Journal of Quantitative Description: Digital Media 4(2024), 1–46 10.51685/jqd.2024.icwsm.8

"You are doomed!" Crisis-specific and Dynamic Use of Fear Speech in Protest and Extremist Radical Social Movements

SIMON GREIPL
JULIAN HOHNER
HEIDI SCHULZE
PATRICK SCHWABL
DIANA RIEGER
Ludwig-Maximilians-University of Munich, Germany

Social media messages can elicit emotional reactions and mobilize users. Strategic utilization of emotionally charged messages, particularly those inducing fear, potentially nurtures a climate of threat and hostility online. Coined fear speech (FS), such communication deliberately portrays certain entities as imminently harmful and drives the perception of a threat, especially when the topic is already crisis-laden. Despite the notion that FS and the resulting climate of threat can serve as a justification for radical attitudes and behavior toward outgroups, research on the prevalence, nature, and context of FS is still scarce. The current paper aims to close this gap and provides a definition of FS, its theoretical foundations, and a starting point for (automatically) detecting FS on social media. The paper presents the results of a manual as well as an automated content analysis of three broadly categorized actor types within a larger radical German Telegram messaging sphere (2.9 million posts). With a rather conservative classification approach, we analyzed the prevalence and distribution of FS for more than five years in relation to six crisis-specific topics. A substantial proportion between 21% and 34% within the observed communication of radical/extremist actors was classified as FS.

Author 1: simon.greipl@ifkw.lmu.de
Date submitted: 2024-04-26

Additionally, the relative amount of FS was found to increase with the overall posting frequency. This underscores FS's potential as an indicator for radicalization dynamics and crisis escalation.

Keywords: fear speech, radicalization, BERT, transformers, classifier, Telegram, far right, COVID-19

Emotions play a pivotal role in shaping our perceptions and interactions, particularly in online communication (Ellis & Tucker, 2022). The ability of language to elicit emotional responses has long been recognized, with Aristotle highlighting the strategic use of emotions as a persuasive tool in his seminal work, *Rhetoric* (Aristotle & Roberts, 2004). This principle finds resonance online, where emotional contagion can lead individuals to unconsciously share and experience the same feelings, thus amplifying the impact of emotional messages on user engagement and content generation (Ellis & Tucker, 2022; Kramer et al., 2014). Such dynamics underscore the potent role of emotional content in social media as a catalyst for user interaction (Goldenberg & Gross, 2020).

The emotion of fear, in particular, demonstrates a unique capacity to propagate digital contagion effects and incite viral panics, highlighting its significance in the digital realm (Lăzăroiu & Adams, 2020). Fear is elicited when people sense a threat or danger to themselves or their significant ingroup (Cohen-Chen et al., 2014) and plays a crucial role in the process of radicalization (Doosje et al., 2013). Extremist groups often exploit perceptions of fear (Marcks & Pawelz, 2022; Schulze et al., 2023), using them strategically in their communication to emphasize narratives like “the great replacement,” which conveys the fear of cultural extinction among the far right (Ziolkowski et al., 2022).

Accordingly, it is theorized that fear not only contributes to the online radicalization process (Greipl et al., 2022; Schulze et al., 2023; Rothut et al., 2022) but also underpins the rhetoric used by these groups. The concept of "dangerous speech," as described by

Benesch (2023), links rhetoric that induces fear to the escalation of violent intergroup conflicts and the potential for imminent violence (Buyse, 2014). Terrorist manifestos, for instance, are prime examples of such speech, illustrating motivations behind violent acts and underscoring fear's pivotal role in extremist ideologies and actions (Wright, 2019).

A particularly concerning manifestation of fear in digital communication is fear speech (FS), which seeks to instill existential fear toward minority groups or particular communities (Buyse, 2014). In early analyses of FS in Facebook communication, mostly minorities or perceived elites were reported as targets of FS and accused of being involved in conspiracies and rumors or were openly discriminated against (Gagliardone et al., 2016), thereby fueling a climate of threat and crisis perception on social media platforms. Fearful narratives create a community of fate (e.g., strengthening ingroup and hardening outgroup boundaries) while they increase the perceived pressure to act and, thereby, the chances of violence (Buyse, 2014; Ziolkowski et al., 2022). The frequency of fear-based discourse on social media may thus correlate with perceptions of a threat or crisis, prompt defensive reactions, and also exacerbate intergroup hostilities.

Recent research suggests that the prevalence of fear speech (FS) may even exceed that of more widely recognized forms of hate speech (Saha et al., 2023). This indicates a potentially more insidious threat due to its covert nature, which makes it challenging to identify and address. Furthermore, fear speech has a profound ability to foster hostility (Buyse, 2014; Saha et al., 2023).

Despite its significance, systematic research on the prevalence, nature, and context of FS remains scarce. This paper aims to fill this gap by providing an (operational) definition of FS and conducting a comprehensive analysis within a notably large segment of a potentially radical German Telegram sphere, comprising three actor types with varying political motivations (far-right, COVID-19 protest, and conspiracy). Utilizing both automated and manual content analysis of messages for more than five years, this study incorporates insights from psychology and political science on fear and anxiety, as well as

the effects of emotional messaging in social media. It aims to clarify the mechanisms by which fear speech (FS) operates and its implications for understanding the dynamics of radicalization and crisis escalation online.

From fear to hostility: The psychological and political dimension of fear and anxiety

Emotions are likewise inextricably linked to political life as they are to the human experience in general. Adhering to affective intelligence theory (AIT) (Marcus et al., 2000), three emotions are relevant in political behavior: enthusiasm, anger, and fear. Albertson and Gadarian (2015) trenchantly describe this, "…enthusiastic supporters return politicians to office, angry citizens march in the streets, a fearful public demands protection from the government" (p. 1). The latter two present a negative valance of emotions.

Fear, a primary emotion, emerges when an individual senses a threat or danger to themselves or a significant ingroup, eliciting physiological and psychological responses that enhance survival in dangerous situations (Öhman, 1993). Fear responses usually appear quickly and automatically and then affect subsequent thought processes, with the potential of occurring at the cost of rational thinking (Jarymowicz & Bar-Tal, 2006; LeDoux, 1995). Fear impacts the way we seek and process information (Marcus et al., 2000). It is believed to heighten individuals' sensitivity to threatening signals, amplify potential threat information, and lead to an overestimation of danger (Bar-Tal, 2013). Importantly, fear responses are associated with cognitive closure and biased information processing. For instance, threatening information is prioritized, selective recall of fear-related information is facilitated, and receptivity to novel ideas is obstructed (Clore et al., 1994; Isen, 1990; LeDoux, 1996). Potentially, very intense fear even induces cognitive inflexibility (Kruglanski, 2004) and can lead to aggressive reactions (Eibl-Eibesfeldt & Sütterlin, 1990; Lazarus, 1991). These impacts extend to the political realm, where threat and resulting fear responses can influence attitude formation and political actions, particularly within the context of intergroup engagement. This is mirrored in risk-averse

political preferences, less creative ideas with regard to conflict resolution, and more resistance when it comes to intergroup negotiation (Sabucedo et al., 2011). Fear seems to be a major driver of conservatism (Jost et al., 2003), e.g., by strengthening conservative viewpoints that maintain the familiar condition of conflict. Fear responses enhance authoritarianism (Canetti et al., 2009), prejudices against minority groups (Sari, 2007), and support for bolstering both domestic and international security policy (Huddy et al., 2007).

