CORE
sCience and human factOr for Resilient sociEty

Deliverable title:	Impact of misinformation on social media on risk perception in a multi-risk environment
Deliverable ID:	D 7.2
Document version:	Final Version
Partner responsible:	ETH Zurich
Due date:	M24 (August 2023)
Status:	Reviewed and submitted

Document Control Sheet

Deliverable number	D7.2
Deliverable name	Impact of misinformation on social media on risk perception in a multi-risk environment
Dissemination Level	Public
Call	H2020-SU-SEC-2020
Topic	SU-DRS01-2018-2019-2020 Human factors, and social, societal, and organizational aspects for disaster-resilient societies
Consortium Coordinator	UNISA
Edition	

Authoring & Approval

Authors of the document	
Name/Beneficiary	Position/Title
Irina Dallo (ETH Zurich)	Postdoctoral associate/ Dr
Michèle Marti (ETH Zurich)	Head of communications / Dr
Giovanni Gugg (ISSNOVA)	Senior researcher / Dr
Carmit Rapaport (CLB)	Director, NIRED-Institute for Regulation of Emergency and Disaster, College of Law and Business / Dr
Abraham Yosipof (CLB)	Dean, Faculty of Information Systems and Computer Science / Dr
Or Elroy (CLB)	PhD student/ MSc
Nadejda Komendantova (IIASA)	Senior research scholar and research group leader/ Prof
Dmitry Erokhin (IIASA)	Researcher / Dr
Rosa Vicari (IIASA)	Guest Research Scholar / Dr
Laure Fallou (EMSC)	Research Officer / MSc

Approved for submission - Representatives of beneficiaries involved in the project		
Name/Beneficiary	Position/Title	Date
Santino Ranieri (UCSA)	Project contributor / Dr	25/07/2023
Raffaele Tortora (UCSA)	Project contributor / Dr	25/07/2023
Anthony Lamaudière (PSCE)	Project officer	02/08/2023
Melissa Scott (PSCE)	Project officer	03/08/2023
Marie-Christine Bonnamour (PSCE)	General secretary / Dr	04/08/2023

ABSTRACT

Misinformation is not a new phenomenon but, through social media, has gained new dynamics such as the rapidity of spread around the world within a few seconds. Past events have shown that misinformation can worsen the response to an emergency by leading to inappropriate behaviours, triggering fear and anxiety, or reducing the credibility of the measures by official actors.

To better understand the dynamics of misinformation on social media or in the press and its effects on people's beliefs and behaviour, we defined six case studies addressing different hazards and time periods. This allowed us to derive recommendations to prevent and fight the spread of and belief in misinformation along the entire communication chain - source, message, channel, receiver, effect, and feedback.

Three of our key results are that: i) official actors should provide information on a regular basis to build credibility and trust, which will allow them to effectively communicate and counter misinformation during emergencies when people are under stressful conditions; ii) fragmented information on social media should be counterbalanced through external links to richer sources, where people can inform themselves about the broader context and details; and iii) the cultural context and ideological debates must be considered to address anchored beliefs and biases when developing strategies to prevent and fight misinformation.

TABLE OF CONTENTS

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CORE	sCience and human factor for Resilient sociEty
EQ	Earthquake
LP-HI	Low-Probability and High-Impact
NIMTOF	Not In My Term Of Office bias
NLP	Natural Language Processing
OLS	Ordinary Least Squares
RISE	Real-time earthquake rIsk reduction for a reSilient Europe
RoBERTa	Robustly optimized BERT pretraining approach
TURNkey	Towards more Earthquake-resilient Urban Societies through a Multi-sensor-based Information System enabling Earthquake Forecasting, Early Warning and Rapid Response actions
URL	Uniform Resource Locator

1 INTRODUCTION

The anthropogenic and natural disaster scenarios targeted in CORE show that multiple factors determine the societal response to an emergency and, consequently, its impacts. What the cases have in common is that there were various obstacles along the entire communication chain – source, message, channel, receiver, effect, and feedback (Berlo, 1960) –, which challenged an effective response. The specific obstacles in the different scenarios are described in the following paragraphs.

The earthquake in L'Aquila in 2009 indicates that the communication of low-probability high-impact events and the related uncertainties is not trivial. The communication of inaccurate, incomplete, and contradictory information about the chance of a possible damaging earthquake immediately before the severe earthquake in 2009 led to a legal trial after the event (Alexander, 2010; Cocco et al., 2015). The trial was also requested by the public since people after a disaster always ask for justice (Lukasiewicz & Baldwin, 2020). This legal process is one of the main reasons why scientists in Italy are nowadays hesitant to communicate any information to the public. Further, the communication responsibilities of the involved institutions were not clearly defined, which has been done in the aftermath of the earthquake (Alexander, 2019).

The industrial accident in Venkatapuram (India) in 2020 showed that unclear responsibilities – vertically (governmental levels) and horizontally (local actors) – inhibit effective disaster risk reduction, and the more actors are involved the more diffuse the lines of accountability get (Amaratunga et al., 2019; Prerna & Abhigya, 2021). This case study further stresses how multiple hazards can worsen an emergency. Problematic was the occurrence during the Covid-19 shutdown, which not only contributed to the incident itself, because the maintenance work was neglected, but also complicated the medical response (Hailwood, 2020; Jadhav et al., 2020). Moreover, even security thresholds were defined by international protocols, no control mechanisms were in place, leading to the neglect of them (Hailwood, 2020; Prerna & Abhigya, 2021).

During the Covid-19 pandemic, societies were challenged by the spread of various conspiracy theories and misinformation on social media (C. Zhou et al., 2021). Especially in the context of the vaccines, various conspiracies circulated: i) going to be used to control and track people (Kricorian et al., 2021); ii) going to be used to alter people's DNA (Kricorian et al., 2021); iii) uncertainty regarding the protection duration (Finney Rutten et al., 2021); and iv) apprehension about adverse effects (Finney Rutten et al., 2021). A further study showed that people who believed that Covid-19 vaccines are unsafe were on average less educated,

had a lower income, and lived in rural areas (Kricorian et al., 2021). Further, especially in times of high uncertainties, lacking authoritative information, or the presence of ambiguous information, fake news and misinformation is more likely believed in and leads to anxiety and fear (Fallou et al., 2020; Peng, 2020). This consequently can generate mistrust leading to societal behaviours that worsen a situation (Chen et al., 2018).

Other crises, such as the floods in Germany in 2021, indicate that a warning system can technically work properly, but if the last step, namely the communication with the affected people does not work, the crisis cannot be prevented (Fekete & Sandholz, 2021). In addition, violence against first responders is increasingly reported, which can worsen an emergency drastically (Fekete & Sandholz, 2021). Further, people in the affected areas reported that missing information was the top problem, followed by misinformation from official sources or from the media. In addition, multiple threats may simultaneously pose challenges to response during and after an event, e.g., insufficient resources in hospitals due to too many COVID-19 patients (Fekete & Sandholz, 2021).

All these cases highlight that at the core of handling an emergency is the appropriate communication along the entire chain from the source to the receiver of the information (Berlo, 1960); also considering the contextual factors (Dallo, 2022). Thus, not only the role of the institutions communicating with the public and the information spread via the different channels matter but also how people handle the information and make decisions based on it. Analysing such events in the aftermath allows to derive needed changes along the disaster cycle and, consequently, to be better prepared for future emergencies (Alexander, 2019). As a consequence, the feedback loop of the communication chain allows one to improve the communication efforts continuously.

The aim of task 7.2 was to analyse the dynamics of misinformation and the role of authoritative voices (on social media) along the communication chain. To this end, we conducted six case studies (see Table 4): i) Misinformation about the link between Covid-19 and 5G on Twitter; ii) Misinformation about earthquake predictions on social media; iii) Fake news about the volcano Vesuvius on general news media; iv) Mining the discussion of Monkeypox misinformation on Twitter; v) Misinformation and the role of media after the Manchester Arena attach; and vi) Authoritative policies to increase societies' resilience to earthquakes – a cross-cultural comparison. We decided to not cover the initial CORE case studies except Covid-19, because the social media evolved significantly in the last years. It was instead considered more appropriate to analyse more recent events to derive recommendations useful for the current communication dynamics. The insights from the CORE specific case studies (summarized above) were used to

define our six cases to make sure to cover different hazards, time horizons (e.g., short-lived vs. on-going), and challenges along the communication chain. for the broader discussion.

2 Theoretical background

In chapter 2, we first introduce the communication chain which builds the backbone of our six case studies. Secondly, we provide an overview of the concepts to understand the belief in and spread of misinformation. Thirdly, we provide an overview of communication strategies, which should allow one to reduce the spread of misinformation during emergencies and minimize its negative impacts.

2.1 Communication chain

Berlo introduced in 1960 a communication process model with the following steps: Source – Message – Channel – Receiver – Effect – Feedback (see Figure 1). In a nutshell, several actors (sources) provide information to the society and communicate it in different formats (messages). The messages are then disseminated via multiple channels to reach the receivers in an efficient way. The receivers differ due to personal as well as contextual factors, which leads to different responses although receiving the same information. The feedback step allows one to analyse these effects and, in turn, adjust the previous steps to foster a desired behavioural response. In the following each step of the communication chain is described in more detail. Even though it is a simplified process, it allows one to keep the overview of the communication networks and reflect on specific issues along the chain.

COMMUNICATION CHAIN [BERLO, 1960]

SOURCE — MESSAGE — CHANNEL — RECEIVER — EFFECT

FEEDBACK

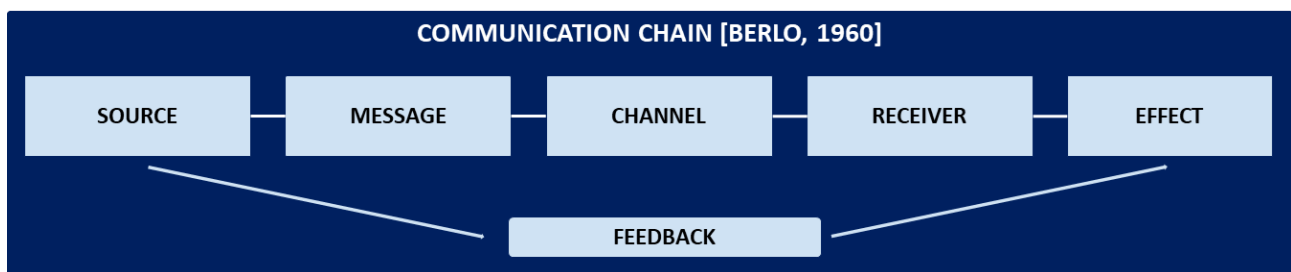*Figure 1 – Overview of the communication process model; introduced by Berlo in 1960*

Due to the technological developments in the last decades, the number of information providers (sources) has increased rapidly. People have different levels of trust in these *sources*, whereby in Switzerland for example the public has the highest trust in authoritative institutions and the lowest in private institutions

and the media (Dallo et al., 2020). During emergencies, people often trust local officials and experts the most (Reuter & Spielhofer, 2017). Trust in the senders of the information influences people's belief in the correctness of the messages (Garcia & Fearnley, 2012; Sullivan-Wiley & Short Gianotti, 2017) and, consequently, their intention to take protective actions.

Messages can have different formats (e.g., textual vs. visual information) and content. Regarding the latter, research indicates that the following five elements are crucial for messages during an emergency: hazard type, affected area, time and duration, source, and behavioural recommendations or instructions (Bean et al., 2015; Mileti & Sorensen, 1990). In addition to this, recent studies have shown that people further desire impact-based warnings (Potter et al., 2021), information about secondary hazards (Becker et al., 2020), and interactive tools such as a sharing button or an 'I am safe' button (Dallo & Marti, 2021). Regarding the format, Kreuzmair et al. (2017) showed that pictograms increase people's compliance to warning messages, and Bossu et al. (2018) that people prefer a combination of visual and textual behavioural guidance.

Beside the traditional dissemination **channels**, such as television and radio, people prefer to receive hazard alerts via modern channels, which allow them to rapidly share the alerts with friends and the family (Maduz et al., 2018). These modern channels include computers, smartphones, and any other digital device. Regarding multi-hazard warning platforms, people for example prefer a mobile application (Dallo et al., 2020) to receive push notifications tailored to their current location (Kotthaus et al., 2016). Additionally, social media platforms have become a channel to seek and share information, allowing to spread messages around the world in a few seconds. In the absence of authoritative information, social media can even help people handle an emergency on their own (Fallou et al., 2020). However, the dynamics on social media also allow people to rapidly spread fake news around the world (Zhou et al., 2018). The good thing is that the public's trust in the content of social media is rather low (Lacassin et al., 2019), which may minimize the belief in misinformation (see section 2.2.1).

Receivers of a message can be characterized by personal and contextual factors. Regarding the personal factors, Kurata et al. (2022) showed that people's perceived behavioural control and attitudes significantly affect their intention to follow flood disaster risk response instructions during typhoons. Further, the contextual factors describe receivers' social and physical environment, i.e., neighbourhood, peers, workplace, or political system. Households, for example, are more likely to take preparedness actions if they observe that others are taking them (Wood et al., 2012). However, selection between various risks, such as the

ecological and social risk on Vesuvius, also influences people's decision to take protective actions or not (Gugg, 2021; 2022).

Different social cognition models try to explain which factors influence how receivers respond to a message (**effect**): i) transtheoretical model (Prochaska et al., 1994); ii) risk information seeking and processing model (Griffin et al., 1999); iii) protective action decision model (Lindell & Perry, 2012); or iv) theory of planned behaviour (Ajzen, 1991). All these models try to identify the factors which allow one to predict people's response to a (warning) message (e.g., self-efficacy, knowledge, prior experiences), which can be addressed in information campaigns and used for the design of actionable warning messages.

With the **feedback** loop the effectiveness of the entire communication process can be evaluated; identifying the critical nodes where changes are needed. By implementing these changes, the effectiveness of people's responses should be increased. Further, the feedback loop also allows one to address new legal frameworks, changing needs from the receivers, or technological innovations. To this end, qualitative and quantitative methods can be used such as interviews, focus groups, or surveys. But also informal feedback/exchanges should not be underestimated, especially to maintain well-established communication networks.

2.2 The concepts of false information

False information is information that is not correct according to the available evidence at the time. Experts thereby differentiate between misinformation, (pure) disinformation, fake news, and conspiracy theories (see Table 1); hereafter referred to as misinformation because a clear separation between the terms is not always possible. For example, it is difficult to check if the person spread certain false misinformation deliberately or not.

Table 1 – Definition of the different types of false information, adapted from [(Dallo et al., 2022) & (Bruns et al., 2022).

Term	Description
Misinformation	Information that is false or misleading and is communicated regardless of an intention to deceive (Komendantova et al., 2021).
Disinformation	False information that is deliberately spread to deceive (Komendantova et al., 2021).
Conspiracies	Is a specific type of misinformation. Conspiracy theories are "a proposed explanation of some historical event in terms of the significant causal agency of a relatively small group of persons – the conspirators – acting in secret (Keeley, 1999, S. 116)".

Fake news	Is a type of disinformation. Fake news is the deliberate presentation of false or misleading claims as news, where the claims are misleading by design (e.g., manipulation of public opinion; Gelfert, 2018).

Misinformation is neither a new problem nor only limited or a result of social media. Further, misinformation is also not present at a much larger scale than we have historically observed (Scheufele et al., 2021). In the following, we first summarize why people believe and spread misinformation and which effects this has on individuals and the entire society. Second, we look at social media and why its dynamics facilitate the spread of misinformation. Third, we provide a list of strategies to fight the spread of and belief in misinformation. These different levels are visually depicted in Figure 2.
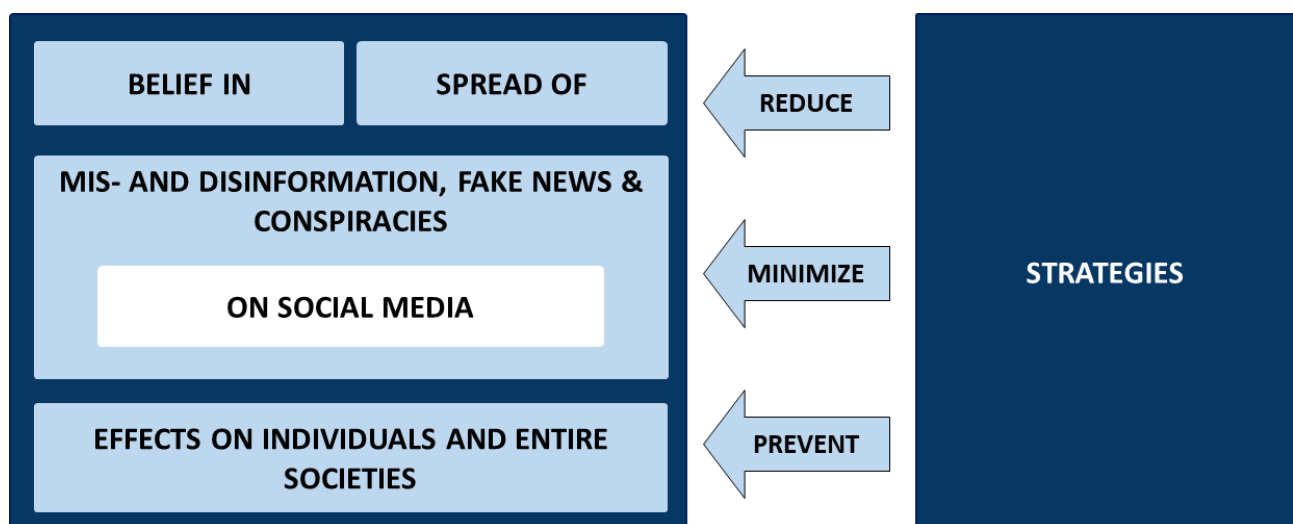
Figure 2 – Overview of the key elements related to the prevention and fight of misinformation in social media.

