

Montana State University
Bozeman, MT

[reinhold, clemente.izurieta, shanahan]@montana.edu
madison.munro@student.montana.edu

INTRODUCTION

Conventional science messaging efforts for hazards, regardless of domain, are ineffective at inducing preparedness (D Jones & Anderson Crow, 2017; Fischhoff & Scheufele, 2013) . One explanation is that scientists and the public do not share a common language to describe risk (Reinhold, Gore, et al., 2023). Scientists use messages that feature probability, uncertainty, frequency, and magnitude, whereas people see themselves in stories, with characters, plotlines, settings, and personal details (Reinhold, Raile, et al., 2023; Shanahan et al., 2019). Risk communication studies find that narrative language is more effective than science language at convincing individuals to adopt protective actions (De Wit et al., 2008; Shen et al., 2015). However, the role and efficacy of personal information interwoven into narrative details has not been studied. Our work seeks to demonstrate how specific narrative details can improve the efficacy of risk communication efforts across hazard domains.

Recent research has examined the inadequacies of conventional risk communication messaging. This includes messages related to: (1) data displays (Dransch et al., 2010; Cao et al., 2017), (2) trust in science (Paton et al., 2008; Pidgeon et al., 2003), and (3) risk perception (Wachinger et al., 2013). Findings suggest that employing narratives, especially ones featuring personal details about the recipients, can change how audiences perceive and adopt directives from risk communication messages that they are presented.

The Narrative Policy Framework (NPF) is a successful application of narratives to risk communications messages (Shanahan et al., 2019). The NPF conceptualizes a narrative as a structure that includes characters, plot, and setting. The characters reflect human and non-human entities engaging in the action of the story (Shanahan et al., 2018). The stories take place in a setting where a hero's actions follow a plot to save a victim from a villain. In the NPF: (1) heroes actively tackle the problem, (2) victims passively suffer fear or harm from a problem, and (3) villains create the problem (Shanahan et al., 2018; Stone, 2012). Researchers have consistently found that narratives featuring a hero are effective at getting individuals to adopt protective actions (Jones, 2014; D Jones & Anderson Crow, 2017; Shanahan et al., 2019; Reinhold, Raile, et al., 2023).

Despite promising foundations and empirical results, employing the NPF is risk communication has been challenging because no automated protocol exists to guide the construction of risk communication messages. The studies employing these messages rely on large amounts of manual effort resulting in a process that is difficult to repeat and replicate. Here, we apply large language models (LLMs) to construct risk communication messages in the NPF in a repeatable, and more automated manner. The application of LLMs to personal data about message recipients creates a tailored and individualized message for the recipient. We hypothesize this individualized message will result in improved risk communication efficacy. Here, we refer to risk communication efficacy as convincing individuals to adopt protective actions against a given hazard (Reinhold et al., 2024). Next, we describe our current study in more detail. Then we present our methodology and describe our evaluation. Finally, we discuss the hypothesized results and predicted effects.

CURRENT STUDY

The goal of our work is to create individualized messages using NPF. To the extent possible we aim to do this in an automated manner. To meet these objectives, we are employing large language models using structured narrative prompts and zero-shot or few-shot learning (Brown et al., 2020; Lynch et al., 2023) to personalize NPF risk communication messages using demographic, interest, and Schwartz value information about individuals (Schwartz,

2012). To determine if our approach is applicable across the spectrum of hazard domains, we are testing messages using the Domain Agnostic Risk Communication (DARC) Framework (Reinhold et al., 2024). We are applying the DARC Framework to three distinct hazard types: natural disasters (i.e flooding), cyber security (i.e. phishing), and homeland security (i.e. active shooter incidents) (Reinhold et al., 2024). In what follows we describe each of these components in more detail. First, we provide an overview of each of the hazards. Next, we describe in detail how conventional science and generic NPF risk communications messages are constructed. Finally, we highlight how the data about the message recipient can be applied in a large language model to create an individualized NPF risk communication message.

Hazards

Flooding on the Elizabeth River

Flooding is the most common natural disaster in the U.S. It can cause power outages, disrupt transportation, damage buildings, and create landslides (FEMA, 2024). The Elizabeth River in Hampton Roads, VA poses flood threats to certain communities and businesses. Figure 1 shows that where the threats are greatest for communities and businesses. This occurs in areas identified as 100-year flood zones. An area with a 1-percent annual chance flood is also referred to as the base flood or 100-year flood. In the figure these areas are designated in teal as Zone AE and in tan as Zone AH (Mitchell, 2021; Hunt et al., 2019; Skaggs & Seltzer, 2002). Given this threat we construct risk communications for flooding on the Elizabeth River in our study to reflect the domain of natural hazards.

Figure 1: Some of the areas in Hampton Roads, VA with a 1-percent annual chance flood (i.e. 100-year flood zone). These areas are designated in teal as Zone AE and in tan as Zone AH.