In its psychological interpretation, fear, as an instinctual response to imminent danger, paradoxically proves inadequate for a comprehensive grasp of FS. This inadequacy in political messaging stems from its limitations, as it typically does not evoke the intense, primal fear associated with immediate, "fight-or-flight" scenarios. Instead, political messages tend to induce anxiety—a persistent sense of apprehension and hypervigilance in situations where a threat might exist, as detailed by Stegmann et al. (2023). Political science often focuses on this concept of anxiety, as outlined by Albertson and Gadarian (2015), describing it as a prolonged reaction to uncertain or ambiguous threats. Anxiety can lead individuals to seek out or avoid information. This behavior serves either to resolve uncertainty or to ignore the source of anxiety (Gross, 2013). Generally, it is assumed in political science that anxiety increases both the desire for information and the active search for it (Albertson and Gadarian, 2015). In sum, fear and anxiety are both emotions characterized by negative valence that arise in reaction to threat, yet they have a partially different neurobiological basis and may result in different responses to different threats[1] (Helminen et al., 2022). Although fear and anxiety represent distinct emotional experiences for the recipient, we use fear as an umbrella term for the current purpose.[2] Given a psychological association in which anxiety can stem from fear, coupled with a preference

[1] Usually conceptualized in exclusive models, accumulating evidence points to additive if not interactive models rather than exclusive models of fear and anxiety indicating facilitated defensive behavior (Stegmann et al., 2023).

[2] Not clearly distinct but also not to be used interchangeably according to biological/neuroscience (Daniel-Watanabe & Fletcher, 2021), they are often used interchangeably in political science (Albertson and Gadarian, 2015; Scheller, 2019).

for parsimony, we opt for FS as the preferred terminology to analyze political messaging at the message or post level.

The complex interplay between relevant negative emotions arising from threats to political attitudes and behavior is under ongoing empirical investigation. According to AIT, responses to a threat differ depending on its qualities, e.g., whether it is identified as unfamiliar or noxious. Most importantly, we assume that this threat perception can transition from unspecified to harmful, for instance, when a threat persists (Berkowitz, 1989). Lazarus (1991) similarly asserts that anger frequently arises subsequent to an appraisal of blame or injustice, a sequence that may be preceded by fear upon initial exposure to a threat. Thus, fear can not only transition to anxiety but also to anger, the arguably more potent emotion in explaining political attitudes and behavior. Fear, for instance, may reduce, whereas anger might increase partisan support, e.g., for the far right (Marcus et al., 2019). Fear or anxiety may even be a precondition or gateway for translating negative feelings into radical political attitudes or actions. Given the major prevalence of fear-eliciting content on social media (Saha et al., 2023), understanding the role of fear and anxiety in political communication is pivotal. We first briefly outline the (political) psychological mechanisms and move on to their broader strategic potential within political communication.

Politically hijacked fears

Fear and anxiety generate vulnerabilities that extremists can exploit for radicalization purposes. While previous research has primarily examined anger and rage for affective mobilization, particularly in the context of extremism and radicalization (Tausch et al., 2024), there is a notable scarcity of studies investigating the prevalence, strategic utilization, and effects of FS in online environments, despite evidence supporting its relevance in the political sphere (Marcks and Pawelz, 2022; Saha et al., 2023).

At the individual level, fear-induced information processing can render individuals susceptible to persuasive messaging, effectively creating openings through which specific ideologies may infiltrate and influence. In such situations, messages that promise certainty or offer simple solutions to complex problems are sometimes especially well-received (Huddy et al., 2007). Illustrative examples of these offerings may come in the form of conspiracy narratives, which can be considered special cases of potentially fear-inducing messages. By default, being threat-based, they suggest that (elite) actors secretly plot against the majority's well-being and prosperity (van Prooijen et al., 2022).

While fear can sometimes make the individual more receptive to otherwise unappealing or far-fetched ideas, the elevated potential to exploit fears and anxieties dissipates at the group level: While being threatful, conspiracy narratives simultaneously create a community of the "enlightened," to the point where even diametrically conflicting, new information is reinterpreted in the perspective of the narrative. For instance, Ziolkowski and colleagues (2022) conducted three case studies comparing different ideologies. They argue that dystopian scenarios and doomsday narratives - intended to incite fear and distrust in political functioning - are of central relevance in most extremist and conspiracy movements, independent of the ideological orientation. By portraying a shared threat from outside, these narratives support ingroup cohesion and create pressure to act so that they might be used to justify violence against the perceived threat. Similar to conspiracy narratives, Ziolkowski et al. (2022) find that most doomsday narratives in extremist ideology can inhabit, to a varying degree, anti-Semitic features.

Alleviating the threat by being and belonging to a specific group, e.g., the enlightened, is beneficial because it entails special ingroup bonds. Beyond this mechanism, fear is assumed to increase the identification with ingroup members and foster the development of demonized images of outgroups (Freiheit & Zick, 2022; Meiering et al., 2018). Radical right parties seem to capitalize on fear, as they offer scapegoating (e.g., blaming immigrants) as a means to create and control the perception of threats or to mobilize for their causes (Rothschild et al., 2012; van der Brug & Fennema, 2007).

Specifically in the context of the far right, fear may indeed be used and even encouraged as the transmissional element to anger (Sauer et al., 2023): The elite and groups like refugees and feminists are blamed for societal losses, channeling fear into anger as a strategy to rally support and restore a sense of lost dominance and agency (Hochschild, 2016).

Immigration is a prevailing issue in this context. It is likewise infused with fear (Kopytowska & Chilton, 2018), presumably fueling anxieties to then exploit the resulting need for protection. According to Albertson and Gadarian (2015), discussions about immigration among political elites in the U.S. often incorporate language connoting danger and threat. Their research also reveals that advertisements and political rhetoric concerning immigration, especially illegal immigration, effectively utilize anxiety-inducing language and ominous imagery to sway public opinion. In Europe, worries about immigration are also a key factor in support for far-right parties (Golder, 2016). Unsurprisingly, then, FS is usually "targeted" (Buyse, 2014), for example, against an entity, institution, group, or person. Conceptually, speech that is aimed at inducing fear is considered FS.

Why you should be worried: Strategic fear speech

One of the first works to introduce the concept of FS was by Buyse (2014), who discusses the role of fear in violent conflict escalation and, attached to that, the legal debate around the interaction of violence and discourse. Regarding violent conflict escalation, Buyse (2014) notes that one factor that leads to violence is the perception of fear inside one's own group. Fear, rather than hatred, is the prime driver in ethnic conflicts (Lischer, 1999). Thereby, fear, "rather than hate speech … may be more relevant when assessing violent conflict escalation" (Buyse, 2014; p. 785). Regarding the legal debate, Buyse states that it is "...one of the most difficult issues in debates on the freedom of expression."

Klein (2021) sees a similar challenge faced by social networks in tackling hate speech, which is often disguised as fear-mongering and identity politics, making it difficult to be flagged by monitors. The study examines six hate crimes from 2019 that were

preceded by social media posts, analyzing the rhetoric used by the assailants. It finds that the expressions of cultural paranoia and fear were more common than direct calls for violence. This result of "cultural paranoia" also supports the idea that while all extremist ideologies have FS in the form of doomsday narratives at their core, the specific narratives vary with respect to the specific ideology (Ziolkowski et al., 2022). FS is thus not homogenous across individuals and groups but depends on the "target," which varies considerably across individuals and also across larger and smaller communities. Thus, although FS can be considered a rather general indicator of radicalization,[3] its use is most likely tied to specific narratives, outgroups, or personal motives.

FS was examined by Sayimer and Derman (2017) as a more in-depth and perhaps hazardous kind of hate speech. They analyzed anti-refugee YouTube videos and explained how fear rationalizes and legitimizes racism, providing the justification for some viewers to feel the need to act. Violence labeled as "self-defense" becomes more appropriate. Most studies on FS focus on threats rooted in group-focused enmity (Heitmeyer 2002), such as Muslims and migrants (Saha et al., 2021; Saha et al., 2023) or refugees (Chitrakar, 2020), while other groups that are often portrayed as threat by radical actors, such as the (perceived) elite (e.g., politicians, scientists, media), have rarely been considered (Gagliardone et al., 2016).