2.2.1 Reasons why people believe in misinformation or not

There are several reasons why people **believe** in misinformation:

- First, the claim comes from a trusted source or a trusted source confirms the information (Jahanbakhsh et al., 2021).
- Second, the news is plausible and consistent with past experiences or observations (Jahanbakhsh et al., 2021).
- Third, when authoritative information is missing but sought by the public, misinformation can fill the gap (Jones et al., 2017; Peng, 2020). Thus, when information from official channels is lacking or disseminated irregularly,

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

- people have a higher risk for exposure to rumours that fill the information void (Jones et al., 2017).
- Fourth, in particular for high-impact and low-probability events, misinformation circulates because of the uncertainties associated with them, such as the Covid-19 pandemic has illustrated (Peng, 2020; UNESCO, 2022).
- Fifth, people actually do not know but want the misinformation to be true because it may help to give sense to a situation (Jahanbakhsh et al., 2021).
- Sixth, misinformation is often written with more ambiguous words and, consequently, can lead to false interpretations (Zhou et al., 2021).
- Seventh, conspiracies are compelling narratives that trigger emotions and are more relatable than scientific texts, i.e. simple language, videos, and images are used (UNESCO, 2022; van Prooijen & Douglas, 2018). People who deliberately share fake news and conspiracy theories use more negative emotions (e.g., attack), anger words, and anxiety links (e.g., fear, threat), and refer to topics such as death, religion, and power (Fong et al., 2021).
- Eighth, conspiracies are often conceptualized in a way to empower people when they feel powerless, meaning that when people feel vulnerable or isolated from politics and economic power, they are more prone to believe in conspiracy theories (UNESCO, 2022). And the challenge is that people who believe in one conspiracy theory are often also more likely to believe in others (UNESCO, 2022).

However, there are many people who **disbelieve** misinformation that is regularly spread due to multiple reasons:

- First, the information contradicts what people know from their trusted sources (Jahanbakhsh et al., 2021).
- Second, the information comes from an untrusted source or appears biased (Jahanbakhsh et al., 2021).
- Third, people with a high level of literacy are more cautious and suspicious about online information and, consequently, critically question it compared to people with low literacy (Chen et al., 2018).
- Fourth, people think that he/she should have heard about it before, thus questioning the correctness of the information (Jahanbakhsh et al., 2021).

These findings indicate that it is indispensable that trusted sources communicate regularly what is true and what not so that people have already come across it and are better prepared to identify misinformation. This also allows for example authoritative institutions to educate people to increase their literacy about relevant topics related to an event/emergency.

2.2.2 Reasons why people share misinformation

There are several reasons why people tend to share misinformation. During health emergencies, people pay more attention to their health conditions so misinformation associated with health suggestions that prevent someone from acquiring diseases are more likely to be shared (Zhou et al., 2021). Further, if updates from authorities or the media are missing, misinformation is more likely shared with others to fill the information void (Peary et al., 2012). In addition, political orientation plays a role too, i.e. conservatives in the U.S. are more likely to share articles from fake news domains than liberals or moderates (Guess et al., 2019). Thereby, especially misinformation which reinforces their socio-political beliefs and identity structures are spread by them (van der Linden, 2022). Further, Facebook users over 65 were identified to share seven times as many articles from fake news domains as the youngest age group (Guess et al., 2019). Moreover, people are sometimes duped by misinformation because they are distracted on social media and are not paying sufficient attention to accuracy cues (van der Linden, 2022). Additionally, narcissists are more likely to share conspiracy theories on social media because these messages are entertaining, elicit strong emotions or contain secrecy and persecutory elements. Or other reasons could be due to gullibility or seeking to manipulate others and blame them for misfortunes (Cichocka et al., 2022). In summary, anxiety, source or content ambiguity, personal involvement, confirmation bias, ease of sharing, illiteracy, attractiveness, attitudes, and social ties influence the (un)intentional spread of misinformation (Muhammed T & Mathew, 2022).

2.2.3 Effects on individuals and entire societies

The spread of misinformation can trigger fear and anxiety, as shown after severe earthquakes (Peng, 2020) and during health emergencies (Zhou et al., 2021) for example. Jones and Silver (2020) have shown that when receiving a false alarm anxiety remained elevated for at least 7 days post alert. Thus, a single all-clear message is not enough to calm people in the aftermath and ensure that they do not believe in the misinformation that is circulated. Further, misinformation can also create unnecessary panic. During the 2018 floods in the South-Indian state of Kerala for example, a fake video on Mullaperyar Dam leakage created panic among the citizens, negatively impacting the rescue operations (Pierpoint, 2018). Another example, in communities unaffected by the earthquake in Hokkaido (Japan) in 2018, stronger bonding and bridging ties encouraged unnecessary evacuation, leading to the spread of rumours during the blackout (Fraser et al., 2021). In addition, misinformation can pose a considerable threat to public health and to a successful management of a global pandemic (van der Linden, 2022). Many people believing in conspiracy theories for example could have been saved

during the Covid-19 pandemic but they refused the vaccines or medical care at the hospital (Brumfiel, 2022).

2.2.4 Past events

The most recent global crisis challenged by misinformation is the pandemic Covid-19 (Peng, 2020). However, misinformation on social media is not new and has already been demonstrated during other major events such as the Ebola outbreaks, the Haitian earthquake in 2010, the earthquake in Nepal in 2015, the hurricane Sandy in 2012, or the Boston Marathon bombing in 2013 (Kolokythas, 2021; Radianti et al., 2016). These examples show that misinformation is not limited to one specific hazard nor present only in certain regions but rather worldwide. In contrast to emergency specific misinformation, political and religious fake news is often characterized by long life cycles or polarized debates between different groups (e.g., pro-government vs. opponents), thus much more difficult to combat since it is deeply rooted (Kwanda & Lin, 2020). A hypothetical scenario further showed that people are under higher acute stress when receiving conflicting information during shootings at school (Jones et al., 2017), and people having direct contact with friends and using Twitter were more exposed to such information (Jones et al., 2017).

2.2.5 Dynamics on social media regarding misinformation

On social media, not only private persons and institutes communicate and share information but also online communities have emerged in the last decades playing an important role in supporting citizens and intervention agencies in humanitarian aid distribution (Arora, 2021). Further, social media has been used to perform vital relief functions, damage information provision, support for disabled individuals, and moral support systems (Peary et al., 2012). And in case of lacking authoritative statements, social media platforms help communities to handle an emergency on their own (Fallou et al., 2020).

However, certain dynamics on social media foster the spread of misinformation. For example, algorithmically-curated social media platforms such as Twitter prioritize user engagement over accuracy (Scheufele et al., 2021), sharing emotion-loaded content more likely. Such content in turn is a key driver of nullifying correction effects (Lee, 2022). Further, a study has shown that fake news is about 70% more likely to be shared and six times faster reaches 1,500 people than true information (Vosoughi et al., 2018). In particular bots (=software robots) are "super-spreaders" of misinformation because they can retweet articles within seconds of their first appearance and often use low-credibility sources (Kolokythas, 2021). Further, programs (e.g., Internet Research Agency) create fake accounts, which can influence public's opinion, as we saw during the U.S. Presidential Election in 2015 (Xia et al., 2019). One exemplary account was 'Jenna

Abrams' impersonating a young, white, American woman with more than 70,000 followers on Twitter (Xia et al., 2019). Especially conservatives are more prone to misinformation (Lee et al., 2021).

2.3 Cognitive and behavioural biases

In general, the "dual-process theory" of reasoning says that people have two separate ways of processing information and reacting to the social environment (Evans, 2003): 1) heuristics-based decision making (=intuitive thinking) and 2) deliberative thinking. The latter refers to the conscious consideration of the benefits and risks of possible choices. In comparison, heuristics are mental shortcuts allowing people to make decisions quickly and efficiently. These heuristics are often used in complex or rapidly evolving situations such as sudden-onset disasters. In some cases, these heuristics lead to acceptable solutions and in others to cognitive and behavioural biases that result in less good decisions (Broomell, 2020).

Especially low-probability and high-impact (LP-HI) risks (e.g., pandemics, earthquakes) are associated with various behavioural biases, which implies that individual behaviour deviates from rational risk assessments by experts and optimal preparedness strategies (Botzen et al., 2021). The simplification bias explains that individuals view the likelihood of LP-HI events as falling below their threshold level of concern and fail to take risk reduction measures, unless they experience the impacts of a disaster according to the availability bias (Botzen et al., 2021). Or the finite pool of worry denotes that due to the health and unemployment consequences of Covid-19, individuals become more concerned about their health and the economy and less concerned about climate change. Nevertheless, a longitudinal panel survey in the UK reveals little evidence for diminishing climate change concern during the Covid-19 pandemic. The findings suggest climate change has become an intransigent concern within UK public consciousness (Evensen et al., 2021). Further, the 'myopia' bias has the effect that individuals insufficiently value the future benefits from actions that reduce risk from climate change, which also applies to politicians according to the Not in my term of office bias (NIMTOF) (Botzen et al., 2021).

These were only some examples of cognitive and behavioural biases. In Table 2, we list the most important behavioural biases from the literature. We used this list of biases to identify any cognitive and behavioural biases in our six case studies.

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

Table 2 – List of cognitive and behavioural biases; in alphabetical order.

Cognitive & behavioural biases	Description
Anchoring bias	Tendency to reply too heavily – to anchor – on one trait of a piece of information when making decisions. It can be differentiated between four types:
	Common source bias: the tendency to combine/compare research studies from the same source, or from sources that use the same methodologies or data.
	Conservatism bias: The tendency to insufficiently revise one's belief when presented with new evidence.
	Functional fixedness: A tendency limiting a person to using an object only in the way it is traditionally used.
	Law of the instrument: An over-reliance on a familiar tool or methods, ignoring or under-valuing alternative approaches (e.g., if all you have is a hammer, everything looks like a nail).
Availability bias	Individuals underestimate LP-HI risks, such as those related to climate change and Covid-19, until they experience the consequences or learn about friends or family who have suffered from the threat. This underestimation is caused by the availability bias (Tversky & Kahneman, 1973).
Backfire effect	Post hoc corrections of beliefs in misinformation can lead to so called 'backfiring effects'; meaning that people end up believing more in the myth as a result of the correction (Lewandowsky et al., 2012).
	The 'worldview' backfire effect thereby refers to the psychological reactance against the correction itself, and the 'familiarity' backfire effect to the repetition of false information (van der Linden, 2022).
Confirmation bias	The tendency to seek or interpret information in a manner that is consistent with existing beliefs or expectations (Marks & Fraley, 2006). Thus, people weigh information supporting their beliefs greater than information countering them (Baron, 2012).
Cry-wolf syndrome	When receiving repeated false alarms, publics lose trust and confidence in the institutions responsible for the emergency (Santoianni, 2007).
Failure to listen to experts	Despite the warnings by scientific experts, government leaders fail to heed these warnings and take needed actions (Kunreuther & Slovic, 2021).
Finite pool of worry	If concern about one issue increases, concerns about other issues decrease because individuals only have a limited pool of emotional resources (Capstick et al., 2015). This has, for example, been used to explain the decline in worry about climate change after major events such as 9/11 and the 2008 financial crisis where, respectively, worries about national security and the economic situation became more important (Weber, 2010).
Herding bias	Individuals' choices are often influenced by other people's behaviour, especially under conditions of uncertainty due to social norms (Meyer & Kunreuther, 2017).
Loss aversion	The tendency is that people are more willing to take risks to avoid a loss than to make a gain, thus 'losses loom larger than gains' (Kahneman & Tversky, 1979).
Myopia	Myopia describes the behaviour that is related to heavy discounting of future risk reduction benefits and overweighting of upfront costs (Gneezy & Potters, 1997).
Myside bias	People evaluate evidence, generate evidence, and test hypotheses in a manner biased toward their own prior opinions and attitudes (Stanovich et al., 2013), which is linked to the actively open-minded thinking concept (Roozenbeek et al., 2022).

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

Not in my term of office (NIMTOF) bias	Situations where politicians fail to undertake expensive measures to limit low-probability risks that are unlikely to happen in their term of office because they obtain insufficient rewards from voters for limiting the impacts of events that do not occur when they are in office (Kunreuther & Useem, 2009).
Optimism and overconfidence	The tendency of people to see situations as less threatening than they are and to see themselves as more capable than they are (United Nations Office for Disaster Risk Reduction, 2022).
Paradoxical effect	Overconfidence in the civil protection system or technological danger control devices (neglection of self-responsibility), which leads to a feeling of security and neglection of the destructive potential of natural phenomena (Ligi, 2009).
Pseudo-inefficacy	The belief that any personal contribution we make towards reducing risks will be insignificant and thus ineffective (Kunreuther & Slovic, 2021).
Salience bias	The tendency to overemphasize information that is more easily remembered or made salient by a specific environment; thus ignoring information that does not grab our attention (United Nations Office for Disaster Risk Reduction, 2022).
Simplification	Many people use threshold models to decide whether to take protective measures in advance of a potential event or not. For example, people often view the likelihood of LP-HI events as falling below their threshold level of concern, whereby no risk-reducing actions are taken (Slovic et al., 1977).
Status quo bias	The tendency to prefer things to stay the same by doing nothing (inertia) or by sticking with a decision made previously (Samuelson & Zeckhauser, 1988).

2.4 Strategies and initiatives to fight mis- and disinformation

2.4.1 Specific strategies

There are several strategies to avoid or fight the spread of misinformation, whereas the cultural and social context determines whether a strategy is successful or not (Fallou et al., 2022). In Table 3, we summarize some main strategies to fight the spread of mis- and disinformation and to prevent/minimize its impacts on individuals and the society. This is not an exhaustive list since current research is working intensively to develop further strategies addressing the diversity of the mis- and disinformation types and dynamics. Besides these strategies, understanding the target audiences' perspectives and perceptions is key to address their concerns and needs with the communication strategies (Dallo et al., 2022), taking into account also behavioural biases (see section 2.3). This is in line with Scheufele et al. (2021, S. 524) who stated that "societal and contextual factors, such as information ecologies and volatile science-society interfaces, might be much more powerful drivers of public attitudes and behaviours than misinformation among individual citizens." A further challenge is that the conspiracies change over time as it was visible during the Covid-19 pandemic, starting with mainly fear mongering and then shifting to vaccine-related narratives (Bruns et al., 2022). Thus, the communication strategies must

be dynamic and flexible to address the existing and over time emerging conspiracies. However, for one hazard the misinformation can be bundled into themes that come up again and again, thus one can reuse already developed information materials (Dallo et al., 2022).

Table 3 – Overview of strategies to fight the spread of and prevent the impacts of mis- and disinformation collected from research studies; starting with the strategies which should be applied before misinformation is spread and ending with the strategies to use when misinformation was spread.

	Strategy	Description
before	Applying the psychological inoculation (Prebunking)	For the psychological inoculation, people are first forewarned that they may be misled by misinformation (active immune system) and, secondly, they are exposed to severely weakened dose of it coupled with strong counters (generate cognitive antibodies), also known as the prebunking phase (van der Linden, 2022). This thus is a strategy 'inoculating' people against false information before they are exposed to it (Abrams, 2021).
	Increasing science literacy	Universities, educational institutions, and scientists should increase the public's understanding of science, allowing them to make decisions that are in line with the scientific consensus (Scheufele, 2013). To this end, scientists should actively communicate what is scientifically correct and for what no scientific evidence exists (Dallo et al., 2022). However, not all people trust scientists and for example vaccine deniers actively ignore scientific corrections of invented links (e.g., link between vaccines and autism) (Klimiuk et al., 2021).
	Increasing self-efficacy	One should disseminate educational material to educate people about how to self-verify information and to increase their awareness of their responsibility to personally make sure that the information they share is accurate and make photographic evidence before sharing the information on multiple social computing platforms (Flores-Saviaga & Savage, 2021). Citizen-driven approaches have shown to be successful in involving the public and jointly fighting misinformation (Flores-Saviaga & Savage, 2021).
	Increasing trust	In some countries the prevalence of misinformation is limited, thus interventions aiming at increasing trust in reliable news sources have a higher effect on the overall quality of the information environment compared to interventions reducing acceptance of spread of misinformation (Acerbi et al., 2022).
	Using clear and consistent wordings	Authorities or other institutions communicating information to the public should avoid the use of ambiguous wordings and provide clear, understandable and accessible information (C. Zhou et al., 2021), using plain language with little to no jargon (Bautista et al., 2021). Further, consistent messages across the different channels is a must to avoid confusion and, in turn, inaction (Weyrich et al., 2019).
during & after	Correcting messages with bots	Fake News bots should classify information on social media and inform users about possible fake news (Flores-Saviaga & Savage, 2021), using corrective algorithms, keywords and hashtags (Bode & Vraga, 2018). Ozturk et al. (2015) for example showed that warnings such as 'this posting may contain misinformation' can decrease users' willingness to repost it.
	Cooperation (media & government)	When high-risk news is spread, it was shown as successful to fight it when the news professionals treated the government as the authority to debunk it, presenting news only after official clarifications; as it was done during the tsunami-genic earthquake in Palu in 2018 (Kwanda & Lin, 2020).
	Debunking	An effective debunking message starts with the facts, presenting them in a simple and memorable fashion. The audience should then be warned about the myths once. Thereby, one should also address the manipulation technique, thus explaining why the myth is misleading. The message should end again with the fact, emphasizing the correct explanation (van der Linden, 2022).

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

Empowering professionals to correct it	Professionals trying to correct misinformation on social media can go through a stepwise procedure: identification and authentication of misinformation, correction preparation and correction dissemination. So, experts verify the truthiness of information, try to understand why people believe the information, prepare a corrective message including reputable references and share it on trusted and credible channels (Bautista et al., 2021).
Providing accurate information	Reducing the sharing of misinformation is possible by providing accuracy assessment and rationale (Jahanbakhsh et al., 2021), coming from a credible source (Oh et al., 2010). This allows people to share the official, verified statements within their network correcting possible false information.
Providing tips to recognize it	The government (or other authoritative/trusted institutions) should provide tips to recognize misinformation and share reliable and accurate information sources; as the Mexico's presidency successfully did after the earthquake in 2017 (Flores-Saviaga & Savage, 2021).
Questioning the source	One should try to not question the message itself but the source because people do not want to feel like they are being manipulated (Brumfiel, 2022). But by making them aware of the inaccurate source and in parallel providing accurate information from an official source may let them critically question the correctness of the other message.
Removing them from the platforms	Social media providers should implement screening tools to detect misinformation quickly and, then, either act against it oneself or inform the responsible authorities who can provide accurate information (C. Zhou et al., 2021). However, decision makers or actors in charge of moderating online content have to find a balance between freedom of expression and preventing harm. A study in the U.S. showed that the majority of the public prefers removing posts and suspending accounts if the consequences are severe and repeated offense over protecting free speech (Kozyreva et al., 2022).