Cyber Phishing

Phishing is a social engineering technique that, through various methodologies, aims to influence the target of the attack to reveal personal information, such as an email address, username, password, or financial information. This information is then used by the attacker to the detriment of the victim (Stavroulakis & Stamp, 2010). The term phishing is derived from the word “fishing”, spelt using what is commonly known as Haxor or L33T Speak. The logic of this terminology is that an attacker uses “bait” to lure the victim and then “fishes” for the personal information they want to steal. The first phishing instance was reported in 1995 when attackers convinced victims to share their AOL account details (Jakobsson & Myers, 2006; Rekouche, 2011). Subsequently, phishing has grown and developed. Attackers have devised new methods and utilized new media, and it is now one of the primary attack vectors used by hackers. Given this threat we construct risk communications for phishing in our study to reflect the domain of cybersecurity hazards.

Active Shooter Events

An active shooter is a term used by law enforcement to describe an event where a shooting is in progress, and this crime requires a set of protocols when responding to the incident (Zhu et al., 2020). During the event the shooter is actively engaged in killing or attempting to kill people in a confined and populated area (Brinsfield & Mitchell Jr, 2015). Most active shooter events are preplanned; the shooter usually has prepared himself well with the aim of killing as many individuals as possible. In most cases, the active shooter has already accepted the fact that he or she may also die during this shootout. The active shooter events usually do not last long; most are over within 19 minutes, either because law enforcement stops the shooter, the guns jam, or the person runs out of bullets (Blair et al., 2013).

Figure 2(a) shows that the frequency of active shooter incidents has continued to increase since 2018 and has tripled since 2011. Most active shooter incidents take place in businesses like the postal service, factories, and corporations. However, as Figure 2(b) shows that no location, even a rural one is immune from an active shooter. In 2021, an active shooter incident occurred in almost half of the states in the U.S (Silva, 2022). Given the increase of events and their geographic ubiquity we construct risk communications for active shooter events in our study to reflect the domain of homeland security hazards.