In one of the first studies with a large data corpus, Saha et al. (2021) used computational methods to study the distribution of FS in Indian WhatsApp groups but concluded that classification of FS is a complex task. Their classifier did not yield satisfactory results. Expanding their approach, in 2023, Saha et al. studied the prevalence, distribution, and interaction on Gab, a platform with no content moderation, and briefly compared it to Twitter and Facebook. They found FS not only more prevalent than hate speech but that it also spread faster and was less frequently detected by toxic speech

[3] Radicalization can be defined as "the increasing challenge to the legitimacy of a normative order and/or the increasing willingness to fight the institutional structure of this order" (Abay Gaspar et al., 2020, p. 5).

classifiers – even on Twitter and Facebook – because of its more subtle nature. “Their nontoxic and argumentative nature makes them appealing to even benign users who in turn contribute to their wide prevalence by resharing, liking, and replying to them” (Saha et al., 2023, p. 1). Comparable to humorous hate speech (Schmid, 2023), the sender of FS can disguise hostilities. Generally, the intention behind a post remains elusive to a third person. This makes it another communication tool to shift or even poison the climate of opinion in favor of radical or extremist aspirations. Most importantly, and irrespective of whether FS is strategically employed or not, it is linked to both hate and the possibility of violence.

To offer a starting point for empirical research and theoretical discussion and furnish a precise definition amid various related terms like hate speech, we define *fear speech (FS) as any deliberate communicative act that explicitly or implicitly portrays a particular entity, e.g., a group or an institution, as inherently and/or imminently harmful on a cultural, societal, or existential level.*

This means that FS communicates information about a threat. It does not directly express any form of contempt, hostility, or even hate speech, but installs fear as the critical transmission and affective backbone (Buyse, 2014; Ziolkowski et al., 2022). This is crucial, since direct threats toward a person or a group are usually subsumed under concepts like intolerance (Kümpel & Unkel, 2023) or toxicity (Kim et al., 2021), which is why we have refrained from coining it *threat speech*, even though FS does not entail direct expressions of fear itself.

While previous work has focused on the type of portrayed threat (Chitrakar, 2020; Klein, 2021), we conceptualize FS in relation to *how* the threat is communicated. The first and necessary, but not sufficient, component is the presence of a detectable threat, and optionally, whether it is coupled with a call for action. Further crucial markers are the use of what we coin *affective flags*. Affective flags are either expressions that convey strong emotional content and subjective views, often used to emotionally charge discourse, e.g., by using cataclysmic terminology (like disaster, terror, or collapse) or mark clear

distinctions between social groups. These mechanisms serve to intensify emotional responses and demarcate ingroup and outgroup boundaries, effectively shaping perceptions and attitudes toward certain subjects or groups.

Building on that, three broader categories of FS can be distinguished. *Indirect FS* communicates a threat, but its harmfulness is either not made explicit or the threat is rather used to build up subjectivity, e.g., antagonisms. As FS messages are often understood as *fear-mongering* or creating identity politics (Klein, 2021), we capture this broadly as threat-based communication in this category (example: "The next big scam?"). *Direct FS*, in contrast, explicitly illustrates a threat as well as its detrimental consequences. The threat is the main focus of the message, which is why we alternatively call this *threat-focused FS* (example: "Planned Parenthood cartoon by UHR encourages kids to turn to deadly puberty blockers if they feel they are transgender." [own translation]). Lastly, we coin *efficacious FS* as a subform, in which direct FS is paired with a call to action to increase the recipient's self-efficacy and to elevate the pressure to act against the threat. Examples of efficacious FS include such calls to action as "Lock them up!" or "Enough… we have to rise up!".

Efficacious FS is grounded in theoretical considerations and qualitative evidence. First, the Extended Parallel Process Model (EPPM) (Witte, 1996) suggests that when individuals are confronted with fear-inducing messages, they process these through two main paths: threat appraisal and efficacy appraisal. Both appraisals shape an individual's response to a fear-inducing message. From a strategic perspective, it seems essential that the fear message is strong enough to capture attention and be perceived as a genuine threat, yet it is also accompanied by a clear, believable, and actionable effective message to empower subsequent action rather than inducing defensive responses. Second, previous qualitative examinations highlight the pressure to act by using fear-inducing communication in dystopian narratives (Ziolkowski et al., 2022).

Specifically for direct FS, we assume that it intentionally creates or promotes a climate of threat, vulnerability, and anxiety in order to justify the author's inclination or

increase the recipient's propensity for radical attitudes (e.g., political agendas, ideologies) and behaviors (e.g., mobilization and violence). One consequence of FS, therefore, resembles other forms of hostile speech; it contributes to a climate of hostility. For hateful messages, research has already demonstrated that it can promote social division, polarization, and ultimately radicalization through verbal violence and the expression of contempt of or even demonization of outgroups (Romero-Rodríguez et al., 2023). A fearful climate may, at the least, provide the breeding ground for these processes, but the demonization of outgroups especially parallels the potential effects of FS, suggesting that some of these findings may be transferable.

The fact that FS promotes us vs. them thinking by picturing a perpetrator reveals group consolidation as its central function (see Ziolkowski et al., 2022). FS, especially in a strategic understanding, may be endorsed to this end depending on how issues provide a clear, common thread around which antagonistic collective identities and agencies can form, even when the outgroup is rather vague (e.g., elites). Thereby, COVID-19 could harbor comparably high strategic FS utilization potential, as it affects a (perceived) large local ingroup severely threatened by a clearly and collectively identified perpetrator (the elites/the government). Conversely, other crisis issues, even when they also affect a very large group of people, like inflation and the energy crisis, are less delineable with respect to the relation of the (size of the) perceived ingroup and the threatening entities, as with the Russian invasion of Ukraine.

Importantly, however, it is difficult to assess FS as a harmful or strategic communication from the outside. Explicit calls for action in FS could be an indicator of the strategic use of fear. Yet, a fearful post may be interpreted as either a sincere expression of paranoia by hysterical people or as a clever trick where bigotry is disguised as fear in order to justify one's hatred. In some instances, it may even require substantial knowledge about the social and group historical context to accurately interpret communication as potentially fear- or anxiety-eliciting. This dual nature makes FS a notoriously difficult subject matter for platform moderation efforts.

In sum, FS has been shown as a possible driver of violent acts (e.g., Buyse, 2014; Marcks & Pawelz, 2022) and may have the same capacity to deliver hate as hate speech itself while provoking fear and cultural paranoia (Klein, 2021). Saha et al. (2021, 2023) found less-moderated platforms, including instant messengers, to be highly effective for FS distribution. Following on previous work, this study focuses on a different platform that has not been considered in relation to FS prevalence but is also well known for its deliberate relinquishment of content moderation practices: Telegram.

Telegram as a potential hotspot for fear speech

According to Benesch (2023), two conditions are necessary to classify a speech act as dangerous: the message must be inflammatory, and the audience must be susceptible. Extremists use a wide array of digital communication opportunities proficiently and efficiently to directly target susceptible audiences with their strategically tailored messages. Currently, the most important platform for target group-specific addresses of radical and extreme speech is the platform Telegram. Telegram offers several advantages over other social media platforms, making it particularly relevant for extremist actors and audiences. Most notably, it advertises itself as a platform for free speech and, as such, rarely deletes extremist content or bans extremist actors while also promising not to cooperate with security authorities. As one of the most used instant messaging platforms, Telegram is often used to talk to family and peers over a personal mobile phone, which creates a highly personalized communication and reception setting that can often blur the lines between highly credible content from trusted sources, such as peers, and strategically framed content, increasing the susceptibility to tailored messages presented on this platform (Schulze et al., 2022).