2.4.2 Initiatives

Various initiatives on a national, European, and international level provide resources to prevent and fight misinformation such as the following ones:

- Code of Practice on Disinformation to curb manipulative behavior to spread disinformation (e.g., fake accounts, bot-driven amplification) and establish stronger collaboration among signatories to combat the challenges associated with such techniques.
- Rapid Alert System (RAS) to facilitate the sharing of insights related to disinformation campaigns and coordinate responses among EU institutions and Member States.
- European Media Literacy Week to underline the societal importance of media literacy and exchange good practices across the EU.
- Fact-checking tools to verify the accuracy of a fact (e.g., Snopes, Truly).
- Tailored campaigns in different EU countries to debunk the local variations of the so-called 'Euromyths' (e.g., the EU determines whether the clocks are changed every autumn or not); such as Euromyty in Slovakia.
- The establishment of a Task Force that reports on and analyses disinformation trends, debunks disinformation narratives, and raises

awareness of the negative impacts of disinformation originating from pro-Kremlin sources.

In the context of CORE, the scientists of WP 7 contributed to the Global Digital Compact by providing a list of key methods to combat the spread of dis- and misinformation derived from WP 7 research. These methods are addressing:

- The relationship and balance between Internet freedom and spread of misinformation/disinformation.
- Classifying information as misinformation/disinformation and censorship decisions.
- Control of and rules for misinformation/disinformation identification and potential censorship.
- Algorithms of misinformation/disinformation identification and potential censorship

Thus, preventing and fighting misinformation is a combination of mitigation measures ranging from self-regulation and control and responsibility rules to preparedness and response efforts of official actors.

3 Methodology

3.1 The six case studies

To understand the dynamic spread of and belief in misinformation, it is important to understand the entire communication chain. The **source** of the information is often decisive whether people believe in the information or not (Jahanbakhsh et al., 2021). Or the content of misinformation **messages** is often more attractive and present on multiple **channels** and, thus, read by more people on various platforms (UNESCO, 2022; van Prooijen & Douglas, 2018). Further, the personal and social factors of **receivers** influence how they react to misinformation messages (**effect**), i.e. past experiences supporting a false statement (Jahanbakhsh et al., 2021).

We thus defined six case studies which address various factors regarding misinformation along the communication chain. An overview of the case studies is listed in Table 4, and a detailed description for each of them is provided in chapter 4.

Table 4 – Source, method, sample size, period, and hazard-focus of the six case studies

Case studies	Data source	Method	Sample size	Period	Hazard

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

Misinformation about the link between Covid-19 and 5G on Twitter	Twitter [English]	• Natural Language Processing methods • RoBERTa • Quantitative analysis	N=331,448	1st January 2020 to 31st December 2021	pandemic
Misinformation about earthquake predictions on social media	Twitter [English]	• Natural Language Processing methods • RoBERTa • Ordinary least squares time series model • Quantitative and qualitative analysis	N=82,129	1st March 2020 to 31st March 2022	earthquake
Fake news about the volcano Vesuvius on general news media	Local online news media [Italian]	• Media analysis • Interviews	N=130 articles	2012-2022	volcano
Mining the discussion of Monkeypox misinformation on Twitter	Twitter [English]	• Natural Language Processing methods • RoBERTa • Quantitative analysis	N=1,440,475	1st May to 24th August 2022	epidemic
Misinformation and the role of media after the Manchester Arena attack	Twitter, worldwide press [English]	•Natural Language Processing methods •Descriptive statistics •Sentiment analysis	3505 press articles 89147 tweets	22nd May 2017 to 13th March 2023	terrorist attack
Authoritative policies to increase societies' resilience to earthquakes - a cross-cultural comparison	Authoritative documents, nation-wide public surveys [Israeli, German]	• Descriptive case study comparison	Surveys CH: N=596 IL: N=920	2020-2023	earthquake

3.2 Link to other CORE tasks and European projects

As summarized in the beginning, the insights from the CORE past scenarios built the framework (communication chain) of our specific case studies addressing different misinformation. Two of our case studies are also part of the CORE past scenarios, namely the Manchester bombing and Covid-19. For the other case studies focusing on volcanoes, Monkeypox, and earthquakes, we used more recent events since social media has evolved drastically in the last years and we wanted to make justice to that. This ensures that our recommendations are

tailored to the current dynamics of social media and, thus, are valuable for authorities or other institutions trying to fight misinformation nowadays.
In collaboration with Task 6.3, Task 7.4, and Task 7.5, we also discussed the role of emerging technologies such as Artificial Intelligence (AI) or automatic bots (Vicari & Komendatova, 2023). Thereby, it is crucial that the societal and ethical aspects are taken into account to ensure that the current digital gap does not increase further, for example (task 6.3). Further, the partners responsible for task 7.1 provided us with the relevant data, and our task 7.2. outcomes are relevant for the subsequent tasks in WP7, which will develop specific tools to fight misinformation (tasks 7.4 and 7.5).

Regarding the link to other European projects, we built on the work done in the context of the two Horizon-2020 projects 'Real-time earthquake rIsk reduction for a reSilient Europe' (RISE) and 'Towards more Earthquake-resilient Urban Societies through a Multi-sensor-based Information System enabling Earthquake Forecasting, Early Warning and Rapid Response actions' (TURNkey). Three authors of this deliverable led a joint effort of these two projects to analyse the most common earthquake myths. The insights from the expert elicitation as well as the developed Communication Guide for professional (societal) stakeholders led the fundament for our case studies (Dallo et al., 2022; Fallou et al., 2022).

4 Results – The six case studies

In chapter 4, we summarize the case studies (Table 4 & Table 5), following the same structure:

1) Problem statement and overarching research question
2) Method
3) Main results
4) Conclusions

The results are further divided into four foci which address the objects of task 7.2. These foci are i) dynamics, constellations, and networks; ii) short- and long-term effects on public's risk perception, attitudes, and biases; iii) risk communication and management (role of authorities/official narratives); and iv) multi-hazard context. Further, we looked at different hazards ranging from short-lived (e.g., earthquakes, bombing) to more long-lived events (e.g., Covid-19, volcano, Monkeypox). From these insights we then derived recommendations for each step of the communication chain (section 5.2).

Table 5 – Overview of the case studies, their related publication, and the involved partners.

Case Study	Publication link	Involved partners
Misinformation about the link between Covid-19 and 5G on Twitter	Elroy, O., & Yosipof, A. (2022). Analysis of Covid-19 5G conspiracy theory tweets using SentenceBERT embedding. In E. Pimenidis, P. Angelov, C. Jayne, A. Papaleonidas, & M. Aydin (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2022* (pp. 186–196). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-15931-2_16	CLB, ETH, IIASA
Misinformation about earthquake predictions on social media	Dallo, I., Elroy, O., Fallou, L., Komendantova, N., & Yosipof, A. (2023). Dynamics and Characteristics of Misinformation related to Earthquake Predictions on Twitter. *[Under review in Scientific Reports]*	ETH, CLB, EMSC, IIASA
Fake news about the volcano Vesuvius on general news media	[in preparation]	ISSNOVA
Mining the discussion of Monkeypox misinformation on Twitter	Elroy, O., Erokhin, D., Komendantova, N., & Yosipof, A. (2023). Mining the Discussion of Monkeypox Misinformation on Twitter Using RoBERTa. In: *19th International Conference on Artificial Intelligence Applications and Innovations*. Springer. https://doi.org/10.1007/978-3-031-34111-3_36	CLB, IIASA
Misinformation and the role of media after the Manchester Arena attack	[in preparation]	IIASA, CLB
Authoritative policies to increase societies' resilience to earthquakes - a cross-cultural comparison	[ready for submission]	ETH, CLB

4.1 Case Study I – Misinformation about the link between Covid-19 and 5G

4.1.1 Problem statement and research question

On social media such as Twitter, information is distributed to a wide range of people around the world. Despite the positive information exchange (e.g., new scientific findings, authoritative information about an emergency), conspiracy theories in different subjects emerge too. The pandemic Covid-19 has been the latest emergency where misinformation was (un)intentionally shared around the global and affected people's behaviours, and thus effective response to save lives. In the context of the pandemic, the World Health Organization then also introduced the term infodemic: "A global epidemic of misinformation - spreading rapidly through social media platforms and other outlets - poses a serious problem for public health (Zarocostas, 2020; p. 395)."

In our study, we chose one of the main Covid-19 conspiracy theories, namely claims that there is a link between 5G and the spread of the virus and the

strength of people's immune system (Elmousalami & Darwish, 2021). Some people believing in the conspiracy reacted negatively and burnt down expensive infrastructure (Ahmed et al., 2020). We thus wanted to analyse the dynamics and characteristics of the conspiracy theory tweets as well as to understand who is supporting and who opposing it, in order to develop effective strategies to fight the claims and, consequently, minimize inappropriate behaviours.

4.1.2 Methods

The data was collected using Twitter API's v2 full search endpoint, which is limited to academic research, and the search query: [covid OR coronavirus OR corona] AND [5G]. In total, we collected and analysed 331,448 tweets related to the discussion on the Covid-19 5G conspiracy January 1, 2020, to December 31, 2022. Of those we hand-labelled 4,291 tweets into 'supporter', 'opponent', and 'neutral/irrelevant'; e.g. supporter. "You can get Covid through 5G".

We then tested five different models and found out that the Voting Ensemble, with both sets of features (sentence embeddings using CT-BERT and SBERT and external features, performs best. We thus applied it to classify the unlabelled dataset, leading to 64,080 supporter tweets, 108,175 opponent tweets, and 159,193 neutral/irrelevant tweets.

4.1.3 Main results aligned to the four foci

Dynamics, constellations, and networks

Regarding the number of tweets, we identified that there are 69% more tweets opposing the conspiracy than tweets supporting it. Further, we show that there are 2.5 times more users opposing the conspiracy (n=89,030) than supporting it (n=35,169). However, supporters on average tweet 1.82 tweets while opponents only 1.22.

Regarding the timely evolution, we identified that when the conspiracy theory emerged in February 2020, significantly more supporter tweets were published. Two months later, it switched and tweets for opponents dominated the discussion on Twitter.

Regarding the content and structure of the tweets, most tweets from supporters and opponents had a neutral sentiment score. In comparison, around 50% of the supporter tweets contained a URL, while only 28% of the opponents used URLs for their tweets. Thus, they include less evidence to refute the conspiracy.

Short- and long-term effects on public's risk perception, attitudes, and cognitive biases

The fact that tweets from supporters of a conspiracy theory contain more often a URL can lead to the common source bias. Thus, people compare arguments from the same source/data origin and, consequently, do not have a balanced information intake.

Further, the confirmation bias may also be an explanation for people believing in the conspiracy theory regarding the spread of Covid-19 and 5G technologies. This means that people who were already critical of 5G technologies are more likely to support this conspiracy theory.

Risk communication and management (role of authorities/official narratives)

From the results, we derived **three main recommendations** for policy processes dealing with the prevention and limitation of the spread of online misinformation. The first one is about dynamics of the spread of misinformation, the second one is about the profiling of social media users, and the third one about regular communication efforts.

The dynamic of tweets during various phases of the Covid-19 pandemic shows that the first phases which are connected with large existing uncertainties and the lack of data are also marked by the broad spread of conspiracy theories. During this phase the supporters of conspiracy theories are much more active than opponents, which changes over time when better data is available or there is less media attention. **The first recommendation thus is that, especially during the first phases of a crisis when information is scarce, efforts are needed to deal with misinformation in social media**.

While looking at the profile of supporters and opponents of the conspiracy theories, it becomes evident that supporters are more active (i.e. posting more tweets) and more often use URLs to support their messages, which may also make them more influential. At the same time, opponents are more numerous but their tweets are frequently missing a link to base their statement. **The second recommendation thus is that authorities or responsible institutions should provide evidence and links which could be used by the opponents of the conspiracy theories.**

To effectively fight misinformation when it is spread during or after an event, authorities or civil protection should have a well-established relationship with the citizens. This ensures that people trust in the authoritative statements and corrections of misinformation. To this end, **we recommend medium- and long-term communication efforts such as**:

- **Introduction of specific teaching modules in schools (e.g., how to verify information in social media).**
- **Regular information events with citizens, especially at a local level, to build relationships between authorities/civil protection and citizens.**

Multi-hazard context

This study did not look at different hazards. However, it stresses the fact that different conspiracies can emerge in the context of an event/emergency. Therefore, authorities need to be flexible and first address the misinformation which might have the highest potential to cause harmful consequences. Further, the relevance of one conspiracy changes over time, thus always different theories are dominating the debates on social media.

4.1.4 Conclusion

This study presents a workflow, which allows one to continuously collect, classify, and analyse tweets related to Covid-19 conspiracies. Especially after an event, such an analysis gives a first overview of the spread and focus of the misinformation on social media. With accurate and clarifying information, authorities or other responsible institutions can then provide opponents with evidence they can embed in their tweets against the false claims. This suggests that the fight against the spread of misinformation and its consequences should be a joint effort between authorities and citizens.

4.2 Case Study II – Misinformation about earthquake predictions on social media

4.2.1 Problem statement and research question

The spread of misinformation on social media can lead to inappropriate behaviours, which can worsen a disaster (Pierpoint, 2018). Examples of earthquake events that were affected by the dissemination of misinformation, and thus by behaviours that challenged emergency responses, include: the 2019 Albania earthquake, the 2018 Palu earthquake, and the 2017 Mexico earthquake (Flores-Saviaga & Savage, 2021; Kwanda & Lin, 2020; Mero, 2019). The most common misinformation are earthquake predictions especially after strong events. However, the precise location, time, and magnitude of the next large event cannot be predicted (Main, 1999).

Especially on social media, misinformation is spread around the world in a few seconds after a (severe) event (Kolokythas, 2021), and, thus, understanding the dynamics and involved actors is key to counteract misinformation. Therefore, we analysed the dynamics and patterns of earthquake prediction statements on

Twitter to answer the overarching research question: *What are the dynamics, temporal fluctuations, characteristics, and users of earthquake prediction misinformation on Twitter?*

4.2.2 Methods

The data was collected using Twitter API's v2 full search endpoint and the following search query: [[predict OR forecast OR warn OR updates OR alert] AND [earthquake OR quake OR [seismic AND event] OR seismicity OR shaking OR EQ]]. In total, we collected and analysed 82,129 tweets related to earthquake predictions from March 1, 2020, to March 31, 2022. Of those we hand-labelled 4,157 tweets, and used RoBERTa to classify the complete dataset (Tarunesh et al., 2021).

Regarding the analysis, we did a descriptive analysis of the frequency of misinformation and not-misinformation tweets, used an ordinary least squares (OLS) time series model to analyse the effect of not-misinformation tweets on the spread of misinformation tweets and vice versa, and used statistical independent sample t-test to compare the tweets and the users in the not-misinformation and misinformation groups.

4.2.3 Main results aligned to the four foci

Dynamics, constellations, and networks

Regarding the amount of misinformation, we showed that there are significantly more not-misinformation tweets (i.e., general earthquake notifications and messages countering misinformation) than tweets about earthquake prediction misinformation. Thereby, the daily peaks often correlate; showing that after a major event and during earthquake sequences the spread of earthquake prediction misinformation increases. Thus, earthquake predictions are continuously present on Twitter but with peaks after felt earthquakes.

Regarding the characteristics of the tweets, we found that URLs are more used than media (pictures & videos). Within the not-misinformation as well as the misinformation group, only about 18% to 24% of the tweets in each group contained media, and no significant differences were found. Regarding the usage of URLs in tweets, 79.1% and 84.8% of the tweets in the misinformation group and the not-misinformation group, respectively, contained one or more URL(s).

Regarding the users, there are more users in the not-misinformation group than in the misinformation group. There are no significant differences in the mean number of total tweets posted or in the mean number of users these users follow between the two groups. In comparison, the mean number of users that follow the users in the not-misinformation group is significantly higher than in the

misinformation group. However, the mean number of retweets, likes, and replies in the not-misinformation group is significantly lower than the mean number in the misinformation group. Additionally, the mean number of days since users in the not-misinformation group were created on the platform is significantly higher than that of users in the misinformation group. This latter insight might be an indication for people to verify the trustworthiness of the information.

Short- and long-term effects on public's risk perception, attitudes, and cognitive biases

We identified that increased discussions about past events on their anniversary day are common. For instance, an increase of interest in the discussion can be seen in the beginning of 2021 in relation to the Fukushima earthquake on February 13, 2011. This indicates that severe events are not forgotten and the public feels the need to share the experiences also after 10 years.

Further, we found out that misinformation about earthquake predictions is not only shared on Twitter but also other social media platforms such as YouTube. There are also private earthquake prediction websites which are linked in the posts on social media. The fact that misinformation is present on several platforms can lead to confirmation biases.

We also saw that people struggle to handle multiple emergencies together such as the Covid-19 pandemic and an earthquake. This might be explained by the finite pool of worry, meaning that people have only a finite capacity to worry about one thing and, if concern about one issue increases, concerns about other issues decreases.

Moreover, we saw that misinformation messages contain more negative wordings, pictures, and links to videos. Since this information is more easily remembered (salience bias), people overemphasize this information.

Risk communication and management (role of authorities/official narratives)

We identified that people struggle to understand how certain technologies work such as earthquake early warning systems[1]. Thus, it is indispensable that authorities or institutions responsible for communication with the public counter these misconceptions as the following tweet did: *"EEW systems cannot predict earthquakes, but they can provide up to tens of seconds of warning by detecting*

---

[1] "Earthquake early warning (EEW) is used to describe real-time earthquake information systems that have the potential to provide warning prior to significant ground shaking. This is possible by rapidly detecting the energy radiating from an earthquake rupture and estimating the resulting ground shaking that will occur later in time either at the same location or some other location (Allen et al., 2009; S. 682)."

an earthquake immediately after it occurs". Thus, **it is crucial that system providers have dedicated information campaigns explaining how the system works and that the warnings, in this case, are not predictions**.

Our results further show that event notifications and general information about earthquakes are dominating the debates on Twitter and, thus, people come across correct information with links to accurate websites of official authorities more often. This is important since misinformation then has less value and does not take over the discussion focus. However, there are only a few tweets that directly clarified that predicting earthquakes is not possible (e.g., national seismological services); **thus there is still potential in communicating this more proactively.**

Some tweeters of misinformation link earthquake notifications from official sources in their earthquake prediction claims. **Thus, official sources should regularly check that they are not linked in such tweets and, if they are linked, ask to be removed.**

Moreover, the not-misinformation tweets are a predictor of the misinformation tweets. Thus, when authorities provide information such as earthquake notifications or other information about earthquakes, **they should also be prepared to react to possible misinformation that is spread in response**.