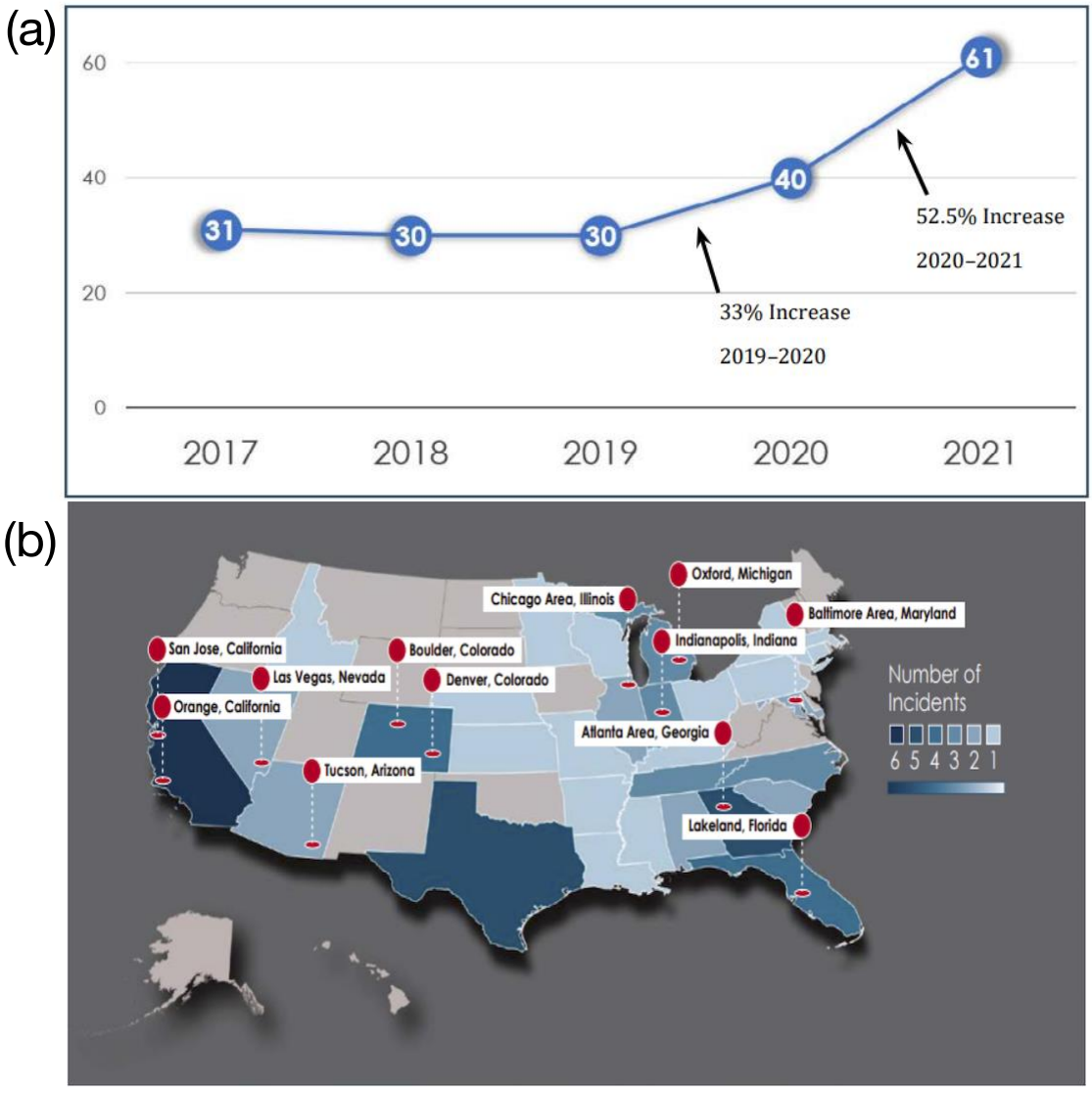

Figure 2: (a) number of active shooter incidents in the United States since 2017-2021 and (b) location of active shooter incidents in the United States in 2021.

Construction of Risk Communication Messages

Conventional Science Messages

Conventional science messages are comprised of three parts: (1) a hazard definition, (2) science information, and (3) a series of protective actions for the recipient to take. The hazard definition is anchored in authoritative preparedness material. In our study we use from guidance from the Federal Emergency Management Agency (FEMA) for protection against flooding, the Federal Trade Commission (FTC) for protection against phishing, and the Department of Homeland Security (DHS) for protection in the event of an active shooter (DHS, 2024; FEMA, 2024; FTC, 2024). The science information in all our messages in this study reflects communicating risk in certainty language. Certainty language states that the identified hazard will happen in the future, even if one does not know exactly when. The protective actions reflect the guidance that authoritative sources want message recipients to take.

Narrative Policy Framework (NPF) Messages

NPF risk communication messages are comprised of four parts. These include the same hazard definitions, science information, and protective actions as conventional science messages. However, they also feature two segments related to the selection of a character narrative. The character narrative selections are: (1) Victim, (2) Victim-to-hero, and Hero. The Hero narrative emphasizes that audience members and their communities can prepare for the hazard. The Victim narrative emphasizes negative outcomes for the audience members and their communities. The Victim-to-hero narrative creates an arc in which the negative outcomes can be reversed by the audience members and their communities.

Given a character selection, construction of NPF risk communication messages consists of four segments. These are: (1) a hazard definition, (2) selected character problem framing, (3) science information, and (4) a selected character in action. The Problem framing segment (#2) introduces the characters and identifies the hazard. The characters in action segment (#4) is the part of the narrative text where the characters provide the action of the story. The protective actions specified in the conventional science messages are woven into the narrative text of the characters in action segment (#4).

The conventional science message, NPF victim, NPF victim-to-hero, and NPF hero messages for an active shooter event are highlighted below. The conventional science and character selection messages for all three hazards are available (Lynch et al., 2024). Recall, the hazard definition, science information, and directives in each message are the same. The conventional science messages created using this approach is:

> ***[Active Shooter definition]*** An active Shooter is an individual who kills or attempts to kill people in a confined and populated area.
> ***[Scientific information]*** Active shooter situations are unpredictable and evolve quickly, ending within 10-15 minutes. The immediate deployment of law enforcement is required to stop the shooting and mitigate harm to targets. Law enforcement's purpose is to stop the active shooter as soon as possible. Upon arrival officers will proceed directly to the area in which the last shots were heard.
> ***[Directive]*** You can improve your chances of survival in an attack by: (1) being aware of your environment and any possible dangers; (2) taking note of the two nearest exits in any facility you visit; (3) if you are in an office, staying there and securing the door; (4) if you are in a hallway, getting into a room and secure the door; and (5) as a last resort, attempting to take the active shooter down.

The Victim character is confronted with the consequences of no preparation:

> ***[Active Shooter definition]*** An active Shooter is an individual who kills or attempts to kill people in a confined and populated area.
> ***[Victim problem framing]*** Many residents in Hampton Roads are concerned with the possibility of being killed or harmed in an active shooter attack at their place of work, an airport, or a public gathering.
> ***[Scientific information]*** Active shooter situations are unpredictable and evolve quickly, ending within 10-15 minutes. The immediate deployment of law enforcement is required to stop the shooting and mitigate harm to targets. Law enforcement's purpose is to stop the active shooter as soon as possible. Upon arrival officers will proceed directly to the area in which the last shots were heard.
> ***[Victim character in action]*** You can improve your chances of survival in an attack by: (1) being aware of your environment and any possible dangers; (2) taking note of the two nearest exits in any facility you visit; (3) if you are in an office, staying there and securing the door; (4) if you are in a hallway, getting into a room and secure the door;

and (5) as a last resort, attempting to take the active shooter down. Without adhering to these guides, you, your family, your friends, and/or your co-workers' lives could be lost as you face difficult and sad times.