However, receiving strategic messages on Telegram does not happen by accident. Unlike Facebook or YouTube, Telegram does not entail algorithmically curated content or recommendations; rather, people must actively seek out and join channels and groups to receive their messages. Often, channels and groups advertise like-minded accounts and

thereby create networks of channels and audiences with similar interests. Since a basic interest in a theme or ideology can be considered a prerequisite for susceptibility to the presented messages, we, therefore, assume that audiences of relevant Telegram channels have at least a minimum susceptibility, which is necessary for fear messages to have an impact.

In sum, FS may be "one of the key dangers emanating from the discourse on alternative platforms online" (Guhl et al., 2022, p. 29) and a significant indicator of radicalization dynamics online (Schulze et al., 2023; Greipl et al., 2022). At times characterized by political upheavals, socioeconomic disparities, and global crises, the ground is fertile for radical ideologies to take root. Against the background that crises increase fear and threat perceptions, fear has a substantial impact on personal and collective motivation and behavior and is strategically exploited by different actors aiming at nurturing uncertainties, we ask *how crisis-specific and prevalent fear speech is in the online communication of extremist or protest movements*. Further, FS is usually tied to a harmful perceived outgroup, which we assume is different across communities. Previous work introduced a useful distinction into three larger identifiable movements with a propensity to exhibit radicalization dynamics on Telegram, namely conspiracy, far-right, and COVID-19-focused actors. This enables us to get differential insights into the community- or group-specific use of FS.

Methods

BERTopic Modeling and Distilbert Automatization

To study the prevalence of FS in the online communication of radical and extremist actors, we analyzed the German language Telegram communication of three different groups of actors known to be highly present on this platform: far-right, COVID-19 protest, and conspiracy (for similar approaches, see Jost & Dogruel, 2023; Schulze et al., 2022). Focusing on the German language, in addition, allowed us to circumvent the pitfalls of

multi-language classification tasks that are specifically pronounced for complex language constructs (Pires et al., 2019). We employed a single platform design to keep platform features and affordances constant, enabling the development of a better functioning instrument and classifier. The most similar design of the observed groups allowed a more precise understanding of the differences in strategic communication with respect to the prevalence of FS. To automatize the FS detection, we collected one large-scale dataset for over a five-year period consisting of over 5.99 million unique posts (see Figure 1).

Data Collection and Actor Classification

To collect Telegram data, it is first necessary to create a list of actors (i.e., channels and groups) active on this platform and relevant to the research interest. The actor collection and classification for this paper relied on previous publications that focused on large-scale analyses of COVID-19 protest groups, far-right or conspiracy actors (Schulze et al., 2022; Rothut et al., 2023; Schulze 2021), resulting in a final list of 3905 actors. All three instances of actor collection and classification followed a similar approach.

First, field research resulted in a manually curated list of relevant actors, which was then used for several instances of snowball sampling based on the mentions and forwards included in the posts. Three iterations of snowball sampling proved sufficient indicating saturation. This approach is widely used in Telegram research. (For a detailed discussion of Telegram actor collection and classification, see Jost & Dogruel, 2023). This resulted in extensive lists of Telegram actors, all of which were manually reviewed using the available actor information (i.e., name, self-description, and posts) based on a prior created and tested classification scheme in the third step. In brief, all actors that affiliated with the COVID-19 protest by name (in Germany, e.g., Querdenken translates to “lateral thinking”) or primarily distributed COVID-19-related protest information including mobilization calls were classified as *COVID-19-focused actors*. All actors that presented clear indications of far-right ideology (e.g., nationalism, authoritarianism, exclusionism following Carter (2018), Hawkins et al. (2018), and Mudde (2002) were classified as *far-right actors*.

Telegram actors were classified as *conspiracy-focused* once it was apparent that the central aim of the actor was to distribute and discuss one or several conspiracy narratives (e.g., QAnon). Conspiracy narratives seek to interpret events by suggesting that powerful institutions, groups, or individuals collaborate covertly to pursue a perceived sinister goal for their own advantage (Schulze et al., 2022; Popper 2003).

Several authors and trained student assistants were involved in the classification process, and all inconsistencies, as well as critical cases, were extensively discussed. Importantly, it must be noted that a perfectly distinct classification of these actor types is not possible, as there are significant overlaps, both rooted in the respective ideology and specifically on Telegram (Rothut et al., 2023; Zehring & Domahidi, 2023). For example, most far-right movements are based on conspiracy narratives (e.g., the Great Replacement) and especially during the COVID-19 pandemic, the all-encompassing topic resulted in these actor types being highly connected through the collective aim of criticizing and destabilizing the government. Therefore, the classification categorized the actor types along the main focus of the channels/groups.

All publicly available content of these actors was scraped in June 2023 using Python and the Telethon library (Lonami, 2019). Considering the quick deletion rates of Telegram content (Buehling, 2023), we merged this dataset with previously collected datasets, leading to the creation of dataset *D0* (for an overview, see Figure 1). Initially comprising approximately 34 million posts, *D0* was reduced to 5.99 million unique posts by 1856 actors after preprocessing, as detailed in the appendix (e.g., duplicate deletion). The time frame spans from the beginning of the channels in 2015 to June 2023, which also serves as the time frame for the subsequent analyses. The pre-processed dataset (*DA*) of 6 million posts was halved into two sub-datasets based on random selection *DB* and *DC*. Next, *DB* was used for domain adaptation to improve the performance of our language model (Sanh et al., 2020), and DC was used for all subsequent topic modeling analyses (Grootendorst, 2022) as well as automated classification, and DD was used to specifically

depict dynamics across the largest possible time frame. From *DC,* various further datasets were created that were used for the next steps in the following chapters.

Operationalizing Fear Speech

We are aware of only two papers that annotated online content for FS (Saha et al., 2021; Saha et al., 2023). However, the annotation guidelines were too domain-specific and not transferable to our research aim. In addition, a binary content classification quickly presented as too simplified to adequately account for the complex nature of FS. Instead, we developed a nominal six-scale annotation scheme (0-5) to capture and distinguish the varying degrees and nuances of FS manifestations. As stated previously, coding depends on the presence of a threat and the use of affective flags. Affective flags are expressions that signal emotional charging (e.g., cataclysmic terminology like "disaster") or demarcation (expressions indicating one's own or the targeted group). "Code 0" was annotated if there was no threat, and therefore, no indication of FS was present. Posts that raised concerns that contained no or only marginal affective flags were coded as 1. Usually, these are rationally laid out arguments or those that raise concerns (1), which *means* rationally laid out depictions of concerns or objective reporting about and expressions of potential threats were coded here. The scale continues with *indirect FS* (2), which refers to posts that often employ mild or ambivalent fear speech, such as fear-mongering that usually contains a threat but provides no or ambivalent threat depiction; the decision is left to the recipient as to why the threat is legitimate. *Direct FS (3)*, in contrast, elaborates on the threat and, e.g., explains why one should be worried. *Efficacious FS (4)* is also direct FS but combined with a specific or unspecific call to action. Coding in stages 3 and 4 can be very similar in length and elaboration as concerns (1) but are marked by affective flags. Lastly, *hostile speech* (5) was annotated if threat perceptions in the posts were present, but explicit expressions of hostility in contrast to fear were the dominant elements of the posts. This is particularly applicable when the language steps away from describing the threat in a fear-inducing manner and leans more toward expressing frustration or anger, indicating a communicative escalation from fear-based to hostility-