Multi-hazard context
Although the focus was on earthquakes, the discussion was often also about cascading effects such as nuclear power plant accidents (Fukushima), the challenge of natural disasters during Covid-19, and tsunamis triggered by the earthquake. People for example indicated in their tweets to be frightened by Covid-19 and, thus, any further disaster/emergency is an additional burden, which can have negative psychological effects. This shows that misinformation on a particular topic is linked in believers' minds to other misinformation on related topics. Various misinformation consequently can form a system of linked misinformation beliefs; demonstrating the importance of considering misinformation in a multi-hazard context.

4.2.4 Conclusion
The internet and social media have recently made it simpler for misinformation messages to spread quickly and widely around the world, which has had important social and political repercussions. This study offers insights into the dynamics of general earthquake notifications and misinformation messages related to earthquake predictions on Twitter and, thus, should support communication experts to better understand the dynamics of earthquake

predictions on social media and, consequently, adjust their communication efforts to counteract them.

4.3 Case Study III – Fake news about the volcano Vesuvius on general news media

4.3.1 Problem statement and research question

The volcano Vesuvius, next to the city of Naples in southern Italy, is more than just a natural phenomenon, it has also become a cultural symbol over the last centuries. As a 'social product', Vesuvius is polysemic because it takes on various cultural meanings: it is a symbol of horror and drama, but also of wonder and discovery, of identity and belonging (Gugg 2022). On a scientific level, it is one of the longest-studied volcanoes and one on which many gazes and reflections dwell, not least because the anthropic and demographic pressure at its foot also makes it one of the most dangerous volcanoes in the world. The first volcanic observatory in the world was established in 1848 on the Vesuvius and, even today, the volcano of Naples attracts considerable media attention, as it remains at the centre of many scientific debates trying to assess the possible consequences of a future eruption. To deal with the Vesuvius risk, in 1995 a 'red zone' was identified with 24 municipalities, i.e. those located in the area at the highest risk in the event of an eruption, in which around 800,000 people live. For these reasons, Vesuvius was defined by «Nature» in 2011 as the «*Europe's ticking time bomb*» (Barnes 2011).

As the social sciences explain, risk is not only a mathematical formula, but also a social product because it is subject to historical dynamics, power tensions, and specific beliefs. More precisely, it is a hierarchical concept because societies attribute it according to priorities. Therefore, the social elaboration of risk is the result of several mutually influencing factors. Among these, a particularly important place is that of information and communication or, more broadly, of the mass media. As Sandra Wallman (2001) says, the mass media contribute to the construction of a 'local filter' which relates personal and global threats (the micro and the macro perspective), so we can say that it is precisely this focus by the press that contributes to the social construction of risk. In other words, the media communication of risk relates not only to the different meanings of the action of social groups but also to general and abstract categories.

4.3.2 Methods

The media discourse on risk changes according to places and times and, in the case of Vesuvius, here we have chosen to focus on the last decade (2012-2022) and only on the written Italian press (present on the internet). There are two

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

reasons for this delimitation of the field: temporal and practical. The first aim was to focus on an extended period including also recent years and the second one to collect many articles, but manageable according to qualitative and ethnographic methodologies.

The Italian articles that talk about Vesuvius in the last decade are hundreds, perhaps thousands[2], but our analysis focused on 130 texts that have had a lot of echoes in the local community, because they are widely shared on social-networks and cited in online discussions and in the tangible space. This meant that, in addition to the literary analysis of each text, its reception on social-networks has also been followed and, over time, its social elaboration has been investigated through ethnographic interviews in the field.

These 130 articles have been divided into three categories according to their source: (a) "professional press" (that of officially registered newspapers with national circulation, in their web version); (b) "local press" (that of the Neapolitan region, not always professionally journalistic, but recognized as such by the community); (c) "alternative press" (essentially composed of blogs or websites that do not have an information function, but above all an opinion function, and which are shared a lot on social-networks). The archive of 130 articles refers to three specific cases that occurred over the last ten years: 1) in 2013, when the National Emergency Plan for Vesuvius was updated (after its first draft, in 1995) with the perimeter of a new 'red zone' (Gugg 2019); 2) in 2015-2016, when the Civil Protection and the Campania Region presented the Evacuation Plan (Gugg 2023); 3) in 2017, when a great wildfire broke out in the summer, burning a third of the entire territory of the Vesuvius National Park (Gugg 2021).

Vesuvius is a topic towards which there is always a lot of media attention, so the quantity of texts is potentially enormous and can be investigated through various methodologies. In this case, an anthropological-qualitative methodology was chosen, weighing the choice of words used in the headlines and articles, the more or less sensationalist tone and the soundness of the statements made in each article. To understand the social effects of the articles and their language, the analysis made use of data collected ethnographically (field interviews) and by netnographic observation (systematic monitoring of online discussions in correspondence with the most shared articles).

---

[2] The total number is uncertain because complete monitoring of all the Italian press is impossible.

4.3.3 Main results aligned to the four foci

Dynamics, constellations, and networks

Vesuvius is a 'bombshell' news because it is a subject of interest, a protagonist who makes the audience. But precisely this enormous visibility also makes it an issue subject to misinformation and disinformation which, as stated in a study by the World Economic Forum in 2013, is «one of the main risks for modern society». The news relating to Vesuvius archived for this study are almost never 'fake' or 'completely fake', they are more frequently 'exaggerated' and 'exasperated'. The disinformation of Vesuvius is therefore not in the substance of the news, but mainly in the tone, that is, in the sensationalism and in the alarmist language. This characteristic is favoured by the concept of risk itself, which is a possibility, a probability, therefore it always has a certain amount of uncertainty. There will be an eruption of Vesuvius, but we do not know exactly how it will be or when it will happen, so we can hypothesize everything, at least on mass-media and social-networks. The most extreme news, among those archived, have a conspiracy approach (such as: «the Italian state does nothing» or «the Americans know, but they don't warn us») or they are apocalyptic (and then they speak of millions of deaths in a few minutes), but they are not the most numerous, nor the most frequent.

In the case of Vesuvius, misinformation is more often an "alteration of information" that proceeds in a circular way: (a) the production on the web of ('alarmist') articles/posts, (b) it involves a search for an audience (often through the clickbait method), (c) which takes place via social-media (especially Facebook), (d) for which there is a return to articles/posts on the web (with related articles of denial or unmasking by debunkers). Several variants can be found in this continuum: the *hoax* (the farce-news), the *fake-news* (in Italian 'bufala', which aims to oppose any unveiling, aiming at ambiguity, at the swamp between the true and the false), the *frovocation* (the false provocation), the *absurd-but-credible news*, the *product of fiction*, the *anti-news* created to make people laugh and not to misinform. To all this must be added further forms of manipulation and adulteration of messages, often deriving from simple sloppiness, carelessness, and superficiality. In any case, these are all ways of polluting information.

News of new discoveries or interviews with experts in the geophysical sector are often reported by the generalist press in a simplified way. To a certain extent it is obvious, but a share of 'metaphor' is added, leading to Vesuvius being described as

«the sleeping giant»[3][1], «a powder keg»[4], «the biggest problem we have»[5], «the most dangerous volcano in the world»[6]. Furthermore, there are examples in which the results of some recent scientific research are presented in a particularly striking manner:

- «*Vesuvius and Campi Flegrei, the expert: "Eruption closer than you think. The invisible volcanoes…"*» (two volcanologists hypothesized a devastation within a few decades)[7].
- «*Vesuvius, the (flop) evacuation plan. Here are the estimated dead in the apocalypse*» (a volcanologist says that the future eruption could "instantly" kill 3 million people)[8].
- «*When Vesuvius decides to start its next eruptive cycle, it will start with an explosion*» (a geologist at the Accademia dei Lincei explains that the time to escape will take a few minutes, at most a few hours)[9].

Finally, to arrive at the more or less apocalyptic headlines spread with great frequency by local web journals in search of an audience, widely shared through social media:

- «*Vesuvius, one million people at risk: the red zone is expanding*».
- «*Vesuvius: more alarms, this time from Japan*».
- «*Alert from the USA: "Vesuvius will explode and kill one million people in 15 minutes"*».
- «*Vesuvius risk lands on RAI* [national television], *the Italian experts: "There will be no escape"*».
- «*Vesuvius, eruption danger: who to believe?*».[10]

In many cases the mass-media underline the lack of a "culture of risk" and of adequate preparation of the population for the possibility of an eruption. The media *hype* on the subject means that the local population is, in fact, informed in one way

---

[3] «Corriere della Sera», 22 June 2018.

[4] «Il Foglio», 2 November 2016.

[5] «Il Denaro», 18 December 2020.

[6] «Corriere dello Sport», 21 December 2021.

[7] «Libero», 18 November 2018.

[8] «Il Tempo», 19 October 2014.

[9] «Il Mattino», 22 June 2012.

[10] Headlines appeared, respectively, on the following webjournals: «Rete News 24» (28 June 2013), «Corso Italia News» (5 September 2013), «Il Fatto Vesuviano» (25 November 2013 and 3 December 2013), «Positano News» (January 10, 2014).

or another about the risk associated with the volcano and that they are constantly reminded of this.

Short- and long-term effects on public's risk perception, attitudes, and cognitive biases

More than disinformation in the strict sense of the word, it is this type of sensationalist language that causes anxiety and can therefore be considered as a form of pollution of information and society: it creates suspicions and contributes to mistrusting institutions (especially scientific ones) and thus it induces citizens not to participate, not to contribute to collective protection. The exasperation of the words and the repetition of alarmed but unjustified tones creates background noise, hence disaffection and disinterest. False alarms make people less receptive to real alarms: from "alarm fatigue" we move towards the "cry-wolf syndrome", which consists in the lowering of the sensitivity to the alarm and in the consequent response delay when the real alarm is raised.

Contrary to the stereotype that wants the Vesuvius inhabitants to be indifferent to risk or fatalistic, the media hype on the subject means that the local population is in fact, in one way or another, informed and constantly reminded about the risk linked to the volcano. The result, however, is probably different from what is expected, as is well exemplified in the words of a 92-year-old woman interviewed in March 2015 in a town in the Vesuvian red area:

- Question: «*Do you like Vesuvius?*»
- Answer: «*Oh, I like it... Well, I mean... But do you know that I can't see it when I see it on television? I feel bad when I see... I don't know, I think about when it's erupting*».

More generally, the people interviewed by Giovanni Gugg during his periodical ethnographies over the last decade often allude to the alarmist tones used for Vesuvius by the press and television:

- «*This news arrives suddenly... "You know, the article came out in the newspaper", or "Television made the documentary that...". My sister, who is always anxious, says, "But I always say, we have to get out of here, we have to buy a house somewhere, we have to sell everything". And for a while we only talk about this. We talk, we talk, we talk... We talk! But then in the end we all know it, we realize that we are talking about it, but we are talking about it for the sake of talking*». (Female, 50, March 2015)
- «*There are moments in which, moments like the earthquake in L'Aquila or moments in which there are rumours about a catastrophe or any*

emergency… moments in which there is turmoil around volcanic activity… discussions begin in the area or in the newspaper, or on television, on the news. For a few days, for a few days we talk about the "Vesuvian danger" and then we return to normal life, to everyday life». (Male, 45, January 2013)

Thus, a real paradox of disasters is generated, that is, an ambiguous relationship is created between science and information, between reassurances and alarms, which disorients and, in the specific case, which blocks and suspends every decision. The residents of the red zone know about the risk, but at the same time they don't know: they see and don't see, they scotomise to keep anxiety under control, somehow. So, strictly speaking, they are not ignorant, they are not indifferent, they do not lack a "culture of risk", but they are suspended: they rely on expert knowledge, but at the same time they are sceptical of it. For example, a woman says about the small earthquakes that are occasionally felt on the volcano: «*You feel those shocks and you think: "What are these? Do I know about Vesuvius or not?" Then you think about it and you reassure yourself and say: "Okay, but could it be that they don't tell us anything?"*». (Female, 62, May 2015)

Like the inhabitants of La Hague, questioned by Françoise Zonabend (1989) about the nuclear waste plant present in their territory, so the residents of the Vesuvian red zone that Giovanni Gugg interviewed need «to be reassured by being told that all the precautions have been taken and that there is nothing to fear».

Risk communication and management (role of authorities/official narratives)
The most recurring theme in information about Vesuvius concerns the emergency plan and its feasibility, reliability, and knowledge among the population, for which the voices of scientists (the volcanologists of the Vesuvius Observatory), technicians (the operators of the Civil Protection), and politicians (the national government and local administrators). This occurs above all in professional journals, while other voices are also cited in 'alternative' publications, those of 'experts' (often foreign, such as American or Japanese scientists) who announce catastrophic eruptions, millions of deaths and, essentially, an apocalypse (which is a rather recurring term).

To give more strength to the news, the articles often refer to the 'experts': they are i) scientists of the Vesuvius Observatory (which is based in Naples and is, obviously, the most important research centre for the knowledge of Vesuvius); ii) scientists of INGV (the national institute of geophysics and volcanology); or iii) scientists who institutionally dialogue with civil protection and with politicians, but sometimes 'dissident' scientists are interviewed (but rather media-oriented, such as G.M.): they

are also structured in Italian academic research centres, but they have made themselves known for their differing opinions.

- «*Vesuvius alert: "The eruption could be a catastrophe. We risk the Apocalypse"*»
- «*Vesuvius, volcanologists' alarm unleashes controversy. Civil Protection: "Emergency plans exist"*»
- «*Scientists against the Vesuvius-Campi Flegrei eruption risk: G.M. delivers a complaint*»
- «*Against Vesuvius and Campi Flegrei, inadequate evacuation plans. We are at the rescue who can*»[11].

Other times, foreign scientists are quoted, as if this gave them greater authority: "American experts" or "Japanese experts", who always sound the alarm about how dangerous Vesuvius is.

The case of an Italian professor (F.D.) who teaches in the United States is very interesting: periodically, from 2012 up to the present day, one of his rather alarmist statements is used to always publish the same article, in which the expert (who he is an engineer, not a volcanologist) says that the next eruption will have enormous power and that 1 or 2 million people will die in 15 minutes. This professor exists, but there is no trace of his presence on the New York University website and he would not appear to be a lecturer or a structured researcher. Above all, he does not appear to be linked to any scientific community, so he almost appears as a "lone hero against the system" like in movies. Beyond the declaration, whose origin is difficult to trace, it is interesting to see its effect on the web journals, which reproduce it almost every year, always the same, as if it were recent, instead it has been circulating for over 10 years.

- «*Sooner or later Vesuvius will explode and destroy everything, says F.D.*»
- «*Vesuvius will kill a million people, says F.D.*»
- «*Vesuvius, one million at risk, says F.D. The red zone is widening*»
- «*Vesuvius, F.D.'s shock words, "it will be a catastrophic eruption"*»
- «*Vesuvius, the latest alarm from F.D.: "Towards an apocalyptic scenario"*»
- «*"Vesuvius? Here's when it will explode", the expert forecasts*»[12]

---

[11] Respectively from the following web journals: «Fidelity News» (13 October 2014), «Il Mattino» (22 August 2015), «Il Mediano» (25 August 2015), «Napoli Today» (19 May 2023).

[12] Respectively from the following web journals: «MeteoWeb» (28 June 2012), «UniStudenti» (29 June 2012), «ReteNews24» (28 June 2013), «InMeteo» (15 July 2013), «Il Mattino» (26 November 2018), «InterNaples» (23 May 2023).

Another interesting case concerns a Japanese professor, N.S., one of the top volcanologists in the world, who on September 5th 2013 gave an interview to the Italian press[13] in which he said two main things. First, Vesuvius is active, thus sooner or later the volcano will explode again. Second, Italians should talk about it more. On the one hand, N.S. specified that the current scientific knowledge regarding the Neapolitan volcano is advanced, but has some gaps on the possibility of predicting when an eruption will occur (i.e. the timing is a general limit of volcanology). On the other hand, N.S. observed that today's political and social attention to the Vesuvius risk is insufficient given that the emergency and evacuation plans would be substantially inadequate.

Some Italian volcanologist colleagues responded to that interview, such as a manager of INGV, P.P., who specified in the «Corriere della Sera» that «the activity of Vesuvius is constantly monitored and there is no sign that it is changing» and that «the plans for mapping the dangerous and evacuation zones exist, but perhaps there is a lack of information for citizens, which we intend to improve shortly»[14].

From these learnings, we derived some **specific recommendations**:

- **Journalists should create spaces for experts/scientists to voice their expertise and communicate evidence-based science.**
- **Journalists should interview scientists and experts from the responsible institutions (e.g., Vesuvius Observatory, INGV).**
- **Scientists/experts should make journalists/the media aware of self-proclaimed experts who in reality have no expertise in the respective field (in this case vulcanology).**
- **International (foreign) experts should not make 'alarmistic' statements about a threat in a country/region where they do not live because they are unfamiliar with the local dynamics and they reactions they might trigger.**
- **Information campaigns on the content of the emergency plan should be organized at a municipal level, for which appropriate resources should be allocated. Although mayors have been responsible for communications since 2021, they should be supported by higher-level authorities because they already have many duties in their day-to-day business.**

---

[13] «Corriere del Mezzogiorno», 5 September 2013.
[14] «Corriere della Sera», 6 September 2013.