The Victim-to-hero language starts with the victim not being able to prevent the attack and turns the character into the hero by implementing protective strategies.

> ***[Active Shooter definition]*** An active Shooter is an individual who kills or attempts to kill people in a confined and populated area.
> ***[Victim-to-hero problem framing]*** Many residents in Hampton Roads are concerned with the possibility of being killed or harmed in an active shooter attack at their place of work, an airport, or a public gathering.
> ***[Scientific information]*** Active shooter situations are unpredictable and evolve quickly, ending within 10-15 minutes. The immediate deployment of law enforcement is required to stop the shooting and mitigate harm to targets. Law enforcement's purpose is to stop the active shooter as soon as possible. Upon arrival officers will proceed directly to the area in which the last shots were heard.
> ***[Victim-to-hero character in action]*** You your family, your friends, and your co-workers can improve your chances of survival in an attack by: (1) being aware of your environment and any possible dangers; (2) taking note of the two nearest exits in any facility you visit; (3) if you are in an office, staying there and securing the door; (4) if you are in a hallway, getting into a room and secure the door; and (5) as a last resort, attempting to take the active shooter down. Following this guide can help save you, your family, your friends, and/or your co-workers from harm.

The hero language shows the character implementing protective strategies to help themself and others throughout the entire message.

> ***[Active Shooter definition]*** An active Shooter is an individual who kills or attempts to kill people in a confined and populated area.
> ***[Hero problem framing]*** In Norfolk, good neighbors like you can educate yourself and others to survive an active shooter situation. While many feel completely safe at work, an airport, or a public gathering there is still the possibility of being killed or harmed in an active shooter attack at these places.
> ***[Scientific information]*** Active shooter situations are unpredictable and evolve quickly, ending within 10-15 minutes. The immediate deployment of law enforcement is required to stop the shooting and mitigate harm to targets. Law enforcement's purpose is to stop the active shooter as soon as possible. Upon arrival officers will proceed directly to the area in which the last shots were heard.
> ***[Hero character in action]*** You your family, your friends, and your co-workers can improve your chances of survival in an attack by: (1) being aware of your environment and any possible dangers; (2) taking note of the two nearest exits in any facility you visit; (3) if you are in an office, staying there and securing the door; (4) if you are in a hallway, getting into a room and secure the door; and (5) as a last resort, attempting to take the active shooter down. Following this guide can help save you, your family, your friends, and/or your co-workers from harm.

Individualized Narrative Policy Framework (NPF) Messages

Since the NPF messages from the previous subsection are not individualized, we refer to them as generic NPF risk communication messages. For each hazard, our study contains a conventional science message (containing a definition, science information, and directives), and three generic NPF risk communication messages. These three generic NPF risk communication messages include one for each of the character selections (victim, victim-to-hero, and hero). In addition, we construct individualized NPF risk communication messages. These three individualized messages are created using large-language models with different subsets of the data about the recipients to personalize the narrative in a generic NPF message.

In our study, we collect demographic, interest/disinterest, and Schwartz value data about each participant. Demographic data includes age, race/ethnicity, gender, military status, marital status, number of children, and state of residence. In addition, participants are asked to indicate which of 19 listed activities (i.e. being physically active; repairing and fixing machines; enjoying music, painting, or writing; being outside; etc.) they are interested in and which activities they are disinterested in. The Schwartz data we collect uses a well-established survey to determine broad values that steer and individuals' behavior (Lindeman & Verkasalo, 2005). The Schwartz values are *self-enhancement* (acting on behalf of oneself), *self-transcendence* (acting on behalf of others), *conservatism* (preferring

traditional norms and experiences), and *openness-to-change* (preferring new norms and experiences) (Schwartz, 2012).

Our approach to using LLMs to create individualized NPF risk communication messages is shown in Figure 3. It leverages recent research in medical, scientific, and engineering communities. These applications demonstrate how LLMs, such as ChatGPT, have been beneficial and challenging across a variety of uses. Recent LLM research summarizes the benefits and challenges of applying these models (Lynch et al., 2023). Benefits show that LLMs perform well when given multimodal (Thirunavukarasu et al., 2023) and validated information (Gilbert et al., 2023) and that consistent and well-structured outcomes can be created with explicit guidance (Filippi, 2023). The identified challenges reflect concerns over validity, uncertainty, bias, and accountability when creating or utilizing LLM outputs.

Our process for creating individualized NPF risk communications is developed with these challenges and concerns in mind. In our study, we utilize GPT 4.0 via OpenAI's ChatGPT API to generate individualized NPF risk communication messages (OpenAI, 2023). Given a generic NPF risk communication message and data about a message recipient, the two NPF message segments that ChatGPT individualizes are: the character problem framing (Segment #2), and the characters in action (Segment #4). This separation is important because: (1) it ensures that both the hazard definition and the science information are not modified, and (2) it enables the characters in the narrative of the risk communication message to be tailored to the individual through an automated means.