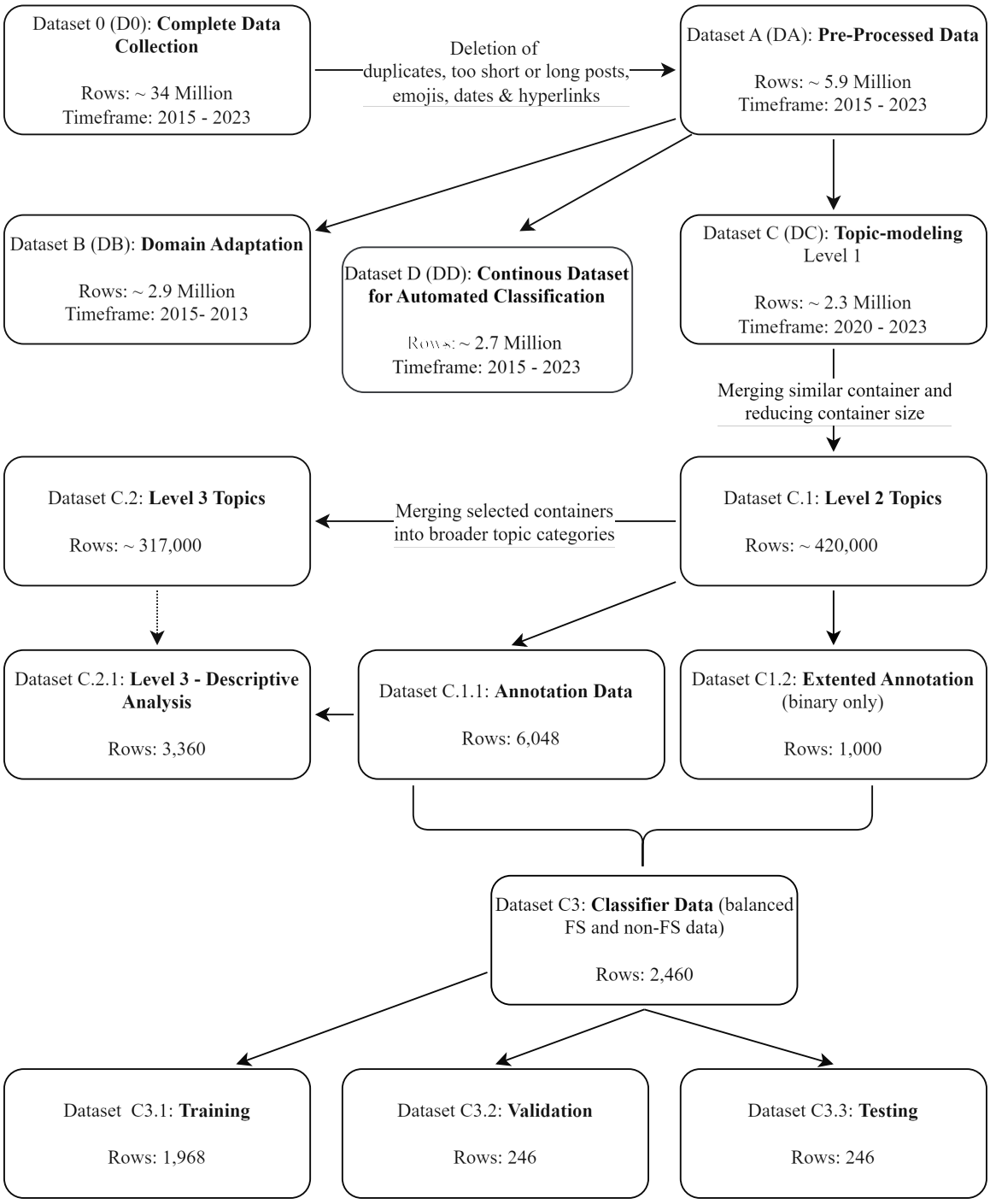

Figure 1. Flowchart of data collection and further pre-processing

centered discourse. At this point, affective flags transmit from dystopic to aggressive and derogatory (the entire codebook including examples and a demonstrator is available in the online Appendix).

We drew 100 posts at random and tested the codebook on two variants: Holsti reached a value of 94% for the six-scale variant (two main coders) and a Holsti value of 88% for the binary version (two main coders and two additional coders for the annotation of further training data) in the pretest. The agreement, precision, and recall for each value of the scale consistently reached values above 87.5% (see Table S1 in the online Appendix).

Topic Modeling

Topic modeling was performed using the BERTopic modeling technique that leverages transformer and c-TF-IDF techniques to cluster words into interpretable topics (Grootendorst, 2022). BERTopic computes these word clusters by first converting the content of Telegram posts into vector representations, reducing the dimensionality using UMAP, and creating bag-of-words representations for each cluster that was extracted from a class-based TF-IDF model. As a language embedding model, All-mpnet-base-v2 was used, which is currently the best-performing model available in the BERTopic library and it also uses posts created on social media as training data. Based on a qualitative inspection of a diverse sample of posts of various lengths, we identified that messages shorter than ten characters typically lacked substantial content, mostly consisting of single words or emoji-related strings. Conversely, posts over 1000 characters frequently covered multiple topics or contained repetitive information, posing challenges in accurately identifying the main topic and classifying fear speech. Consequently, to balance data comprehensiveness and analytical precision, we excluded posts shorter than ten characters and those exceeding 1000 characters and removed emojis, numbers, and hyperlinks. We also limited the time for topic creation to 2020-2023 to reduce the number of possible containers calculated.

Only a very small part of our data was published before 2020, but the longer time of five years from 2015 to 2020 would have resulted in a disproportionate increase in containers.

This reduced the number of posts for the topic model from 2.81 million to approximately 2.26 million posts in *DA*. The model parameters were kept on default except for reducing the number of components in UMAP (from 5 to 2) and changing the minimum (15) and maximum (0.9%) occurrence of terms in all posts in the count vectorizer to keep the size and of our TF-IDF model within a manageable limit (see online Appendix).

Topic creation was set to "auto" and incrementally reduced from 22,500 to 1,500 containers while we observed the automatic container conversion in each reduction step. At 1,500 containers, the smallest container became meaningful to interpret, and hence, we started to validate and label these as a specific topic qualitatively. Qualitative inspection included the decision to merge similar containers and to delete containers that could not be connected to a political topic (e.g., containers containing recurring elements, such as weekdays, signatures, and advertisements). After manually merging similar containers, we reinitiated the topic model, relabeled all containers, and constructed base topics, extracting 18 of the most prevalent political topics (*C.1;* level 2 extraction; n ~420,000). To reduce the complexity of the descriptive analysis, we finally limited the number to six final topics by deleting or collapsing the topics of level 2 (*C.2;* level 3 extraction; n ~ 370,000).

Coding Sampling

After validating the annotation scheme and selecting relevant topics, we conducted a stratified random sampling to create the sample for manual coding. We assumed that FS differed from topic to topic. To maximize FS diversity, we sampled the dataset so it had equal shares of topics derived from the topic modeling and then sampled each group to account for possible semantic differences. Additionally, we stratified the sample by time, specifically creating quarterly time slices, resulting in 6,048 posts that were manually coded in *C1.1*. The allocation of FS to non-FS was unequally distributed (~13% FS), and deemed insufficient, especially the FS-coded training data. Thus, we conducted a second

round of annotation of an additional 1,000 posts (*C1.2)* to increase the number of FS posts in our training dataset. Finally, we compressed the six-scale coding scheme into a dichotomous scale where 0 equals no FS and 1 equals FS. This resulted in a training dataset that consisted of 2,460 equally split posts in dataset *C3*.

Domain-Adaptation

Similar to the topic modeling approach, we used a transformer-based approach to classify the second half of our data (Vaswani et al., 2017).[4] DistilBERT is a variant of the BERT transformer model and is trained on data from Wikipedia, books, subtitles, and news crawls. It is reduced in model size while retaining 97% of its language understanding capabilities (Devlin et al., 2019). However, we assumed the language to differ in accordance with the platform on which it was posted, the authors, and their ideological background. Thus, a language model focusing on far-right German language Telegram data would have been more suitable for our use, but at the time we conducted this research, it did not exist.