Multi-hazard context

During the months of June and July 2017, a series of fires burned one third of the surface area of the Vesuvius National Park, established in 1995. An event totally attributable to human actions whose contingent causes, direct culprits and possible motivations are unknown (Gugg, 2018). At times, especially in the upper area of Torre del Greco, the flames came dangerously close to some houses, forcing residents to evacuate, in an atmosphere of bewilderment and anger. In the excitement of those hours, one of the first explanations for the disaster that spread on social media was that of the "incendiary cats" or "fire cats", an urban legend according to which the criminals would have used live cats, sprinkled with petrol and then set alight to spread the fire.[15] Although denied by the Carabinieri Forestali and declassified as "fake news"[16], the rumour was widely circulated, managing - at least for a certain number of days - to trigger anger and frustration towards unknown men, considered diabolical because terribly wicked:

- «*I have no words, having so much wickedness can't be human, I refuse to think so… it's from Satan in person*»;
- «*My God, it's terrible. To these accursed ones the same treatment they reserved for those poor puppies*»;
- «*No, that's enough, now I'm not a good person anymore! Now, I have become a beast! And I wish you, yes, you cursed who have made this inhuman gesture, to die! And die slowly and suffering the pains of hell!*»;
- «*In addition to being criminals, they are also sadists and scoundrels, where prison is too comfortable a place for them.*»[17]

Within the news archive considered for this study, this one of the "fire cats" is the only case of complete 'fake news' and brings to mind an ancient Vesuvian popular legend, according to which demons carried hay to the top of the crater to feed the energy of the eruption, and that in a 1632 version that volcanic fire was set off by animals full of sparks: «one sees sparks of fire coming out of the animals, either living or dead […], from among the hairs of horses, others even in the

---

[15] Trione M., 2017: Incendio sul Vesuvio: trovati 8 inneschi, sono carcasse di gatti bruciati vivi, «Il Fatto Vesuviano», 12 July.

[16] «Corriere della Sera», 2017, No, i piromani non hanno usato gatti vivi per l'incendio sul Vesuvio, 12 July.

[17] These are comments left under a post on Facebook published by the "Carabinieri Forestali" account, false but considered plausible by many, on 12 July 2017 at 17:17: https://www.facebook.com/carabinieriforestale/posts/ 1960907107458278

darkness quickly rubbing their cats, or dogs have also seen the sparks of fire»[18] (Castelli, 1632: 38).

Of course, in ten years of journalistic news concerning Vesuvius there are many also linked to other topics, such as for example the seismic tremors on the volcano[19] (almost always of very low magnitude, often imperceptible to humans, but 'visible' through online seismographs), or like the new geological discoveries[20] (for example a 'gas dome' at the bottom of the Gulf of Naples) or, again, when something anomalous occurs[21] (this is the case of a series of bubbles on the sea in March 2015, but it was just a malfunction of an underwater pipeline).

4.3.4 Conclusion

Talking about Vesuvius means talking about a cultural universe, not just about a natural, catastrophic event that will take place in a future that we cannot currently identify with certainty. The invitation ten years ago by the Japanese professor to talk more about Vesuvius must be declined in the right way. There is much dialog and communication about Vesuvius, sometimes too much, to the point that in certain situations a real infodemic is created (Manfredi, 2015), which in turn can lead to confusion. Thus, society rather needs to talk better about Vesuvius. The analysis of journalistic articles of the last decade allows one to understand that the question is not 'how much' we talk about Vesuvius, but 'how' we talk about it and 'who' does it.

As other studies on online information show (Tandoc, 2014; Van der Linden, 2022), headlines and texts of articles are currently often influenced by web analytics. Editors use them to identify strategies for further increasing web traffic, i.e. they are monitored to understand which stories are doing well and, therefore, placed more prominently on the homepage and promoted on social media to attract traffic. Vesuvius is clearly a topic that moves web metrics, so it always has a high probability of generating traffic. However, to be more 'attractive' than its competitors, the title and text must always be 'stronger', 'striking', and 'sensationalistic'.

The journalism sector, threatened by the contraction of its economic capital, has in its online version, which is still dependent on an advertising-based model, the

---

[18] The original Italian version: «si vede uscir dall'animali, ò vivi, ò morti [...], uscir scintille di fuoco trà i peli delli cavalli, altri pur nelle tenebre sfregolando velocemente le gatte, ò i cani hanno anche veduto le scintille del fuoco» (Castelli, 1632: 38).

[19] «InMeteo» (11 February 2014), «Cronache di Napoli» (6 May 2015), «Fanpage» (22 April 2023).

[20] «Corriere del Mezzogiorno» (2 March 2016), «Nanopress» (2 March 2016), «News In» (23 July 2016).

[21] «Fanpage» (24 March 2015), «VesuvioLive» (24 March 2015), «Il Meteo» (3 April 2015).

need to chase online traffic. As Tandoc writes, «to attract an audience no longer loyal to mainstream news, journalism dances defiantly – publishing stories about the wildest celebrities, uploading adorable cat videos, highlighting salacious headlines – hoping to attract the attention, to increase traffic» (Tandoc 2014: 12). But this is a choice that greatly disqualifies journalism, which is instead a profession that relies heavily on respect and reputation.

On a social level, the effect of this shouted and striking language is that the wave of curiosity, concern, and anxiety is momentary, because it then subsides into silence and sinks into a karstic path, which will resurface with the next 'warning'. This is not exactly the "cry-wolf syndrome"[22], but rather it is a form of habituation. The continuous succession of substantially equivalent news turns into a kind of background noise, something whose distant echo is felt but which touches no one or, if it does, lasts just an instant. This attitude recalls that of the inhabitants of the nineteenth-century metropolis mentioned by Georg Simmel in a famous essay on the urban experience: bombarded by too many stimuli, they do not react and, on the contrary, assume a "blasé attitude" (Simmel, 1979). It is not a matter of indifference properly so called, but of detachment, or rather of a form of neutralization or scotomisation, which allows them to keep away the anguish of dissolution (or even just of its eventuality), to avoid what Ernesto de Martino defines "crisis of presence" (de Martino, 2008).

4.4 Case Study IV – Mining the discussion of Monkeypox misinformation on Twitter

4.4.1 Problem statement and research question

The monkeypox outbreak in 2022 resulted in the dissemination of false information on social media platforms. This misinformation can have detrimental consequences, including poor decision making and even loss of life. The speed at which misinformation spreads on social media exacerbates the negative impacts during emergencies. Therefore, it is essential to tackle and combat misinformation on these platforms.

The research sought to analyse the discourse surrounding misinformation about monkeypox on Twitter. It aimed to develop classifiers that can distinguish between tweets that propagate misinformation and those that counter it. Furthermore, the study aimed to gain insights by examining the behavioural patterns and domains used in tweets containing misinformation and counter

---

[22] When receiving repeated false alarms, publics lose trust and confidence in the institutions responsible for the emergency (Santoianni, 2007).

misinformation. The ultimate goal was to provide policy recommendations for addressing and mitigating the spread of misinformation on social media platforms.

4.4.2 Methods

A dataset of tweets related to monkeypox was collected using Twitter's academic research API. The search query included English tweets containing the term "monkeypox" between May 1, 2022, and August 24, 2022, excluding retweets. The dataset was pre-processed, and tweets longer than 350 characters were removed. The final dataset consisted of 1,440,475 tweets from 505,163 users. To train a classifier, 3,218 tweets were hand-labelled into three categories: misinformation, counter-misinformation, and neutral.

For classification, word embeddings were calculated and transformed into sentence embeddings using RoBERTa and Sentence-BERT, respectively. Several machine learning classifiers were tested, including Random Forest, k-Nearest Neighbours, Support Vector Machine, Naïve Bayes, and XGBoost. A Voting Ensemble combining all five models was also evaluated using stratified 5-fold cross-validation.

In addition, a fine-tuned RoBERTa model with a classification layer was trained. The model was evaluated using the same cross-validation approach but with a further split for validation during fine-tuning. The model was fine-tuned for 10 epochs with specific parameters. The performance of each model was assessed based on average F1, precision, and recall scores on the test sets. The fine-tuned RoBERTa model achieved the best results and was used to classify the unlabelled dataset.

4.4.3 Main results aligned to the four foci

Dynamics, constellations, and networks

Dynamics: The results show that there is a dynamic shift in the conversation on Twitter regarding misinformation related to monkeypox. At the beginning of the outbreak, tweets spreading misinformation dominated the conversation. However, during the second peak of new confirmed cases, there was a shift in dominance, possibly indicating the last cycle of misinformation.

Constellations: The analysis of users participating in the discussion revealed that the interest in the discussion on misinformation related to monkeypox is driven by new users joining the conversation. This suggests that the interest is authentic and not artificially created by a small number of users.

Networks: The analysis of domains referenced in misinformation and counter-misinformation tweets provides insights into the networks involved. Tweets countering misinformation frequently referenced authoritative sources such as the U.S. Centres for Disease Control and Prevention (CDC), the World Health Organization (WHO), and established news agencies. On the other hand, tweets spreading misinformation often referenced websites that allow users to upload and publish their own content, such as YouTube. These findings highlight the different networks and sources of information involved in spreading and countering misinformation on monkeypox.

Overall, the dynamics, constellations, and networks discussed in the results shed light on the patterns of conversation, the involvement of different users, and the sources of information related to misinformation and counter-misinformation on Twitter regarding monkeypox.

Short- and long-term effects on public's risk perception, attitudes, and cognitive biases

Short-term effects on public's risk perception, attitudes, and cognitive biases:

Increased awareness of misinformation: The analysis reveals that a significant portion of tweets (30%) discuss misinformation related to monkeypox. This can lead to increased awareness among the public about the presence and spread of misinformation on the topic.

Counteracting misinformation: The presence of tweets countering misinformation, referencing authoritative sources such as the CDC and WHO, indicates that efforts are being made to address and correct the false information. This can contribute to reducing the impact of misinformation on the public's risk perception and attitudes.

Formation of cognitive biases: The dominance of tweets spreading misinformation at the beginning of the outbreak may contribute to the formation of cognitive biases, such as confirmation bias, where individuals are more likely to believe and share information that aligns with their pre-existing beliefs. This can further perpetuate the spread of misinformation and affect the public's risk perception.

Long-term effects on public's risk perception, attitudes, and cognitive biases:

Shift in risk perception: The shift in dominance from spreading misinformation to countering misinformation at the beginning of the second peak suggests a possible decline in the influence of misinformation over time. This shift can

contribute to a more accurate risk perception among the public, aligning it with reliable information from authoritative sources.

Trust in authoritative sources: The frequent referencing of authoritative sources in tweets countering misinformation, such as the CDC and WHO, can enhance the public's trust in these sources. This trust can have long-term effects on the public's attitudes towards health information and their willingness to rely on credible sources for accurate information.

Mitigation of cognitive biases: The presence of tweets countering misinformation and referencing reliable sources can help mitigate cognitive biases by providing alternative perspectives and reliable information. This can encourage critical thinking and reduce the impact of cognitive biases, leading to more informed decision-making by the public.

Overall, the short- and long-term effects of the analysed results suggest a potential influence on the public's risk perception, attitudes, and cognitive biases. The presence of misinformation and efforts to counter it can shape the public discourse and influence how individuals perceive and respond to health-related information.

Risk communication and management (role of authorities/official narratives)
Risk communication and management, specifically the role of authorities and official narratives, can be understood from the following results:

Identification of misinformation: The analysis identifies a significant number of tweets spreading or supporting misinformation related to the monkeypox virus. This information is crucial for authorities to be aware of the presence and extent of misinformation circulating on social media platforms.

Counter-misinformation efforts: The results also reveal a substantial number of tweets countering misinformation related to monkeypox. Importantly, these counter-misinformation tweets frequently reference authoritative sources such as the U.S. Centres for Disease Control and Prevention (CDC), the World Health Organization (WHO), and established news agencies like NBC and the New York Times. This highlights the role of authorities in providing accurate information and actively countering misinformation.

Importance of reliable sources: The analysis shows that tweets countering misinformation tend to reference reliable sources, indicating the significance of authoritative information in mitigating the spread of false information. It suggests that authorities should continue to provide frequent updates on

authoritative websites to support those who are countering misinformation and indirectly contribute to the fight against it.

Platforms hosting user-generated content: The domains referenced in tweets spreading misinformation are primarily websites that allow users to upload and publish their own content, with YouTube being the most frequently referenced platform. This finding emphasizes the need for social media platforms to adopt stricter community guidelines and monitor content more effectively. Authorities can play a role in encouraging these platforms to take proactive measures to reduce the sharing of misinformation.

Enhancing user awareness: To minimize the unaware echoing of misinformation, **it is recommended to notify social media platform users about the sources of content and their credibility**. This can help users distinguish reliable information from misinformation and prevent the inadvertent amplification of false narratives.

Overall, the results highlight the importance of authorities and official narratives in risk communication and management during an outbreak. Authorities should continue to provide accurate and timely information through trusted channels, actively counter misinformation, collaborate with social media platforms to address content moderation, and promote media literacy to empower users to critically evaluate information sources.

Multi-hazard context

The multi-hazard context of the results discussed in the analysis is the outbreak of the monkeypox virus and the associated misinformation circulating on Twitter. Monkeypox is a zoonotic viral disease that can cause illness in humans. The analysis focuses on the spread and counter of misinformation related to monkeypox on social media, specifically Twitter. The following points highlight the multi-hazard context:

Outbreak of monkeypox: The analysis mentions two major peaks in the discussion of monkeypox on Twitter, which likely correspond to the epidemiological evolution of the virus. These peaks align with periods when new cases of monkeypox were reported globally. The discussion of misinformation is situated within the broader context of the monkeypox outbreak.

Misinformation spreading: The analysis reveals that a significant number of tweets, 180,259 in total, spread or support misinformation related to the monkeypox virus. This misinformation poses a risk by potentially disseminating

false information and contributing to public confusion and panic during an outbreak.

Counter-misinformation efforts: In response to the spread of misinformation, there are also tweets countering the false narratives. A total of 152,522 tweets are identified as countering misinformation related to monkeypox. These counter-misinformation efforts play a crucial role in mitigating the impact of false information and providing accurate and reliable information to the public.

User behaviour and participation: The analysis examines the behaviour of users participating in the discussion on Twitter. It notes that the discussion is predominantly driven by users who participate for the first time. This suggests that there is genuine interest in the topic and not just the result of a small group of users artificially generating the discussion. Understanding user behaviour is essential for effective risk communication and management.

Referenced domains: The domains referenced in tweets spreading misinformation and countering misinformation provide additional context. The domains associated with countering misinformation often include authoritative sources such as the U.S. Centres for Disease Control and Prevention (CDC), the World Health Organization (WHO), and established news agencies. In contrast, the domains associated with spreading misinformation are often platforms that allow user-generated content, such as YouTube. These platforms have been linked to conspiracy theories and extreme free speech.

Overall, the analysis explores the discussion on Twitter surrounding monkeypox and the presence of misinformation. It highlights the multi-hazard context of an ongoing outbreak, where misinformation can pose risks to public health and emphasizes the importance of countering false narratives with accurate and authoritative information.

4.4.4 Conclusion

In this study, the focus was on investigating the discussion of misinformation related to the monkeypox virus on Twitter. A RoBERTa model was trained and fine-tuned to classify tweets as misinformation, counter-misinformation, or neutral. The results showed that only one-third of the tweets discussed misinformation, while two-thirds were neutral. The analysis of user behaviour revealed that new participants drove the interest in the discussion.

Regarding countering misinformation, tweets often referenced authoritative sources like the CDC and WHO. The study suggests providing more frequent updates from reliable sources to support users countering misinformation. On

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

the other hand, tweets spreading misinformation frequently referenced platforms allowing user-generated content, such as YouTube. Encouraging these platforms to monitor content and enforce stricter guidelines could help reduce the dissemination of misinformation. **It is also recommended to inform social media users about content sources and credibility to minimize the unwitting spread of misinformation.**

The study acknowledges limitations, such as focusing only on Twitter and a short period of analysis. Future research could expand to other social media platforms and analyse the discussion over a longer duration.

4.5 Case Study V – Misinformation and the role of media after the Manchester Arena attack

4.5.1 Problem statement and research question

The May 22, 2017, Manchester Arena attack received substantial media attention and fuelled public discourse via traditional media and social platforms. Analysing the role and impact of traditional media and social media during and after the attack is vital for understanding information dissemination, public engagement, and the spread of misinformation.

Staton (2020) explores how traditional and social media platforms vary in tone and purpose. Berubé et al. (2020) identify multiple topics discussed on Twitter after the attack, including resilience, reporting, requesting, responding, remembering, and reheating. Zhao and Zhan (2019) investigate the impact of emotional appeal, framing, images, and informal tone on behaviour and engagement on Twitter. Additionally, the context examines hate speech and misinformation related to the attack. Innes (2019) explains three techniques of disinformation and misinformation: spoofing, truthing, and social proofing. Press sources and Reed & Ingram (2020) highlight six fake news or rumours circulated during the attack, including false claims about the police tracking the attacker and allegations against Ariana Grande. Khanum and Khan (2020) reveal a significant presence of hate speech and anti-Islam, anti-Muslim discourse on Twitter following the attack.

The comprehensive investigation of the extent, impact, and response to misinformation through traditional media and social media platforms remains an unexplored area in the existing scientific literature. Additional exploration and analysis are necessary to address the following question: what was the extent, impact of misinformation and response to it through traditional media and social media platforms? This research question aims to explore the presence and

consequences of misinformation surrounding the attack. It seeks to analyse the types of misinformation circulated, the channels used for dissemination, and their effects on public understanding and perception. Additionally, the question investigates how traditional media and social media platforms addressed misinformation. By examining the roles of traditional media and social media, this research question contributes to a comprehensive understanding of information dynamics during crisis situations. It also provides insights into the influence of these media platforms on shaping public narratives, offering valuable lessons for enhancing information integrity and resilience in similar contexts.

4.5.2 Methods

This research encompasses several key steps. Firstly, tweets and press articles were extracted using a predefined set of search keywords relevant to the Manchester Arena attack. The temporal distribution of publications was then compared to understand the timeline of information dissemination. Subsequently, tweets and articles containing keywords associated with fake news, as identified in the literature, were extracted. Descriptive statistics were generated based on the different types of fake news, and a comparison was made between retweet and like counts. Additionally, sentiment analysis was conducted to assess the emotional tone of the collected data. These steps collectively contribute to a comprehensive analysis of misinformation and its impact through traditional media and social media platforms.

4.5.3 Main results aligned to the four foci

Dynamics, constellations, and networks

The dynamics of misinformation surrounding the Manchester Arena Attack exhibit interesting patterns. One notable finding is the divergent temporal distribution between the press and Twitter, except for the first two annual anniversaries. The impact of recent publications, reports, public inquiries, and new regulations, such as the "2022 Report finds 'fatal' errors after Manchester terror attack" and the introduction of Martyn's Law in 2022, is evident in the press but not on Twitter.

In terms of fake news, two types have emerged with significant yet distinct impacts on Twitter. The rumour about children being sheltered in Holiday Inn and the rumour about the bomber being a refugee gained considerable attention and engagement on the platform. However, these same fake news stories had a limited impact in the press.