Figure 3 also highlights that in our study each individualized NPF risk communication generated by ChatGPT will be manually inspected to determine that it is appropriate to present to a participant. It is important to note that this manual appropriateness check is not a quality assurance effort that would potentially bias the results of our study. Instead, it is meant to screen for vulgar, hostile, and profane language generated by ChatGPT which would be dehumanizing to our participants.

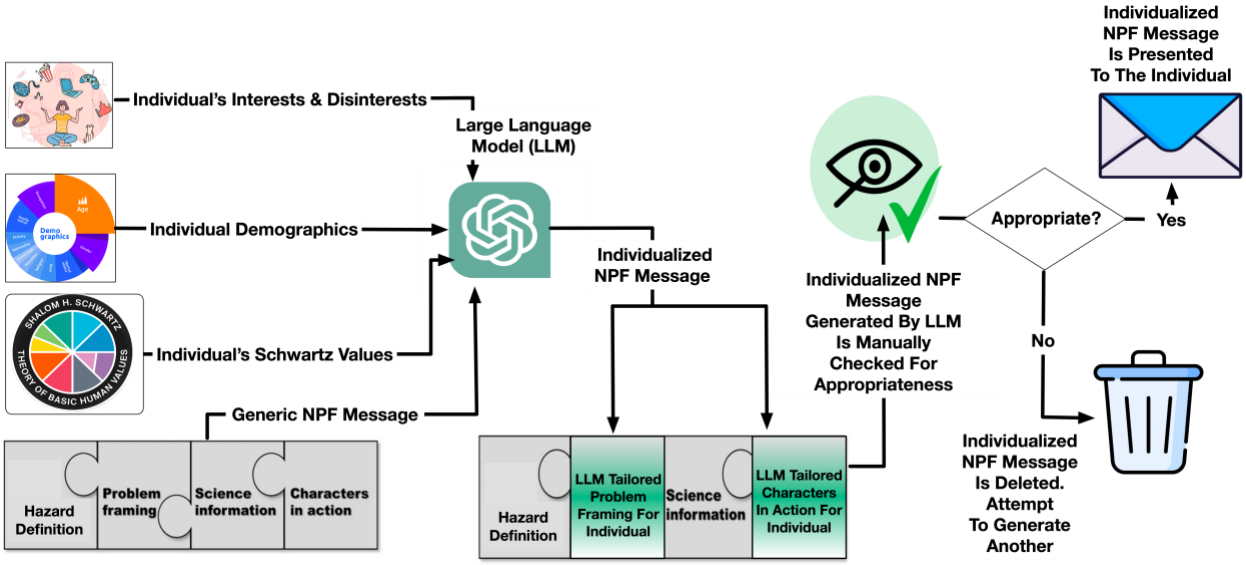

Figure 3: A general overview of the workflow of the inputs, application of individual's data, and use of large language models to create an individualized risk communication message in the NPF.

An example of individualized NPF risk communication messages for an active shooter hazard for a victim-to-hero character helps elucidate the predicted efficacy of our work. The individualized message for this character is shown below. The data about the individual, the generic NPF message, and the script used to prompt ChatGPT for this example are available (Lynch et al., 2024). This individual does not reflect an actual study participant. Instead, this is one of several synthetic examples we have constructed to incrementally test and improve our framework before deploying it in our study.

This is the individualized NPF victim-to-hero character message for a single, 18-year-old, woman with self-enhancing and conservative values. The individual likes to be physically active, enjoys the outdoors, and likes music. The individual dislikes fixing or repairing things, and dislikes writing. We highlight ChatGPT's individualization of the message by underlining the modifications it made to the text based on the individual's data.

> ***[Active Shooter definition]*** An active Shooter is an individual who kills or attempts to kill people in a confined and populated area.

[Victim problem framing] Many women in Hampton Roads, are concerned with the possibility of being killed or harmed in an active shooter attack at their place of work, an airport, or a public gathering like a music concert. Attacks could occur in outdoor locations.
[Scientific information] Active shooter situations are unpredictable and evolve quickly, ending within 10-15 minutes. The immediate deployment of law enforcement is required to stop the shooting and mitigate harm to targets. Law enforcement's purpose is to stop the active shooter as soon as possible. Upon arrival officers will proceed directly to the area in which the last shots were heard.
[Hero character in action] You can improve your chances of survival and protect yourself in an attack by: being aware of your environment and any possible dangers; (2) taking note of the two nearest exits in any facility you visit, it is not necessary to write them down; (3) if you are in an office, staying there staying there and securing the door; (4) if you are in a hallway, getting into a room and secure the door; and (5) as a last resort, attempting to take the active shooter down. While taking the shooter down may seem like an aggressive last resort you can summon your physical strength to do it. Following this guide can help you save yourself from harm.