Thus, before training our model with our annotated training data, we used the remaining half (~3 million posts) of our initial dataset (*D0*) to conduct a process called domain adaptation (*DB*; Tunstall et al., 2022). Domain adaptation can be an intermediate step in classification tasks to pretrain an "off-the-shelf" language model to the language used in the training data. It assumes that "machine-learning models can learn 'language knowledge' and 'task knowledge' during a pre-training phase and store this 'knowledge' in their parameters" (Laurer et al., 2023, p. 2). When conducting the actual fine-tuning of the model using the manually annotated training data, the model can transfer the in-domain generated knowledge by the domain adaptation to increase the performance of the classification tasks (see the Appendix for the link to the model).

[4] The language model is available at: https://huggingface.co/distilbert-base-german-cased.

Classification

In the last step, we used the domain-adapted DistilBERT model within the Python transformers library and built the classification algorithm. Only *direct* (3) and *efficacious FS (4)* were considered FS postings for the classification algorithm. The rationale behind this decision was to keep the classification of FS as conservative as possible. We assumed that including every potentially threatful message would lead to an overestimation of FS within our dataset and, thus, it would reduce classification precision overall. Similar problems in which increased ambiguity or covertness led to reduced classification quality were previously evident for hate speech (Benikova et al., 2018; Zhang & Luo, 2019). In contrast, direct or efficacious FS, as clearly threat-focused communication, is much less ambiguous in terms of how the message is crafted. We did a random 8:1:1 (n=1968; 246; 246) training (*C3.1*), validation (C3.2), and test (C.3.3) data split. After hyperparameter optimization on the validation set, we achieved a macro-average F1 score of .82 on the validation set and .79 on the final test set with consistently balanced and robust precision and recall metrics above .76 (see Table 1 for details; the link to the classifier is available in the Appendix).

Table 1. Detailed performance metrics for the fear speech (FS) classification algorithm

	validation set			test set			
	precision	recall	f1-score	precision	recall	f1-score	support
no FS	0.84	0.78	0.81	0.81	0.76	0.79	123
FS	0.80	0.85	0.82	0.78	0.82	0.80	123
accuracy			0.82			0.79	246
macro avg	0.82	0.82	0.82	0.79	0.79	0.79	246

Results

Overview and Topic Modeling

Considering our aim to investigate the use of FS across salient political events, we focused on the five crisis-related themes most prevalent in the data, coinciding with the most salient crisis themes during our observation period.[5] Additionally, we created a sixth topic container consisting of conspiracy-related content to inspect the classifier's performance and the role of FS in conspiracy narratives. Our final topic model consisted of six topics revolving around the COVID-19 pandemic, conspiracy narratives, as well as the Russian invasion of Ukraine (RioU), the energy crisis, inflation, and migration.

These topics aggregated to 317,299 posts in total. By far, the largest topic in our sample consisted of *COVID-19*-related posts (N=132,501), followed by posts about *RioU* (N= 68,518) and with a greater gap *Conspiracy Narratives* (N=43,216), *Energy Crisis* (N=29,053), *Inflation* (N=26,562) and *Migration* (N=17,449). In terms of temporal dynamics, COVID-19 was the most dominant topic in 2020, 2021, and early 2022. Around February 2022, the salience of COVID-19 was incrementally replaced by the topics *RioU, Energy Crisis, and Inflation*. *Conspiracy Narratives* and *Migration* were evenly distributed across the whole observation period.

Inspecting the distribution of posts by classified actors in Table 2, conspiracy-focused actors were responsible for the majority of posts on these topics, followed by far-right actors, and COVID-19-focused actors with a greater distance. Accounting for prevalence in these topics, conspiracy-focused actors had a higher relative share when posting about COVID-19 (16%), RioU (10%), and Conspiracy Narratives (9%). The far-right actors, in contrast, were more diversified as they had the highest shares, especially in Energy Crisis (5%) and Migration (4%) while having equal shares in COVID-19 (16%)

[5] See online Appendix for a detailed list of topics and their proportions.

with the conspiracy-focused actors. Lastly, the COVID-19-focused actors mainly focused on COVID-19 (10%) and only had minor shares on the other topics.

Fear Speech Proportions (manual coding)

Table 2. Proportions of manually coded fear speech (FS) posts (Dataset C.2.1) by selected crisis-related topics and actor-focus

Topic	Actor-focus	no FS	raising concerns	indirect FS	direct FS	efficacious FS	hostile speech
COVID-19	Far-right	27.2	23.7	39.6	7.7	1.5	0.3
	COVID-19 Protest	18.5	34.5	32.4	10.4	3.3	0.9
	Conspiracy	20.7	25.4	17.7	21.0	15.0	0.3
Conspiracy	Far-right	13.9	20.8	43.9	15.4	3.6	2.4
	COVID-19 Protest	31.2	10.1	49.4	7.7	0.9	0.6
	Conspiracy	48.4	6.6	30.7	11.3	1.2	1.8
Inflation	Far-right	21.9	39.5	22.8	13.2	0.9	1.8
	COVID-19 Protest	24.1	24.1	31.2	15.2	5.4	0.0
	Conspiracy	21.8	40.0	23.6	10.9	3.6	0.0
Migration	Far-right	9.7	12.4	62.8	9.7	1.8	3.5
	COVID-19 Protest	29.5	22.3	41.1	4.5	1.8	0.9
	Conspiracy	35.1	39.6	14.4	8.1	1.8	0.9
Russia's Invasion of Ukraine	Far-right	44.2	20.4	30.1	4.4	0.0	0.9
	COVID-19 Protest	50.9	24.1	17.0	4.5	2.7	0.9
	Conspiracy	42.3	32.4	16.2	7.2	0.9	0.9

Topic	Actor-focus	no FS	raising concerns	indirect FS	direct FS	efficacious FS	hostile speech
Energy Crisis	Far-right	25.7	24.5	31.9	7.1	0.9	0.0
	COVID-19 Protest	25.9	40.2	22.3	10.7	0.9	0.0
	Conspiracy	30.6	29.7	13.5	20.7	4.5	0.9

In the detailed exploration of the manual FS coding across diverse topics, a nuanced landscape emerged, revealing variations in the proportions of different levels of FS (see Table 1).

Within the *COVID-19* topic, a spectrum of FS was also observed across types. Approximately 22.1% of posts exhibited no FS, while a higher proportion of 27.9% was categorized as raising concerns. Indirect FS was represented by 29.9% of posts, and a similar proportion of 13.0% depicted direct FS. Efficacious FS and hostile speech were less prevalent, constituting 6.5% and 0.5% of posts, respectively. Conspiracy-focused actors had the highest direct and efficacious FS shares, with 21.0% and 15.0%, respectively.

Transitioning to the *Energy Crisis* topic, approximately 27.4% of posts were labeled as no FS, and 31.5% as raising concern. The presence of indirect FS was noted in 22.6% of posts, while direct FS accounts were 12.8%. Efficacious FS and hostile speech were comparatively minimal, at 2.1% and 0.9%, respectively. Conspiracy-focused actors had the highest shares of direct and efficacious FS (20.7% and 4.5%, respectively).

In the realm of *Inflation*, the data unveiled 22.6% of posts with no FS and a substantial 34.5% raising concern. Indirect FS was present in 25.8% of posts, with direct FS at 13.1%. The efficacious FS and hostile speech proportions were lower, at 3.3% and 0.6%, respectively. Group shares were comparably large across coding categories.

Examining the *Migration* topic, 24.8% of posts did not contain FS, and 24.8% were categorized as raising concerns. The proportion of indirect FS was relatively high at 39.4%, whereas direct and efficacious FS were relatively low at 7.4% and 1.8%, respectively. Together with the topic Conspiracy Narratives, Migration elicited higher proportions of hostile speech at 1.8%. Indirect FS and hostile speech were especially driven by far-right-focused actors (62.8% and 3.5%, respectively).