This disparity underscores the diverse dynamics of misinformation spread and reception across various media platforms. This deeper understanding of

dynamics then allows for an examination of the short- and long-term effects on the public's risk perception, attitudes, and cognitive biases.

Short- and long-term effects on public's risk perception, attitudes, and cognitive biases

The dynamics of misinformation dissemination and reception across different media channels can have significant short- and long-term effects on the public's risk perception, attitudes, and cognitive biases. The rumour about children sheltered in Holiday Inn gained a substantial number of likes and retweets within a short time frame, indicating a rapid impact. On the contrary, the rumour about the bomber being a refugee generated extensive and prolonged debates, accumulating a high number of likes over an extended period (until 2023). The last rumour seems to be more intricate and challenging to address and clarify. Moreover, it was preceded by a conducive environment that fostered the prolonged debate over several years.

These fake news stories had a comparatively limited impact in the press. The detailed information provided in articles allowed for better contextualization, mitigating cognitive biases and avoiding misinformation's negative consequences.

Understanding the short- and long-term effects of misinformation on public perception and cognitive biases is crucial for assessing the broader impact of such incidents. It emphasizes the necessity of implementing effective communication strategies to counteract misinformation and address cognitive biases during the disaster, in the immediate aftermath and in the long term.

Risk communication and management (role of authorities/official narratives)

Based on the research results highlighting the impact of long-term misinformation reinforcing hate speeches in incidents like the Manchester Arena bombing, **policymakers must prioritize evidence-based communication strategies.** To combat narratives linking migration to threats, it is essential to emphasize accurate information, challenge stereotypes, and promote inclusive narratives for social cohesion. **Collaborating with community leaders, organizations, and media outlets is crucial for shaping public discourse and countering divisive narratives.** Policymakers should also invest in educational initiatives promoting media literacy and critical thinking to discern reliable information. **During the emergency, it is crucial to prioritize and dedicate significant efforts to enhance communication and coordination among the public, individuals responsible, and enforcement authorities to ensure public safety**, as stated in the "Manchester arena inquiry" (2022). Additionally, there

should be a focus on promoting education to foster the development of an inclusive society that embraces diversity in terms of race and religion. Implementing these policy recommendations mitigates the negative effects of misinformation, fostering an informed and cohesive society.

Multi-hazard context

The occurrence of the Manchester Arena bombing on May 22, 2017, can be classified as an independent incident. Numerous false information circulated on social media regarding this incident. Nevertheless, it was a singular event and does not fit into the category of multi-hazard.

4.5.4 Conclusion

In conclusion, the Manchester Arena Attack generated significant impacts through the circulation of rumours on social media. The rumour regarding hotels sheltering children had a notable short-term effect in terms of retweets, indicating the rapid spread of misinformation. Conversely, the rumour suggesting the attacker was a refugee had a lasting impact with substantial likes and comments, reflecting a sustained interest and discussion. Furthermore, ongoing investigations are examining the correlation between the temporal distribution of press articles and tweets, offering valuable insights into the dynamics of information dissemination. It is crucial to delve deeper into press articles that reference these rumours, as their full content may provide further understanding and context.

4.6 Case Study VII – Authoritative policies to increase societies' resilience to earthquakes: a cross-cultural comparison

4.6.1 Problem statement and research question

Earthquakes are considered a major hazard, which might have devastating short- and long-term consequences on affected regions and entire countries (Crowley et al., 2021). Therefore, earthquake mitigation, preparedness, response, and recovery policies are needed to improve and ensure societies' resilience and coping capacity. Such policies include building codes, retrofitting efforts, operation of warning systems, or campaigns to increase the public's ability to respond to strong shaking (Marti et al., 2020; Porter, 2021).

Several studies postulate that one policy might be effective at one place but not at another due to cultural, sociodemographic, or political differences. We thus compared two earthquake-prone countries - Switzerland and Israel. This comparison allowed us to identify how countries manage the trade-offs between national and local constraints when implementing various mitigation and

preparedness policies for four dimensions: 1) hazard and risk assessments for locations, populations, communities; 2) mitigation of building and infrastructure collapse risk; 3) warning systems; and 4) hazard and risk communication with the public. Further, we explored citizens' awareness and reactions after an earthquake to identify whether the translation from awareness to protective actions was successfully triggered by the policies (Sechi et al., 2022).

We thus answered the following overarching research question: *What are the similarities and differences between the earthquake and preparedness policies in Israel and Switzerland, and how do the citizens react to those policies?*

4.6.2 Methods

The comparison between the two countries, Israel and Switzerland, is based on their similar seismic hazard levels (Danciu et al., 2021). We applied a case study approach as means to obtain an in-depth appreciation of the seismic risk governance (Crowe et al., 2011), defining the two countries - Israel and Switzerland - as cases (Schoch, 2020). We chose this approach because our findings should enable others to apply the principles and lessons learned to other countries leading to transferability (Schoch, 2020). We compared official documents and websites from the institutions responsible in the two countries to assess earthquake hazard and risk and provide the basis for earthquake policies. Further, we used insights from recent public surveys conducted in both countries to compare the public's risk awareness, attitudes, and preparedness levels (Dallo et al., 2022; Kirschenbaum, Rapaport & Canetti, 2018; Yaar, 2015).

Although the size of the population is fairly similar (ca. 9 million citizens in both countries), the GDP per capita is twice as high in Switzerland compared to Israel. Both countries have advanced emergency response systems, providing healthcare and civil and environmental protection. Further, Israel's population is composed of various ethnic groups and, thus, is more culturally diverse than Switzerland. These similarities and differences are crucial and lead to challenges when considering the capabilities of citizens to comply with the government policies.

4.6.3 Main results aligned to the four foci

Dynamics, constellations, and networks

When providing hazard and risk information, the collaboration between all involved actors is key. To this end, connections must be established and actively maintained. In Switzerland, this network has been successfully built and through for example working groups and regular exchange continuously strengthened. This network also ensures that joint efforts are possible (→ See section Multi-hazard context).

The choice of appropriate communication means should depend on the end-users' needs. We identified that people in both countries prefer to receive rapid earthquake messages on their smartphone (for example on an app); followed by public announcements such as loudspeakers or sirens, and mass media such as radio or TV.

Short- and long-term effects on public's risk perception, attitudes, and cognitive biases

People in both countries tend to believe that the earthquake risk is higher for the entire country than for them personally. This belief can be explained by the optimism cognitive bias "It will not happen to me", which lets people ignore the actual risk. We further identified that people living in hazard-prone regions perceive the risk as higher, which indicates that risk perception is proportional to the actual hazard level at a place. In Switzerland, we additionally showed that people in high hazard regions more often looked for information on how to respond to strong shaking or how to prepare for an emergency.

A crucial focus of communication is to educate people about which actions to take to manage earthquakes effectively. Although information on behavioural recommendations before, during, and after an event may be available and communicated to the public at various occasions, the behaviours need to be internalized to become automatic. The challenge is to facilitate the transfer from just knowing what one can do to behaviours being tacit knowledge. Our analysis shows that people who participate in drills also take these actions during (strong) shaking. Thus, practical exercises/drills at schools and workplaces are indispensable.

Risk communication and management (role of authorities/official narratives)

One relevant issue of disaster risk communication is to provide people with behavioural recommendations for protective actions they can take before, during, and after an event. These recommendations differ from region to region since they depend on, among others, the vulnerability of the buildings. In Israel the first recommended action thus is to get out of the building during the shaking, whereas in Switzerland it is to protect oneself on the spot. This is not only due to the building's vulnerability but also the fact that Israel has an earthquake early warning system in place, which should give people some seconds to get out of the buildings. Thus, the communication also needs to be aligned to the available systems and technologies in a country. Israel additionally also provides recommendations for vulnerable groups such as people in a wheelchair to be as inclusive as possible.

Regarding the communication of event-related earthquake information, we recommend to have a 'single voice' which provides the scientific correct information (e.g., seismological service), and that this information is disseminated via various channels to reach as many people as possible (e.g., civil protection platform, government website). Further, a multi-hazard platform may facilitate the access to information for all relevant hazards in a country/region. To this end, well-established networks between the information provider and the dissemination entities are indispensable.

Regarding the policy implementation approach. The main difference is in the centralization of the decision making authorities. In Israel, the mitigation and preparedness policy is centralized and is being applied nationally, while in Switzerland it is decentralized, and certain policies are implemented only by high-risk cantons. This has led to differences which were found in both countries' citizens' awareness, risk perceptions, and preparedness. In Israel and in Swiss cantons which enforce mitigation actions (i.e. building codes), more citizens knew the instructions, perceived the risk as higher, and in general were more aware of the earthquake risk than those who were not exposed to hazard and/or risk communication campaigns (e.g., drills at school).

Multi-hazard context

In Switzerland, earthquake notifications from a certain magnitude are also distributed via the Swiss natural hazard platform and the Swiss weather app to reach more people. The access data of the weather app actually shows that many users of the app check the earthquake notifications and are successfully forwarded to the website of the Swiss Seismological Service. Several national services further showed that people wish to have a multi-hazard platform where they can access information about relevant hazards. Thus, preparedness but especially messages for immediate hazards should be communicated via dedicated multi-hazard platforms.

Neither in Switzerland nor in Israel are the risks of secondary hazards currently taken into account in the earthquake risk models. However, the interactions with diverse stakeholders revealed that they would like to have this information. For example in coastal regions there is the risk of tsunamis or in mountain areas there is the risk of avalanches and landslides. Thus, future efforts are needed to address also the secondary risks in the model and, consequently, the communication efforts.

4.6.4 Conclusion

Earthquake preparedness and mitigation policies are complicated processes, involving earthquake-resistant building constructions, hazard and risk assessment, and efforts to increase societies' awareness and coping capacities. Given this, there is a need for a collaborative approach, which integrates national forces with local entities in order to increase the capabilities and efficiency of all systems, from the central government to the local authorities, as well as other sectors such as the private sector and civil society organizations.

5 Synthesis and Recommendations

Chapter 5 is divided into two sections. In section 5.1, we provide an overview of the insights from the six case studies aligned to the steps of the communication chain as well as the three main topics of this task in CORE:

> i) Dynamics, constellations, and networks;

> ii) Short- and long-term effects on publics' risk perception, attitudes, and cognitive biases;

> iii) Risk communication & management and the role of authorities/official narratives.

To this end, we organized a virtual synthesis workshop on June 1st, 2023, where representatives of each case study jointly discussed the common insights and derived recommendations to fight the spread of and belief in misinformation (Annex A). The synthesis and the recommendations emanating from this workshop are provided in section 5.2.

5.1 Overview of the insights across the six case studies

In Table 6, we provide an overview of the insights per case study aligned to the communication chain and the three topics defined above. These insights allowed us to derive recommendations to prevent and fight misinformation (see section 5.2).

Table 6 – Overview of the insights from the case studies aligned to communication chain (rows) and divided into three topics (columns). The insights are color-coded based on the case studies: conspiracies about the link between Covid-19 and 5G on Twitter, misinformation about earthquake predictions on social media, authoritative policies to increase societies' resilience to earthquakes, fake news about the volcano Vesuvius on general news media, mining the discussion of Monkeypox misinformation on Twitter, and misinformation and the role of media after the Manchester Arena attack.

		Topics		
		Dynamics, constellations, and networks	Short- and long-term effects on publics' risk perception, attitudes, and cognitive biases	Risk communication & management and the role of authorities/official narratives
Communication chain	Source	• People who believe in the 5G conspiracy theory • People that counter the conspiracy • Bots • Official news agencies • Self-proclaimed experts • People who experiences an earthquake and need to make sense of what is happening • People who already believe in other conspiracy theories • Unvoluntary source because a misinformation provider links official sources • People can be both on Twitter: a source as well as a receiver of the information • Especially during an event, it is key that the information pathways are well established and people know where to access accurate information (trusted source). • Experts • Influencers	• Could increase the use of Twitter as an alternative source of information • Information confusion among the public: often the historical and authoritative newspaper is put on the same level as the 'alternative' web journal • Erosion of trust in risk mitigation measures • A fertile context previously existed for hate speech	• Authorities should provide information on a regular basis • Use official sources and work for them to be identified as such • Partner with other institutions to have a consistent message • Official sources are dependent on the policies and owners of the social media platforms and do not have much influence on the decisions made by them • Single voice people trust. This information can then be spread via various channels • Official sources should not only communicate in case of an emergency but also in quiet times to build credibility/trust and a relationship with the public • In Italy, civil protection does not have social media accounts, although it is beginning to be present. Until now, this void has been filled by others, not always experts or authorities

Message

- People who tent to believe in other conspiracy theories
- People want to refer to reliable sources
- Twitter users who produce original tweets and bots
- The press cites official sources
- Authorities
- Occasionally social media
- Press agencies
- Link between 5G and Covid.19 discussion
- Earthquake predictions (on a daily basis or after a significant earthquake)
- The choice and design of the communications should depend on the end-users' needs.
- The sources talk about the volcanic risk but there is a difference between them, depending on the source and context.
- The national articles are more professional, the local and 'alternative' ones are more striking and conspiratorial.
- There are no 'real' fake news but there are sensationalist and sometimes alarmist tones.
- Monkeypox discussion
- According to the type of rumour, the same message is simply repeated or it generates a rich debate with a wide variety messages
- Press: a rich an detailed context description reduces risk biases

- Dramatic and drastic messages are more 'successful'
- Unique messages, not something trivial, usual or regular
- Messages that would mean drastic consequences
- Salience bias: misinformation is often more attractive (e.g., use of videos) and, thus, better remembered and overemphasized
- Often the tone of the articles is sensationalist or alarmist, so the credibility of the message decreases
- Various messages about the source and spread of
- According to the type of rumour, we observe short-term or long-term impacts
- Simple rumours during an emergency can be fought with clear facts afterwards immediately.
- More complex rumours also targeting other discussed topics in the society (e.g. immigration and violence) are more difficult to fight.

- Build credibility in the quiet times, so that people have trust in the institutions in charge during an emergency and follow the instructions
- Provide a list of reliable sources of information
- Provide context
- Highlight how information is based on evidence
- Label: introduce a label that allows people to recognise reliable sources

- Clear indication of the source
- Clear and consistent message
- State what you know, what you do not know, why you do not know, and, if possible, when you will know it
- Avoid information void
- Understandable and actionable messages, which were tested with the end-users
- Indication of the source to stress the reliability of the message
- Timely information, especially right after the alarm or the actual experience of an event
- The messages must be comprehensible to non-technical people and should not be ambiguous
- Science often has uncertainties but, in the event of a risk, the public wants to know what is happening and what to do.
- Clear, understandable message about source and spread of virus

Channel

- Twitter with link to other platforms (e.g. YouTube)
- Twitter with link to external websites and YouTube
- Misinformation also present on traditional media but only rarely
- Follow a multi-channel strategy to ensure the entire society receives the information.
- The reliable information should come from a single-voice.
- Twitter
- Twitter
- Worldwide press

- Confirmation bias and selection of channel depend on where the false information is spread
- Confirmation bias: the same misinformation on multiple channels can let people believe in it
-
- The impacts of fake news was more limited in the press than on Twitter

- The fragmented information on Twitter can be counterbalanced through external links to richer sources / embedded videos / tweet threads.
- Social media
- Unique website or app for emergencies
- Use channels were misinformation is present if possible (Twitter yes, but what about messaging apps for instance).
- Adopt a multi-channel strategy
- Do not neglect websites where information is available a all time (vs. social media where information is more timely, which is also necessary).
- Take publics' expectations into account where they think they can access reliable information
- Address multiple channels and check them regularly
- Be present where the public is: on social media.
- Be open to dialogue: one-way (or top-down) communication does not work; it is even rejected by the public.
- Digital forums of the media outlets
- Moderated discussions
- Journalism ethics could be a source of inspiration for social media guidelines
- A reflexion on social media platform business model is needed to counterbalance polarisation of information

Receiver

- People who are looking for information on Covid-19
- People interested in earthquakes in general or because they felt one
- People react differently to hazard and risk information and, thus, have different levels of preparedness
- People who are looking for information on Covid-19

- The receiver may lose trust in authorities and, consequently, believe less in preparedness messages.
- People may adopt dangerous behaviours if they believe false earthquake predictions (Albania)
- Optimism cognitive bias: People tend to believe that earthquake risk is higher for the entire country
- People who are concerned with the spread of the virus
- Chaos in reaching the victims in the aftermath of the attack → Request for travel ban
- Biases: Backfire effect, finite-pool-of-worry, herding, salience bias, confirmation bias

- Social media platform could certify official sources as reliable sources, hence help them to gain visibility
- Prevent misinformation with education and prebunking + inoculation
- Providing advice to recognize it
- Increasing science literacy and self-efficacy
- Communication should target the entire population but tailored to different societal groups: children, elderly, families, etc.
- Citizen
- Journalists
- Influencers
- Digital literacy: long-term educational process
- Reflexion on interface design is needed
- Help (re)build trust
- People want to know, understand, and talk
- We need to be able to establish a dialogue with the public; not sporadic but continuous
- Stimulation of critical thinking
- Raising trust in risk mitigation measures

Effect

- People do not believe anymore in authoritative messages
- Misbehaviour
- Distrust towards the authorities
- Confusion between different information services (e.g., EEW)
- Low trust in authorities and low risk perception levels can decrease the effectiveness of communication networks.
- Increasing social anxiety
- Leading to indifferences to potential real alarms
- Depending of the rumours, they can be short- or long-term

- Burning cell towers
- Fear of technology
- On a long term, it could lead to general distrust toward institutions.
- Finite-pool-of-worry: People weigh the relevance of different hazards differently based on, for example, past earthquake experiences or risk perception.
- Alarm fatigue & cry-wolf effect: False alarms make people less receptive to real alarms and lead to response delays.
- Erosion of trust in capacities of the government to manage risk
- Long-term effect is the tendency to lose trust in authorities over time, which makes people more

Feedback

- Need to understand the dynamics to counteract
- Work with partners
- Create a network of institutions
- If new tools/technologies emerge, it is important to embed them in the existing communication network
- User testing to make sure that the communications are co-designed with the end-users
- Contemporary information system challenging societies
- Need in development of AI tools for raising attention about misinformation

likely/vulnerable to belief in conspiracies related to other topics

- Negative impacts on rescue services on victims' families
- Spread of hate speech
- A direct and continuous dialogue with the population by volcanologists and civil protection would be necessary.
- Measures to raise trust in official sources of information and capacities of the government to manage the risk.
- Clarification about a rumour can be simple or complex (children in hotels vs. refugee bomber)

- At EMSC, we see that automatic prebunking is important but not sufficient; manual tweets are more often shared
- Feedback is double-sided: Both authorities and the public should gain feedback from each other.
- Local authority plays an important role as a mediator
- Empowering professionals to correct it
- Contemporary communication and information is increasingly individualized, making it more difficult to understand and manage
- The difference is given by the credibility in the speaker, and the credibility over time.
- Feedback and moderated discussions by public authorities but also communication professionals
- Public authorities and scientific institutions should contribute to the Twitter debate (e.g. answer, comment to informal individual users)

5.2 Recommendations to prevent and fight misinformation

This summary of the key insights and the derived recommendations per communication chain step are intended to support institutions, scientists and practitioners who are communicating (hazard and risk) information to the public and who are trying to prevent and fight the spread of misinformation. The recommendations are valid for natural and anthropogenic hazards as well as the multi-hazard context.