This message shares a significant amount in common with the generic NPF victim-to-hero character active shooter risk communication message shown previously. The language starts with the victim not being able to prevent the attack and turns the character into the hero by implementing protection strategies. However, in the problem framing and character in action segments ChatGPT makes modifications, often at the beginning or end of a sentence / phrase, that directly reflects data about the recipient. The use of the word 'women' as opposed to 'people' in the first sentence of the problem framing demonstrates ChatGPT has recognized and attempted to incorporate the individual's age and gender into the message. Furthermore, the problem framing ends with a recognition of the individual's interest in 'music' and the 'outdoors'.

Then, in the character in action section, the Schwartz values (i.e. self-enhancing and conservative) of the individual are considered. The message focuses on only the survival and protection of the recipient, not others. This text reflects the self-enhancing values of the individual and the demographic data indicates that the individual is single. ChatGPT also recognizes the conservative nature of the individual and urges them, as a last resort, to use strength cultivated from their enjoyment of physical activities to overcome it.

However, ChatGPT does not make effective use of all the data about the recipient in the individualized message. Its modifications and additions are most limited when attempting to leverage data about the dislikes of the individual. There is not any content that appears related to the individual's dislikes of fixing or repairing things. One reaction is that fixing or repairing things seems unrelated to an active shooter event. However, a similar argument can be made about enjoying music. Furthermore, when ChatGPT does attempt to leverage data about the individual disliking writing things down it is not effective. The text that emphasizes that 'it is not necessary to write down' the nearest exists in a facility could be distracting and take away from the importance of the protective action given to the recipient in the message.

It is important to note that this is one generated individualized NPF risk communication message, for one character and for only one hazard. We have chosen to highlight it in this paper because it demonstrates the potential efficacy of our approach and obstacles we still need to overcome. Some individualized NPF message examples we have generated seem less effective than this one. There are other examples that we have generated that seem to integrate more data about the recipient but do so less subtly.

Overall, this example reflects iterative refinement of prompting ChatGPT for individualized NPF messages. In the refinement process we have continually constrained the degrees of freedom for ChatGPT additions and modifications to the generic NPF message. In our experience, ChatGPT is most effective at creating individualized NPF messages when given explicit directions about where within the character problem framing and character in action segments it should provide its additions and modifications.

METHODS, EVALUATION, AND HYPOTHESES

This research was deemed exempt from Institutional Research Board (IRB) review by Old Dominion University IRB # 2106298-3. Currently, we have received interest from 109 participants, 65 of whom have already completed our demographic, interest / disinterest, and Schwartz value questionnaires and been assigned a NPF character selection for each of the three hazards featured in our study.

Our study will employ a within-subjects design in which participants will be exposed to five risk messages for each hazard domain: one conventional science message, one generic NPF message and three individualized NPF messages. The character selection for each hazard will be the same in all five risk messages they receive. Each of the individualized NPF messages will use different subsets of the data about the individual. One individualized NPF

message will use their demographic data, and their interest / disinterest data. Another individualized message will use their demographic data, and their Schwartz value data. The third individualized message will use all their data: demographic, interest / disinterest, and Schwartz values.

Evaluation

After each message, an individual will be asked to identify the extent to which they will follow four protective actions related to the hazard. The extent to which the individual will follow these protective actions determines the message's efficacy. For each of the protective actions, the participant will indicate the extent to which they plan to take the action on a 0 (never plan to take the action) to 100 (always plan to take the action) slider scale. For each hazard, we will run a reliability analysis to compute the agreement among the protective actions. If one or more protective actions results in low agreement, we will remove it from analysis. This will be done to ensure participant responses reflect the true efficacy of the risk communication message and are not compromised by the inclusion of asking them to take polarizing protective action(s).

Research Hypotheses and Predicted Effects

We have three hypotheses related to our study. For each hypothesis we have specified the predicted effect. It is important to note that we are still in the process of collecting and evaluating data. However, highlighting the hypotheses and predicted effect for our study helps elucidate the motivation for our work and increases its transparency and rigor.

Hypothesis and Predicted Effect 1: NPF Messages Are More Effective Than Conventional Science Message Agnostic of Domain

Across all hazard types, risk communication messages constructed using the NPF narrative structure are more effective than conventional science messages because narrative transportation is an effective means of engaging a target audience. If the narrative mechanism of character is an important driver of narrative transportation across hazard types, then the target audience will report a greater willingness to take protective actions when responding to narrative messages than straight-science messages. Responses to take protective actions will be consistently higher for narrative messages across all hazard types – ranging from natural disasters (i.e. flooding), cyber security (i.e. phishing), and homeland security (i.e. active shooter incidents).

Data from our study that supports this hypothesis will provide further evidence that risk communication messages constructed using the NPF narrative structure are more effective than conventional science messages (D Jones & Anderson Crow, 2017; Reinhold, Raile, et al., 2023; Shanahan et al., 2019). This has been already established and additional support will continue to make this clear.