Delving into the *Russian Invasion of Ukraine*, approximately 45.8% of posts were identified as having no FS, and 25.6% as raising concerns. Indirect and direct FS proportions were 21.1% and 5.4%, respectively, while efficacious and hostile FS were less common, constituting around 1.2% and 0.9% of posts, respectively. RioU was the least threatful topic in our sample, and only actors with a far-right focus had heightened shares in indirect FS (30.1%).

Lastly, the *Conspiracy Narratives* topic presented a diverse distribution with 31% of posts exhibiting no FS and 13% raising concerns. A notable 41% of posts were classified as indirect FS, and 12% as direct FS. Efficacious FS was observed in 1.9% of posts and hostile speech in 1.6%. Surprisingly, the highest amount of FS - across all categories - was published by the far-right-focused actors, while conspiracy-focused actors linked this topic the least to FS.

In summary, this detailed analysis elucidates the heterogeneous nature of FS across these various topics. The distribution of proportions for each level of FS revealed distinctive patterns, highlighting the diversity and complexity inherent in the discourse across different types and topics.

FS dynamics and proportion (automated analysis)

Regarding the classified continuous dataset (Dataset DD), the overall dynamics of the posting frequency seemed to correspond closely to the dynamics of overall FS postings

(see Figure 2). Both showed an increase from the start of 2020 and peaked in the last quarter of 2021 and the first quarter of 2022. As the dynamic increased overall, not only did the overall FS frequency increase but also its relative proportion. Starting in 2018 with less than 5% FS in all postings, there was a steady increase until the last quarter of 2021, reaching a maximum FS share of over 25%, which stayed above 20% until the end of the observations in 2023. At 21%, the FS share in the first half of 2023 corresponded to the average amount of direct FS in the whole dataset (~2,7M unique posts).

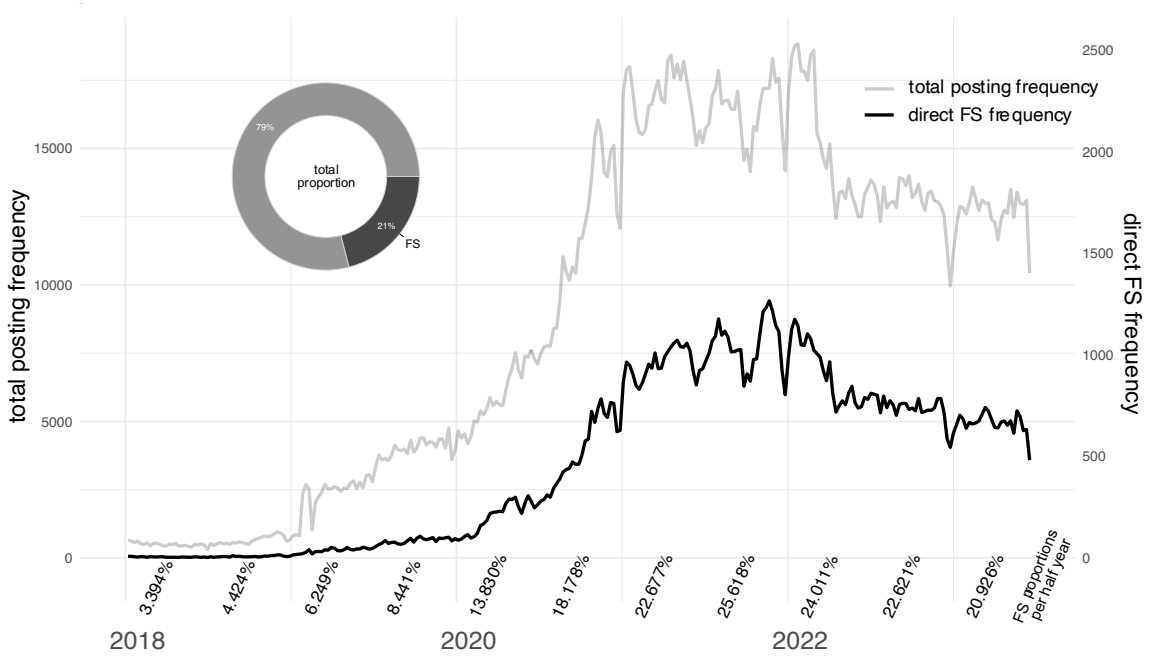

Figure 2. Weekly aggregated posting frequency of the continuous dataset (DD)

Note. The left scale depicts the overall posting frequency, and the right scale depicts the automatically FS classified posts. The pie chart depicts FS classified post proportion in the full dataset.

Regarding the specific topic by actor-focus proportions (see Figure 3) in the *Conspiracy* topic, COVID-19-focused actors stand out with a higher proportion of 43.8% of posts indicating FS, while the conspiracy-focused actors and the far-right actors exhibit proportions of FS at 30.8% and 32.1%, respectively.

For the topic of *Energy Crisis*, all three actor types demonstrated lower proportions of FS, with the highest being 28.1% in COVID-19-focused actors.

Within *Inflation*, the proportions of FS were lower across all types, with COVID-19-focused actors having the highest proportion at 31.8%.

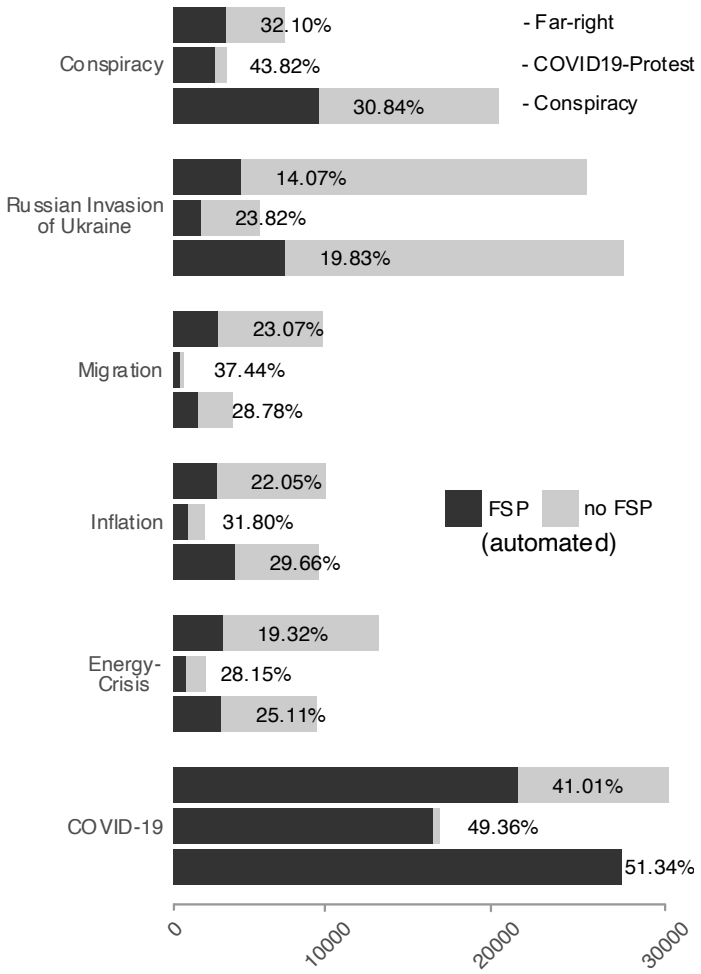

Figure 3. Automated classification of direct fear speech in posts (Dataset C.2; n=317K) regarding selected crisis-related topics.

COVID-19 presented a nuanced scenario in which conspiracy-focused actors and COVID-19-focused actors exhibited nearly equivalent proportions of posts with and without FS, around 51.3% and 49.4%, respectively, indicating a higher presence of FS than other topics.

Regarding *the RioU*, despite the overall lower proportions of FS across all types, far-right actors had a particularly low proportion of FS at 14.1%, making it one of the lowest across all topics and types.

Lastly, in the *Migration* topic, COVID-19-focused actors surfaced with a considerable proportion of FS at 37.4%.