Regarding the **source**, there are different players who provide information and exchange on the currently existing communication platforms/networks. Regarding the actors who spread misinformation, one can differentiate between the following actors[23]: i) people who just believe in misinformation; ii) self-proclaimed experts; iii) involuntary sources because a misinformation spreader is linking them in his/her messages (can also happen to official sources); iv) bots; and v) people experiencing an emergency and trying to make sense to what is happening. Thus, misinformation is shared intentionally as well as unintentionally. The actors fighting misinformation are mainly: i) official sources (e.g., civil protection, authorities); ii) press agencies; iii) (scientific) experts; iv) citizens willing to correct a false rumour; and v) influencers. The importance is that these sources are trusted by the society, because only then is their provided information perceived as reliable. From these insights we derived three main recommendations:

☞ Official actors (e.g., authorities, civil protection) should provide information on a regular basis to build trust and credibility, which will ensure that people perceive them as a reliable source during an emergency.

☞ International or national authorities should provide a list of reliable sources of information to help the public access accurate information.

☞ Official actors (e.g., authorities, scientific institutions) should provide reliable and updated information about an emergency, which allows citizens to link this information in their messages on the social media platforms and other communication channels (e.g., WhatsApp).

The fact that misinformation messages are more attractive, leads to the fact that they are better remembered.

Regarding the **message**, official messages by authorities or the press have an objective and neutral tone and contain the context of the provided information. In comparison, misinformation messages are often written in a dramatic way, address drastic consequences, and use attractive design elements (e.g., videos, pictures), which increases people's attraction. The fact that misinformation messages are more attractive, leads to the challenge that they are better remembered by the people and thus overemphasized (salience bias). Additionally, misinformation can be simple (one clear statement) or complex (addressing other public debates), which influences the efforts needed to correct them. From these insights we derived three main recommendations:

---

[23] This is not an exhaustive list, but a listing of the key players identified in our case studies.

👉 The source of the information should be clearly and well visible stated in the message, allowing people to access more detailed information on the official websites.

👉 When educating people about certain misinformation in quiet times, the corresponding messages should be designed in an attractive way using for example pictures, videos etc.

👉 Fragmented information on social media, such as Twitter, should be counterbalanced through external links to richer sources, embedded videos, and tweet threads. This allows one to also give the context of a statement.

Regarding the **channel**, misinformation is often present on multiple social media platforms and other communication networks, which can lead to confirmation biases. Thus, people check the accuracy of the information on one channel with available information on other channels and start believing it when it is also present there. In general, misinformation is more often spread on social media than in the general press. From these insights we derived three main recommendations:

👉 Official actors should be present on the social media platforms to continuously provide accurate information and immediately counteract the spread of misinformation (especially during an emergency).

👉 Official actors should offer two-way-communication tools, allowing people to ask their questions directly to experts (e.g., moderated discussions).

👉 Social media platforms should certify official sources as reliable, hence help them to gain visibility. Further, journalism ethics could be a source of inspiration for social media guidelines.

Regarding the **receivers**, one can differentiate between people who actively look for information related to an event (e.g., earthquake, pandemic), people who just came across the misinformation, and people who follow certain accounts and thus received the misinformation message. To ensure that these different groups understand that a certain information is not correct, different strategies are needed. For example, to convince people, who are part of conspiracy networks and consume related information more frequently, is more challenging than providing clarifying information to people who just came across a certain information. In the case of emergencies, people's emotions and fears must be considered since these feelings influence their thinking and behaviour in these exceptional moments. Thus, although uncertainties are always there, reassuring messages are key for people under stress.

👉 Official actors should invest in science literacy and increase people's self-efficacy and capacities to critically reflect information they read (e.g., prebunking). Consequently, people are enabled to identify misinformation by themselves.

👉 When designing communications to counter misinformation, affected people's characteristics and environments (e.g. peers, knowledge) must be taken into account.

☞ Resources and capacities to interact with the public/receivers of an information should not be underestimated and must be guaranteed when aiming to communicate with the public.

Regarding the **effect**, it depends on the circumstances of an event as well as on the receivers of the information. In our case studies, we identified that misinformation can lead to i) inappropriate behaviours (e.g., burning cell towers, unnecessary evacuations); ii) neglection of real alarms due to alarm-fatigue triggered by too many false messages; iii) erosion of trust in authorities and their capacities to manage risk; iv) negative impacts on rescue services; and v) spread of hate speech (against innocent people). From these insights we derived three recommendations:

☞ Although challenging, authorities who lost their credibility should try, with support from partner institutions, to rebuild credibility and trust.

☞ When developing strategies to fight misinformation, the cultural context and ideological debates and discussions in the targeted societal groups must be considered. Only then can rooted and anchored beliefs and biases be addressed.

☞ Self-responsibility and critical thinking should be strengthened since official actors do not have the resources to monitor and take part in any public debate with misinformation present.

Regarding the **feedback**, the case studies revealed that a well-established network with all relevant actors is the basis to effectively prevent and fight misinformation. Understanding the dynamics of these networks allows one to implement the strategies at the nodes where they have the highest positive impact. However, contemporary information systems challenge societies (e.g., information overload, short-lived information). Moreover, emerging technologies, such as AI tools, can either be a supporter in the fight against misinformation or an accelerator of its spread. From these insights we derived three main recommendations:

☞ All relevant actors should actively create and, afterwards, maintain a well-structured communication network, allowing them to quickly, jointly, and effectively respond to misinformation.

☞ The potential of AI can and should be used (e.g., automatic prebunking), but it is not sufficient. Manual messages cannot be replaced entirely, because each event is different and needs precise and tailored wording.

☞ Emerging technologies and tools should be reflected ethically and from a societal perspective before being implemented and used in communication.

6 CONCLUSIONS

With our efforts in task 7.2, we assessed the dynamics, constellations and networks of social media (especially Twitter) and the general press in order to understand the evolution of misinformation over time for multiple hazards. Further, we analysed the effect of misinformation on these platforms on receivers' attitudes, perceptions, and behaviours. Besides the identification of cognitive and behavioural biases, such as the confirmation and salience bias, we provide a better understanding on the short- and long-term effect of misinformation (e.g., erosion of trust in official sources). We further identified the importance of authoritative narratives and provision of accurate information, which gives context and allows people to support their statements against misinformation with links to reliable sources. Overall, we conclude that official actors should create and maintain a rooted (communication) network with the public in quiet times to establish credibility and trust. This allows them to effectively communicate in case of an emergency and fight misinformation also under circumstances of uncertainties, stress, and anxiety.

ANNEX A

Screenshots of the Mural board of the interactive synthesis workshop hold on June 1st, 2023, with all case study representatives. The agenda of the workshop was as follows:

10.00 – 10.10	Introduction of the case studies	Irina
10.10 – 10.15	Procedure and aim of the workshop	Irina
10.15 – 10.45	Topic I – Dynamics, constellations, and networks	All
10.45 – 11.15	Topic II – Short-term and long-term effects on publics' risk perception, attitudes, and cognitive biases	All
11.15 – 11.45	Topic III – Risk communication & management and the role of authorities/official narratives	All
11.45 – 11.55	Final round	Irina

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

REFERENCES

Abrams, Z. (2021). Controlling the spread of misinformation. *Monitor on Psychology*, *52*(2), Article 2. https://www.apa.org/monitor/2021/03/controlling-misinformation

Acerbi, A., Altay, S., & Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-87

Ahmed, W., Vidal-Alaball, J., Downing, J., Seguí, F.L (2020).: COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of medical internet research*., 22(5), e19458.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Alexander, D. E. (2019). L'Aquila, central Italy, and the "disaster cycle", 2009-2017. *Disaster Prevention and Management: An International Journal*, *28*(4), 419–433. https://doi.org/10.1108/DPM-01-2018-0022

Allen, R. M., Gasparini, P., Kamigaichi, O., & Bose, M. (2009). The status of earthquake early warning around the world: An introductory overview. *Seismological Research Letters*, *80*(5), Article 5. https://doi.org/10.1785/gssrl.80.5.682

Amaratunga, D., Haigh, R., & Hettige, S. (2019). *Accountability in the context of disaster risk governance*. UNDRR.

Arora, S. (2021). Post-disaster communities on social media: Citizen participation in crisis communication after the Nepal earthquake, 2015. *Journal of Applied Communication Research*, 1–18. https://doi.org/10.1080/00909882.2021.1964572

Barnes K., 2011: *Europe's ticking time bomb*, in *Nature*, vol. 473, May 12, 140-141.

Baron, J. (2012). Beliefs about thinking. In *Informal Reasoning and Education* (S. 187–204). Routledge.

Bautista, J. R., Zhang, Y., & Gwizdka, J. (2021). Healthcare professionals' acts of correcting health misinformation on social media. *International Journal of Medical Informatics*, *148*, 104375. https://doi.org/10.1016/j.ijmedinf.2021.104375

Bean, H., Sutton, J., Liu, B. F., Madden, S., Wood, M. M., & Mileti, D. S. (2015). The study of mobile public warning messages: A research review and agenda. *Review of Communication*, *15*(1), 60–80. https://doi.org/10.1080/15358593.2015.1014402

Becker, J. S., Potter, S. H., Vinnell, L. J., Nakayachi, K., McBride, S. K., & Johnston, D. M. (2020). Earthquake early warning in Aotearoa New Zealand: A survey of public perspectives to guide warning system development. *Humanities and Social Sciences Communications*, *7*(1), 138. https://doi.org/10.1057/s41599-020-00613-9

Bérubé, M., Tang, T. U., Fortin, F., Ozalp, S., Williams, M. L., & Burnap, P. (2020). Social media forensics applied to assessment of post–critical incident social reaction: The case of the 2017 Manchester Arena terrorist attack. *Forensic science international*, 313, 110364. https://doi.org/10.1016/j.forsciint.2020.110364

Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, *33*(9), 1131–1140. https://doi.org/10.1080/10410236.2017.1331312

Bossu, R., Roussel, F., Fallou, L., Landès, M., Steed, R., Mazet-Roux, G., Dupont, A., Frobert, L., & Petersen, L. (2018). LastQuake: From rapid information to global seismic risk reduction. *International Journal of Disaster Risk Reduction*, *28*, 32–42. https://doi.org/10.1016/j.ijdrr.2018.02.024

Botzen, W., Duijndam, S., & van Beukering, P. (2021). Lessons for climate policy from behavioral biases towards COVID-19 and climate change risks. *World Development*, *137*, 105214. https://doi.org/10.1016/j.worlddev.2020.105214

Broomell, S. B. (2020). Global–local incompatibility: The misperception of reliability in judgment regarding global variables. *Cognitive Science*, *44*(4), e12831. https://doi.org/10.1111/cogs.12831

Brumfiel, G. (2022, April 24). Their mom died of COVID. They say conspiracy theories are what really killed her. *NPR*. https://www.npr.org/sections/health-shots/2022/04/24/1089786147/covid-conspiracy-theories

Bruns, H., Dessart, F., & Pantazi, M. (2022). *COVID-19 misinformation: Preparing for future crises* (Publications Office of the European Union). https://doi.org/10.2760/41905

Capstick, S., Whitmarsh, L., Poortinga, W., Pidgeon, N., & Upham, P. (2015). International trends in public perceptions of climate change over the past quarter century. *WIREs Climate Change*, *6*(1), 35–61. https://doi.org/10.1002/wcc.321

Castelli P., 1632: *Incendio del monte Vesuvio*, Giacomo Mascardi, Rome.

Chen, X., Hay, J. L., Waters, E. A., Kiviniemi, M. T., Biddle, C., Schofield, E., Li, Y., Kaphingst, K., & Orom, H. (2018). Health Literacy and Use and Trust in Health Information. *Journal of Health Communication*, *23*(8), 724–734. https://doi.org/10.1080/10810730.2018.1511658

Cichocka, A., Marchlewska, M., & Biddlestone, M. (2022). Why do narcissists find conspiracy theories so appealing? *Current Opinion in Psychology*, 101386. https://doi.org/10.1016/j.copsyc.2022.101386

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. BMC Medical Research Methodology, 11(2), Article 2.

Crowley, H., Dabbeek, J., Despotaki, V., Rodrigues, D., Martins, Silva, V., Romão, X., Pereira, N., Weatherill, G., & Danciu, L. (2021). European Seismic Risk Model (ESRM20) [EFEHR Technical Report 002 V1.0.0]. Eucentre. https://doi.org/10.7414/EUC-EFEHR-TR002-ESRM20

Dallo, I. (2022). *Understanding the communication of event-related earthquake information in a multi-hazard context to improve society's resilience* [Doctoral dissertation]. ETH Zurich.

Dallo, I., Corradini, M., Fallou, L., & Marti, M. (2022). *How to fight misinformation about earthquakes? - A Communication Guide*. https://doi.org/10.3929/ethz-b-000530319

Dallo, I., & Marti, M. (2021). Why should I use a multi-hazard app? Assessing the public's information needs and app feature preferences in a participatory process. *International Journal of Disaster Risk Reduction*, *57*, 102197. https://doi.org/10.1016/j.ijdrr.2021.102197

Dallo, I., Marti, M., Clinton, J., Böse, M., Massin, F., & Zaugg, S. (2022). Earthquake early warning in countries where damaging earthquakes only occur every 50 to 150 years – The societal perspective. International Journal of Disaster Risk Reduction, 83, 103441. https://doi.org/10.1016/j.ijdrr.2022.103441

Dallo, I., Stauffacher, M., & Marti, M. (2020). What defines the success of maps and additional information on a multi-hazard platform? *International Journal of Disaster Risk Reduction*, *49*, 101761. https://doi.org/10.1016/j.ijdrr.2020.101761

Danciu, L., Nandan, S., Reyes, C., Basili, R., Weatherill, G., Beauval, C., Rovida, A., Vilanova, S., Sesetyan, K., Bard, P.-Y., Cotton, F., Wiemer, S., & Giardini, D. (2021). The 2020 update of the European Seismic Hazard Model: Model Overview. EFEHR Technical Report 001, v1.0.0. https://doi.org/10.12686/A15

de Martino E., 2008: *Il mondo magico. Prolegomeni a una storia del magismo* [1948], Bollati Boringhieri, Turin.

Elmousalami, H. H., Darwish, A., & Hassanien, A. E. (2021). The truth about 5G and COVID-19: basics, analysis, and opportunities. *Digital transformation and emerging technologies for fighting COVID-19 pandemic: innovative approaches*, 249-259.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Fallou, L., Bossu, R., Landès, M., Roch, J., Roussel, F., Steed, R., & Julien-Laferrière, S. (2020). Citizen seismology without seismologists? Lessons learned from Mayotte leading to improved collaboration. *Frontiers in Communication*, *5*, 49. https://doi.org/10.3389/fcomm.2020.00049

Fallou, L., Marti, M., Dallo, I., & Corradini, M. (2022). How to fight earthquake misinformation: A communication guide. *Seismological Research Letters*. https://doi.org/10.1785/0220220086

Fekete, A., & Sandholz, S. (2021). Here comes the flood, but not failure? Lessons to learn after the heavy rain and pluvial floods in Germany 2021. *Water*, *13*(21), 3016. https://doi.org/10.3390/w13213016

Flores-Saviaga, C., & Savage, S. (2021). Fighting Disaster Misinformation in Latin America: The #19S Mexican Earthquake Case Study. *Personal and Ubiquitous Computing*, *25*(2), 353–373. https://doi.org/10.1007/s00779-020-01411-5

Fong, A., Roozenbeek, J., Goldwert, D., Rathje, S., & van der Linden, S. (2021). The language of conspiracy: A psychological analysis of speech used by conspiracy theorists and their followers on Twitter—Amos Fong, Jon Roozenbeek, Danielle Goldwert, Steven Rathje, Sander van der Linden, 2021. *Group Processes & Intergroup Relations*, *24*(4), 606–623. https://doi.org/10.1177/1368430220987596

Fraser, T., Morikawa, L., & Aldrich, D. P. (2021). Rumor has it: The role of social ties and misinformation in evacuation to nearby shelters after disaster. *Climate Risk Management*, *33*, 100320. https://doi.org/10.1016/j.crm.2021.100320

Garcia, C., & Fearnley, C. J. (2012). Evaluating critical links in early warning systems for natural hazards. *Environmental Hazards*, *11*(2), 123–137. https://doi.org/10.1080/17477891.2011.609877

Gelfert, A. (2018). Fake News: A Definition. *Informal Logic*, *38*(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645. Scopus. https://doi.org/10.1162/003355397555217

Griffin, R. J., Dunwoody, S., & Neuwirth, K. (1999). Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. *Environmental Research*, *80*(2), S230–S245. https://doi.org/10.1006/enrs.1998.3940

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Gugg G., 2018: *«Con lingue di foco ei par che gridi». Il Vesuvio, fucina di natura e immaginazione*, in B. Terracciano (ed by), *Geoaffetti. Narrare la nostra terra*, CMEA (Centro Meridionale di Educazione Ambientale), Sorrento: 77-93.

Gugg G., 2019: *Anthropology of the Vesuvius Emergency Plan: history, perspectives and limits of a dispositive for volcanic risk government*, in L. Antronico, F. Marincioni (eds by), *Natural hazards and Disaster Risk Reduction policies*, book of "Geographies of Anthropocene", Il Sileno, Cosenza.

Gugg G., 2021: *Guarire un vulcano, guarire gli umani. Elaborazioni del rischio ecologico e sanitario alle pendici del Vesuvio*, in "Antropologia Medica", vol. 51.