Hypothesis and Predicted Effect 2: Personalized NPF Messages Are More Effective Than Generic NPF Messages Agnostic of Domain

Across all hazard types, risk communication messages constructed using the NPF that are personalized for the target audience (i.e. individualized NPF messages) are more effective than risk communication messages constructed using the NPF for a general audience (i.e. generic NPF messages) because personalization is an important driver of audience engagement with hazard messaging. If personalization is an important driver of audience engagement, then participants will report a greater willingness to take protective actions when responding to personalized narrative messages than narrative messages written for a broad audience. Responses to take protective actions will be consistently higher for personalized messages across all hazard types - ranging from natural disasters (i.e. flooding), cyber security (i.e. phishing), and homeland security (i.e. active shooter incidents).

Previous research has shown that providing community specific information in risk communication messages constructed using the NPF is effective (Reinhold, Raile, et al., 2023). Data from our study that supports this hypothesis will provide additional evidence that the efficacy of NPF risk communication messages is increased when individuals see themselves, their interests, and/or their values in them.

Hypothesis and Predicted Effect 3: Using All Personal Information Available To Construct A Personalized NPF Messages Creates More Effective Messages Than Only Using A Subset of Personal Information Available Agnostic of Domain

Across all hazard types, risk communication messages constructed using the NPF narrative structure that use all the personal information available about an individual are more effective than risk communication messages constructed using the NPF narrative structure which only use a subset of the personal information about an individual. If

personalization is an important driver of audience engagement, then participants will report a greater willingness to take protective actions when responding to personalized narrative messages that use all the information about them as opposed to risk communication messages constructed using the NPF narrative structure that only use a subset of the information.

Recall, in our study we will provide participants with 3 individualized NPF messages. Each message will use a different subset of the personal data to create the tailored message. This hypothesis reflects that we expect the individualized NPF message created will all the participant's data to result in responses where protective actions will be consistently higher than the individualized NPF messages created using only some of the participant's data.

CONCLUSION

To date employing the NPF in risk communication has been very effective but also labor intensive. In part this is because no protocol exists to guide the construction of risk communication messages. The studies employing these messages rely on large amounts of manual effort resulting in a process that is difficult to repeat and replicate. Here, we apply large language models (LLMs) to construct risk communication messages, leveraging data about recipients, in the NPF in a more repeatable and automated manner. We hypothesize this tailored and individualized message for the recipient will result in improved risk communication efficacy that will save lives agnostic of the risk domain.

ACKNOWLEDGEMENTS

This study was funded by the Office of Enterprise Research and Innovation at Old Dominion University (#300916-010).

REFERENCES

Blair, J. P., Nichols, T., Burns, D., & Curnutt, J. R. (2013). *Active shooter events and response*. CRC Press.

Brinsfield, K. H., & Mitchell Jr, E. (2015). The Department of Homeland Security's role in enhancing and implementing the response to active shooter and intentional mass casualty events. *Bulletin of the American College of Surgeons*, *100*(1 Suppl), 24-26.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Cao, Y., Boruff, B. J., & McNeill, I. M. (2017). Towards personalised public warnings: harnessing technological advancements to promote better individual decision-making in the face of disasters. *International Journal of Digital Earth*, *10*(12), 1231-1252.

D Jones, M., & Anderson Crow, D. (2017). How can we use the 'science of stories' to produce persuasive scientific stories? *Palgrave Communications*, *3*(1), 1-9.

De Wit, J. B. F., Das, E., & Vet, R. (2008). What works best: objective statistics or a personal testimonial? An assessment of the persuasive effects of different types of message evidence on risk perception. *Health Psychology*, *27*(1), 110.

DHS. (2024). *Active Shooter: How To Respond*. Deptartment of Homeland Security. Retrieved February 5, 2024 from https://www.dhs.gov/xlibrary/assets/active_shooter_booklet.pdf

Dransch, D., Rotzoll, H., & Poser, K. (2010). The contribution of maps to the challenges of risk communication to the public. *International Journal of Digital Earth*, *3*(3), 292-311.

FEMA. (2024). *Be Prepared for a Flood*. Department of Homeland Security. Retrieved February 5, 2024 from https://community.fema.gov/ProtectiveActions/s/article/Flood

Filippi, S. (2023). Measuring the impact of ChatGPT on fostering concept generation in innovative product design. *Electronics*, *12*(16), 3535.

Fischhoff, B., & Scheufele, D. A. (2013). The science of science communication. *Proceedings of the National Academy of Sciences*, *110*(supplement_3), 14031-14032.

FTC. (2024). *How to Recognize and Avoid Phishing Scams*. The Office of Inspector General. Retrieved February 5, 2024 from https://consumer.ftc.gov/articles/how-recognize-and-avoid-phishing-scams

Gilbert, S., Harvey, H., Melvin, T., Vollebregt, E., & Wicks, P. (2023). Large language model AI chatbots require approval as medical devices. *Nature Medicine*, *29*(10), 2396-2398.