Gugg G., 2022: *Ordinary Life in the Shadow of Vesuvius: Surviving the Announced Catastrophe*, in B. Świtek, A. Abramson, H. Swee (eds by), *Extraordinary Risks, Ordinary Lives. Logics of Precariousness in Everyday Contexts*, Palgrave Macmillan, London.

Gugg G., 2023: *Afar from Vesuvius, but still at risk. The unstoppable urbanization of the Naples volcano's yellow zone*, in G. Forino (ed by), *Disasters and Changes into Society and Politics. Contemporary Perspectives from Italy*, Bristol University Press (Forthcoming).

Hailwood, M. (2020). *Release of hazardous vapours at LG Polymers chemical plant in Visakhapatnam, Andhra Pradesh, India*. 5.

Innes M. (2019) Techniques of disinformation: constructing and communicating soft facts after terrorism. Br. J. Sociology, 71

Jadhav, B., Gulecha, V., & Zalte, A. (2020). Synthesis of Triazolothiadiazole derivatives. *EPRA International Journal of Research & Development (IJRD)*, *5*(8). https://eprajournals.com/jpanel/images/pseditor/IJRD%20AUGUST%202020%20FULL%20JOURNAL.pdf#page=8

Jahanbakhsh, F., Zhang, A. X., Berinsky, A. J., Pennycook, G., Rand, D. G., & Karger, D. R. (2021). Exploring Lightweight Interventions at Posting Time to Reduce the Sharing of Misinformation on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–42. https://doi.org/10.1145/3449092

Jones, N. M., & Silver, R. C. (2020). This is not a drill: Anxiety on Twitter following the 2018 Hawaii false missile alert. *The American Psychologist*, *75*(5), 683–693. https://doi.org/10.1037/amp0000495

Jones, N. M., Thompson, R. R., Dunkel Schetter, C., & Silver, R. C. (2017). Distress and rumor exposure on social media during a campus lockdown. *Proceedings of the National Academy of Sciences*, *114*(44), 11663–11668. https://doi.org/10.1073/pnas.1708518114

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Keeley, B. L. (1999). Of conspiracy theories. *The Journal of Philosophy*, *96*(3). https://doi.org/10.2307/2564659

Khanum N., Khan F.M. (2020) Islamophobia and digital spaces: case of Twitter response on new stories of Manchester attack 5/22. International Journal of Science and Research, 76(4).

Kirschenbaum, A. A., Rapaport, C., & Canetti, D. (2017). The impact of information sources on earthquake preparedness. International Journal of Disaster Risk Reduction, 21, 99-109.

Klimiuk, K., Czoska, A., Biernacka, K., & Balwicki, Ł. (2021). Vaccine misinformation on social media – topic-based content and sentiment analysis of Polish vaccine-deniers' comments on Facebook. *Human Vaccines & Immunotherapeutics*, *17*(7), 2026–2035. https://doi.org/10.1080/21645515.2020.1850072

Kolokythas, A. (2021). What do the aftermath of the 2010 Haiti earthquake, Hurricane Sandy, the Boston Marathon bombing, the 2013 Ebola outbreak, and the COVID-19 pandemic have in common? *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *132*(4), 371–372. https://doi.org/10.1016/j.oooo.2021.06.016

Komendantova, N., Ekenberg, L., Svahn, M., Larsson, A., Shah, S. I. H., Glinos, M., Koulolias, V., & Danielson, M. (2021). A value-driven approach to addressing misinformation in social media. *Humanities and Social Sciences Communications*, *8*(1), 33. https://doi.org/10.1057/s41599-020-00702-9

Kotthaus, C., Ludwig, T., & Pipek, V. (2016). Persuasive system design analysis of mobile warning apps for citizens. *ECSW@ PERSUASIVE*, 12.

Kozyreva, A., Herzog, S., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2022). *Free speech vs. harmful misinformation: Moral dilemmas in online content moderation*. PsyArXiv. https://doi.org/10.31234/osf.io/2pc3a

Kreuzmair, C., Siegrist, M., & Keller, C. (2017). Does iconicity in pictographs matter? The influence of iconicity and numeracy on information processing, decision making, and liking in an eye-tracking study: Does iconicity in pictographs matter? *Risk Analysis*, *37*(3), 546–556. https://doi.org/10.1111/risa.12623

Kricorian, K., Civen, R., & Equils, O. (2021). COVID-19 vaccine hesitancy: Misinformation and perceptions of vaccine safety. *Human Vaccines & Immunotherapeutics*, *0*(0), 1–8. https://doi.org/10.1080/21645515.2021.1950504

Kunreuther, H., & Slovic, P. (2021, Januar 26). *Learning from the COVID-19 pandemic to address climate change*. MBR Journal. https://mbrjournal.com/2021/01/26/learning-from-the-covid-19-pandemic-to-address-climate-change/

Kunreuther, H., & Useem, M. (2009). *Learning from catastrophes: Strategies for reaction and response*. Pearson Prentice Hall.

Kurata, Y. B., Prasetyo, Y. T., Ong, A. K. S., Nadlifatin, R., & Chuenyindee, T. (2022). Factors affecting perceived effectiveness of Typhoon Vamco (Ulysses) flood disaster response among Filipinos in Luzon, Philippines: An integration of protection motivation theory and extended theory of planned behavior. *International Journal of Disaster Risk Reduction*, *67*, 102670. https://doi.org/10.1016/j.ijdrr.2021.102670

Kwanda, F. A., & Lin, T. T. C. (2020). Fake news practices in Indonesian newsrooms during and after the Palu earthquake: A hierarchy-of-influences approach. *Information, Communication & Society*, *23*(6), 849–866. https://doi.org/10.1080/1369118X.2020.1759669

Lacassin, R., Devès, M., Hicks, S. P., Ampuero, J.-P., Bossu, R., Bruhat, L., Daryono, Wibisono, D. F., Fallou, L., Fielding, E. J., Gabriel, A.-A., Gurney, J., Krippner, J., Lomax, A., Sudibyo, Muh. M., Pamumpuni, A., Patton, J. R., Robinson, H., Tingay, M., & Valkaniotis, S. (2019). *Rapid collaborative knowledge building via Twitter after significant geohazard events* [Preprint]. Geoscience engagement/Co-creation and co-production. https://doi.org/10.5194/gc-2019-23

Impact of misinformation on social media on risk
perception in a multi-risk environment
D7.2

Lee, J. (2022). When web add-on correction comes with fear-arousing misinformation in public health crisis: Focusing on the role of risk perception in belief in misinformation. *Journal of Applied Communication Research*, *50*(1), 70–90. https://doi.org/10.1080/00909882.2021.1964574

Lee, J., Choi, J., & Britt, R. K. (2021). Social media as risk-attenuation and misinformation-amplification station: How social media interaction affects misperceptions about COVID-19. *Health Communication*, 1–11. https://doi.org/10.1080/10410236.2021.1996920

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lindell, M. K., & Perry, R. W. (2012). The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis*, *32*(4), 616–632. https://doi.org/10.1111/j.1539-6924.2011.01647.x

Lukasiewicz, A., & Baldwin, C. (Hrsg.). (2020). *Natural hazards and disaster justice: Challenges for Australia and its neighbours*. Springer Singapore. https://doi.org/10.1007/978-981-15-0466-2

Maduz, L., Roth, F., Prior, T., & Wolf, A. (2018). *Individuelle Katastrophenvorsorge: Gefährdungswahrnehmung, Kenntnisse und Informationsbedürfnisse der Schweizer Bevölkerung* (Risk and Resilience Report). Center for Security Studies (CSS). https://www.researchgate.net/publication/326200547_Individuelle_Katastrophenvorsorge_Gefahrdungswahrnehmung_Kenntnisse_und_Informationsbedurfnisse_der_Schweizer_Bevolkerung

Manfredi G., 2015: *Infodemia. I meccanismi complessi della comunicazione nelle emergenze*, Guaraldi, Rimini.

Marks, M. J., & Fraley, R. C. (2006). Confirmation Bias and the Sexual Double Standard. *Sex Roles*, *54*(1), 19–26. https://doi.org/10.1007/s11199-006-8866-9

Marti, M., Stauffacher, M., & Wiemer, S. (2020). Anecdotal evidence is an insufficient basis for designing earthquake preparedness campaigns. Seismological Research Letters, 91(4), Article 4. https://doi.org/10.1785/0220200010

Meyer, R., & Kunreuther, H. (2017). *The ostrich paradox: Why we underprepare for disasters*. University of Pennsylvania Press.

Mileti, D. S., & Sorensen, J. H. (1990). *Communication of emergency public warnings: A social science perspective and state-of-the-art assessment*. Oak Ridge National Lab., TN (USA).

Muhammed T, S., & Mathew, S. K. (2022). The disaster of misinformation: A review of research in social media. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-022-00311-6

Oh, O., Kwon, K. H., & Rao, H. R. (2010). *An exploration of social media in extreme events: Rumor theory and twitter during Haiti earthquake 2010*. 15.

Ozturk, P., Li, H., & Sakamoto, Y. (2015). Combating Rumor Spread on Social Media: The Effectiveness of Refutation and Warning. *2015 48th Hawaii International Conference on System Sciences*, 2406–2414. https://doi.org/10.1109/HICSS.2015.288

Peary, B. D. M., Shaw, R., & Takeuchi, Y. (2012). Utilization of Social Media in the East Japan Earthquake and Tsunami and its Effectiveness. *Journal of Natural Disaster Science*, *34*(1), 3–18. https://doi.org/10.2328/jnds.34.3

Peng, Z. (2020). Earthquakes and Coronavirus: How to Survive an Infodemic. *Seismological Research Letters*, *91*(5), 2441–2443. https://doi.org/10.1785/0220200125

Pierpoint, G. (2018, August 20). Kerala floods: Fake news „creating unnecessary panic". *BBC News*. https://www.bbc.com/news/world-asia-india-45245999

Porter, K. A. (2021). Should we build better? The case for resilient earthquake design in the United States. Earthquake Spectra, 37(1), Article 1. https://doi.org/10.1177/8755293020944186

Potter, S. H., Harrison, S., & Kreft, P. (2021). The benefits and challenges of implementing impact-based severe weather warning systems: Perspectives of weather, flood, and emergency management personnel. *Weather, Climate, and Society*. https://doi.org/10.1175/WCAS-D-20-0110.1

Prerna, S., & Abhigya. (2021, Dezember 10). *The making of Vizag gas leak disaster: Procedural lapses or regulatory design?* https://blog.castac.org/2020/08/the-making-of-vizag-gas-leak-disaster-procedural-lapses-or-regulatory-design/

Prochaska, J. O., Redding, C. A., Harlow, L. L., Rossi, J. S., & Velicer, W. F. (1994). The transtheoretical model of change and HIV prevention: A review. *Health Education Quarterly*, *21*(4), 471–486. https://doi.org/10.1177/109019819402100410

Radianti, J., Hiltz, S. R., & Labaka, L. (2016). An Overview of Public Concerns During the Recovery Period after a Major Earthquake: Nepal Twitter Analysis. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 136–145. https://doi.org/10.1109/HICSS.2016.25

Reed A., Ingram H.J. (2020) Towards a framework for post-terrorist incident communications strategies. *Global Research Network on Terrorism and Technology Paper,* 12, 4-5.

Reuter, C., & Spielhofer, T. (2017). Towards social resilience: A quantitative and qualitative survey on citizens' perception of social media in emergencies in Europe. *Technological Forecasting and Social Change*, *121*, 168–180. https://doi.org/10.1016/j.techfore.2016.07.038

Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., & Sultan, M. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making*, *17*(3), 27.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Santoianni F., 2007: *Disaster Management. Protezione Civile*, Accursio Edizioni, Florence.

Savoia, E., Harriman, N. W., Piltch-Loeb, R., Bonetti, M., Toffolutti, V., & Testa, M. A. (2022). Exploring the association between misinformation endorsement, opinions on the government response, risk perception, and COVID-19 vaccine hesitancy in the US, Canada, and Italy. *Vaccines*, *10*(5), 671. https://doi.org/10.3390/vaccines10050671

Scheufele, D. A. (2013). Communicating science in social settings. *Proceedings of the National Academy of Sciences*, *110*(supplement_3), 14040–14047. https://doi.org/10.1073/pnas.1213275110

Scheufele, D. A., Krause, N. M., & Freiling, I. (2021). Misinformed about the "infodemic?" Science's ongoing struggle with misinformation. *Journal of Applied Research in Memory and Cognition*, *10*(4), 522–526. https://doi.org/10.1016/j.jarmac.2021.10.009

Schoch, K. (2020). Case study research. In Research design and methods: An applied guide for the scholar-practitioner (pp. 245–258).

Sechi, G. J., Lopane, F. D., & Hendriks, E. (2022). Mapping seismic risk awareness among construction stakeholders: The case of Iringa (Tanzania). International Journal of Disaster Risk Reduction, 82, 103299. https://doi.org/10.1016/j.ijdrr.2022.103299

Simmel G., 1979: *La metropoli e la vita spirituale* [1903], in T. Maldonado (ed. by), *Tecnica e cultura. Il dibattito tedesco tra Bismarck e Weimar*, Feltrinelli, Milan.

Slovic, P., Fischhoff, B., Lichtenstein, S., Corrigan, B., & Combs, B. (1977). Preference for insuring against probable small losses: Insurance implications. *The Journal of Risk and Insurance*, *44*(2), 237–258. https://doi.org/10.2307/252136

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, *22*(4), 259–264. https://doi.org/10.1177/0963721413480174

Staton A.R. (2020) Evolving crisis communication in social media: analysis of tweets and news stories of the Manchester Arena bombing. U. Southern Miss. Thesis, May.

Sullivan-Wiley, K. A., & Short Gianotti, A. G. (2017). Risk perception in a multi-hazard environment. *World Development*, *97*, 138–152. https://doi.org/10.1016/j.worlddev.2017.04.002

Tandoc E., 2014: *Journalism is twerking? How web analytics is changing the process of gatekeeping*, in "New Media & Society", vol. 16 (4): 1-17.

Tarunesh, I., Aditya, S. & Choudhury, M. (2021). Trusting roberta over bert: Insights from checklisting the natural language inference task. arXiv preprint arXiv:2107.07229

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

UNESCO. (2022). *Addressing conspiracy theories: What teachers need to know*. the United Nations Educational, Scientific and Cultural Organization,. https://unesdoc.unesco.org/ark:/48223/pf0000381958

United Nations Office for Disaster Risk Reduction. (2022). *Global Assessment Report on Disaster Risk Reduction 2022: Our World at Risk: Transforming Governance for a Resilient Future*.

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, *28*(3), 460–467. https://doi.org/10.1038/s41591-022-01713-6

van Prooijen, J.-W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, *48*(7), 897–908. https://doi.org/10.1002/ejsp.2530

Vicari, R., & Komendatova, N. (2023). Systematic meta-analysis of research on AI tools to deal with misinformation on social media during natural and anthropogenic hazards and disasters. *Humanities and Social Sciences Communications*, *10*(1), Article 1. https://doi.org/10.1057/s41599-023-01838-0

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wallman S., 2001: *Global threats, local options, personal risk. Dimensions of migrant sex work in Europe*, "Health, Risk & Society", n. 3, vol. 1.

Weber, E. U. (2010). What shapes perceptions of climate change? *WIREs Climate Change*, *1*(3), 332–342. https://doi.org/10.1002/wcc.41

Weyrich, P., Scolobig, A., & Patt, A. (2019). Dealing with inconsistent weather warnings: Effects on warning quality and intended actions. *Meteorological Applications*, *26*(4), 569–583. https://doi.org/10.1002/met.1785

Wood, M. M., Mileti, D. S., Kano, M., Kelley, M. M., Regan, R., & Bourque, L. B. (2012). Communicating actionable risk for terrorism and other hazards. *Risk Analysis*, *32*(4), 601–615. https://doi.org/10.1111/j.1539-6924.2011.01645.x

World Economic Forum, 2013: *Digital Wildfires in a Hyperconnected World*, http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/

Xia, Y., Lukito, J., Zhang, Y., Wells, C., Kim, S. J., & Tong, C. (2019). Disinformation, performed: Self-presentation of a Russian IRA account on Twitter. *Information, Communication & Society*, *22*(11), 1646–1664. https://doi.org/10.1080/1369118X.2019.1621921

Yaar E, Brandes-Yaacov O, Elkalay Y, Gelman A, Bramnis (2015). A. Assessing perceptions in the Israeli population regarding a strong earthquake in Israel, preparing for it and the ability to cope with its consequences. Jerusalem: Henrietta Szold Institute (Hebrew).

Zarocostas, J.: How to fight an infodemic. The lancet 395, 676 (2020)

Zhao X., Zhan M.M. (2019) Appealing to the heart: How social communication characteristics affected users' liking behaviour during the Manchester terrorist attack. *International Journal of Communication, 13*, 22.

Zhou, C., Xiu, H., Wang, Y., & Yu, X. (2021). Characterizing the dissemination of misinformation on social media in health emergencies: An empirical study based on COVID-19. *Information Processing & Management*, *58*(4), 102554. https://doi.org/10.1016/j.ipm.2021.102554

Zhou, L., Wu, X., Xu, Z., & Fujita, H. (2018). Emergency decision making for natural disasters: An overview. *International Journal of Disaster Risk Reduction*, *27*, 567–576. https://doi.org/10.1016/j.ijdrr.2017.09.037

Zonabend, F. (1989): *La presqu'île au nucléaire*, Editions Odile Jacob, Paris.

CARE
sCience and human factOr for Resilient sociEty

UNIVERSITÀ DEGLI STUDI DI SALERNO

ISSNOVA
Institute for Sustainable Society and Innovation

ETH zürich

IIASA

University of HUDDERSFIELD
Inspiring global professionals

HANKEN

SAHER EUROPE

PSCEurope
Public Safety Communication Europe

Science Ethics
Institute for Science & Ethics

COLLEGE OF LAW & BUSINESS

MTOSafety

RESALLIENCE by sixense

CSEM EMSC

VIGILI DEL FUOCO CORPO NAZIONALE

Croce Rossa Italiana
Comitato di Vicenza

UCSA
Ufficio Comune per la Sostenibilità Ambientale

Hochschule für den öffentlichen Dienst in Bayern

ÖSTERREICHISCHES ROTES KREUZ

parco archeologico campi flegrei