Hunt, S., Green, K., Craig, C., Crisman, P., Ye, M., League, L. T. C., & Watch, W. (2019). Creating Coastal Resilience through Community Engagement: A Plan for Lake Taylor.

Jakobsson, M., & Myers, S. (2006). *Phishing and countermeasures: understanding the increasing problem of electronic identity theft*. John Wiley & Sons.

Jones, M. D. (2014). Cultural characters and climate change: How heroes shape our perception of climate science. *Social Science Quarterly*, *95*(1), 1-39.

Lindeman, M., & Verkasalo, M. (2005). Measuring values with the short Schwartz's value survey. *Journal of personality assessment*, *85*(2), 170-178.

Lynch, C. J., Gore, R., & Jensen, E. J. (2024). *Large Language Model-Driven Narrative Generation Study Data: ChatGPT-Generated Narratives, Real Tweets, and Source Code* Version 3). https://doi.org/10.17632/nyxndvwfsh.2

Lynch, C. J., Jensen, E. J., Zamponi, V., O'Brien, K., Frydenlund, E., & Gore, R. (2023). A Structured Narrative Prompt for Prompting Narratives from Large Language Models: Sentiment Assessment of ChatGPT-Generated Narratives and Real Tweets. *Future Internet*, *15*(12), 375.

Mitchell, A. P. (2021). *Impacts of coastal flooding on watersheds in Hampton Roads, VA* Virginia Tech].

OpenAI. (2023). *ChatGPT*. https://chat.openai.com

Paton, D., Smith, L., Daly, M., & Johnston, D. (2008). Risk perception and volcanic hazard mitigation: Individual and social perspectives. *Journal of Volcanology and Geothermal Research*, *172*(3-4), 179-188.

Pidgeon, N., Kasperson, R. E., & Slovic, P. (2003). *The social amplification of risk*. Cambridge University Press.

Reinhold, A. M., Gore, R. J., Ezell, B. C., Izurieta, C. I., & Shanahan, E. A. (2023). From Cyclones to Cybersecurity: A Call for Convergence in Risk and Crisis Communications Research. https://doi.org/10.5281/zenodo.10067626

Reinhold, A. M., Munro, M. H., Shanahan, E. A., Gore, R. J., Ezell, B. C., & Izurieta, C. (2024). Embedding Software Engineering in Mixed Methods: Computationally Enhanced Risk Communication. https://doi.org/10.5281/zenodo.10519749

Reinhold, A. M., Raile, E. D., Izurieta, C., McEvoy, J., King, H. W., Poole, G. C., Ready, R. C., Bergmann, N. T., & Shanahan, E. A. (2023). Persuasion with precision: Using natural language processing to improve instrument fidelity for risk communication experimental treatments. *Journal of Mixed Methods Research*, *17*(4), 373-395.

Rekouche, K. (2011). Early phishing. *arXiv preprint arXiv:1106.4692*.

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online readings in Psychology and Culture*, *2*(1), 11.

Shanahan, E. A., Raile, E. D., French, K. A., & McEvoy, J. (2018). Bounded stories. *Policy Studies Journal*, *46*(4), 922-948.

Shanahan, E. A., Reinhold, A. M., Raile, E. D., Poole, G. C., Ready, R. C., Izurieta, C., McEvoy, J., Bergmann, N. T., & King, H. (2019). Characters matter: How narratives shape affective responses to risk communication. *PLoS One*, *14*(12), e0225968.

Shen, F., Sheer, V. C., & Li, R. (2015). Impact of narratives on persuasion in health communication: A meta-analysis. *Journal of advertising*, *44*(2), 105-113.

Silva, J. R. (2022). Mass shooting outcomes: A comparison of completed, attempted, failed, and foiled incidents in America. *Deviant Behavior*, *43*(12), 1487-1506.

Skaggs, L. L., & Seltzer, C. L. (2002, 2002//). COST EFFECTIVENESS AND INCREMENTAL COST ANALYSES FOR WATERSHED DECISION-MAKING: APPLICATION TO THE ELIZABETH RIVER, VIRGINIA, ENVIRONMENTAL RESTORATION. Watershed Conference 2002,

Stavroulakis, P., & Stamp, M. (2010). *Handbook of information and communication security*. Springer Science & Business Media.

Stone, D. A. (2012). *Policy paradox: The art of political decision making* (Vol. 3). WW Norton & company.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, *29*(8), 1930-1940.

Wachinger, G., Renn, O., Begg, C., & Kuhlicke, C. (2013). The risk perception paradox—implications for governance and communication of natural hazards. *Risk analysis*, *33*(6), 1049-1065.

Zhu, R., Lucas, G. M., Becerik-Gerber, B., & Southers, E. G. (2020). Building preparedness in response to active shooter incidents: Results of focus group interviews. *International Journal of Disaster Risk Reduction*, *48*, 101617.