Discussion

In the current work, we analyzed the prevalence of FS in a large sample of Telegram posts (2.9 million) published by three different radical or extremist movements (COVID-19 protest, far-right, and conspiracy-focused) within a selection of six crisis-related topics over more than five years (2015-2023) via manual and automated content analytical methods. For the manual coding, we introduced a new conceptualization of FS, i.e., the strategic instilling of fear through threat-based communication, and we distinguished three types of FS: 1) Indirect FS as threat-based, 2) Direct FS as threat-focused (threat is the primary subject of the message) and a subform of direct FS, and 3) efficacious FS, where direct FS is coupled with a call to action to increase the recipient's self-efficacy elevating the pressure to act against the threat. We applied the coding concept to a large sample of Telegram posts for manual content analysis and used the identified direct and efficacious FS posts as training data for a BERT-based classification model to scale the quantification of FS to the whole dataset. Thus, the analysis focused on broader aspects of FS use, such as prevalence across topics, groups, and time, of Telegram posts and provided evidence for four different aspects of FS usage. Meaningful coherence between manual and automated coding revealed a high prevalence of FS across the dataset, a close relation between posting-volume and FS use dynamics, and group- and topic-specific FS.

In the manually coded sample, COVID-19, Inflation, Energy Crisis, Conspiracy, Migration, and RioU appeared in descending order of FS proportions. In the automated FS coding, COVID-19 and RioU were again ranked at the top and bottom, respectively, but Conspiracy, Inflation, Migration, and Energy Crisis was the in-between order. This difference is most likely due to the sampling strategy of the manually coded data, as it was stratified across groups, topics, and time; furthermore, it was too small to tightly pick up on the true distribution of FS in each topic. However, COVID-19 and RioU consistently marked the endpoints of all FS proportions. One explanation could be a threat-proximity relation. By applying a coding scheme that centers on the portrayed threat, the relevance of the threat to the individual or collective (geographically and/or personally/ideologically) may determine the amount of FS attached to a topic. Additionally, the outgroup attributability, e.g., the possibility of identifying a concrete enemy as the source of a threat, seemed to be reflected here. COVID-19 was the overarching issue in our dataset, with the highest relevance and proximity and, therefore, the highest FS proportions. In contrast, RioU is, geographically, the most distant topic, which is why it is more difficult to attribute to some "dangerous" outgroup. In addition, the threat is not as imminent and individually perceivable as for topics that are linked to RioU but far more relevant (e.g., Energy Crisis), thus showing the least amount of FS. Further, it must be noted that parts of the posts on RioU contained information and reports from the conflict and, thus, presented new and factual information concerning the current situation, more so than other FS-related reinterpretations of these events, further explaining the relatively low amount of FS in this topic. Reversing the perspective, the amount of communication without threats (coding=0) places topics in a similar rank as with FS, with COVID-19 having the least amount of communication without threats and RioU having the most, further substantiating the correspondence between manual and automated coding.

Manual and automated coding revealed group differences. Group-specific posting frequency is, across topics, roughly concurs with overall posting shares by the group. Unsurprisingly, the only deviances are disproportionately occurring COVID-19-related

posts by COVID-19-focused actors and disproportionately many Conspiracy-related posts by conspiracy-focused actors. Regarding FS shares in the automated analysis, conspiracy-focused actors and COVID-19-focused actors tended to exhibit more direct FS than the far-right actors. However, the manual content analysis is rather inconspicuous overall in terms of the proportion of group-specific direct FS. Conspiracy-focused actors only show noticeably greater shares with respect to the topics of COVID-19 and Energy Crisis.

This changes when indirect FS is included. The far-right actors seemed to be more inclined to use indirect FS, which often indicates "fear-mongering" (see, e.g., Klein et al., 2021) or identity-priming antagonisms. In any case, such statements build on or are accompanied by a noticeable threat. Regarding the topics COVID-19, Energy Crisis, and especially Migration, far-right actors have the highest shares of indirect FS. Including the observation that the far-right actors usually have the highest shares of hostile speech in the annotated sample, this completes the impression that communication around fear and threats among this group may be more firmly grounded on identity, demarcation, and, finally, hostility. The distinct use of FS among various radical and extremist groups on Telegram may impact our understanding of group dynamics and communication strategies. Far-right actors predominantly engage in indirect FS to subtly foster a narrative of identity and demarcation, indirectly reinforcing group boundaries and hostility without pinpointing specific threats. This strategy effectively embeds fear within a broader cultural and identity preservation narrative, mobilizing support through a nuanced "us-versus-them" framework. In contrast, conspiracy-focused and COVID-19-focused groups lean toward direct FS, aiming to create a sense of immediate crisis and urgency. This form of communication may be designed to prompt action more directly by making threats feel immediate and response efficacy tangible, thereby rallying followers around a clear cause and fostering group cohesion.

Looking at the overall FS shares across the whole dataset of 2.9 million posts, our results in the automated content analysis reveal that a substantial proportion of 21% of posts were classified as direct FS. Within the topic modeling dataset with selected crisis

topics (~317K posts), the amount of FS climbed to 34%. This is substantially greater than the manual content analysis revealed, which is 13.3%, and considerably more than others found on Twitter (10%) or Facebook (4%) (Saha et al., 2023). Extrapolating this perspective by including indirect FS of the manually annotated dataset, we can scale fear-based communication up to around 40-60%, comparable to what Saha and colleagues (2021) found in a preselected, manually annotated sample on WhatsApp.[6] This is worrisome, as we can assume that both forms of FS contribute somewhat to a climate of fear and, consequently, hostility.

Depending on the sector or community within the Telegram sphere, communication seemed thus, to a large extent, characterized by fear, and, in other words, it elaborated on threats as well as drawing lines between "them and us." This not only validates the methodological advancements of our study but also raises critical questions about the role of platform affordances in the amplification of FS, suggesting that Telegram's unique features might facilitate a higher prevalence of FS compared to more public and moderated platforms like Twitter and Facebook. Our findings thus extend the empirical investigation of FS by demonstrating significant variances in FS prevalence and typologies across different digital environments, contributing to a deeper understanding of how technological and contextual factors influence the dynamics of online radical discourse.

If the radical Telegram sphere, at least to some degree, is driven by fear, it will follow that the overall posting volume, as it signals importance (e.g., of a topic), transmits increasing absolute but relative amounts of FS. This seemed to be the case in our data. As the posting volume increased steeply in 2020, the absolute number of FS increased as well. Importantly, the relative amount of FS increased from around 10% to almost 26% at the end of 2022. As the Telegram online sphere gained radical momentum through and with the COVID-19 pandemic (Schulze 2021), this supports the argument that FS is a useful indicator of radicalization dynamics as previously assumed (Greipl et al., 2022; Schulze et

[6] Note that, unlike Saha and colleagues (2021), our dataset represents a strategically collected rather than a convenience sample.

al., 2023; Buyse 2014). Indeed, as a fearful climate stimulates pressure to act, FS may even indicate the potential for crisis escalation. However, while this pattern of FS volume coherence is surprisingly clear on the macro-level, this pattern becomes volatile on more detailed levels, e.g., with respect to single topics. Again, the example is the COVID-19 topic, which is the reference with up to 51% FS (conspiracy), while the RioU topic has the lowest value with 14% FS (far-right actors). This may resemble the idea that endorsement of FS varies with the issue's potential to allow for the creation of cohesive and antagonistic collective identities (and the malleability of the threat). Whereas COVID-19-related fear speech posts often specifically address the threat and outgroup (example: "The PHARMA [MAFIA] is planning another mRNA [attack] on humans!"), in FS related to the Russian invasion of Ukraine, the crisis issue needs to be embedded in a larger threat narrative, such as the "Great Reset" (example: "NATO is at war with the EU against Europe!").

By extension, because the COVID-19 crisis (and other FS-fueled issues to a lesser extent) has also been used for protest mobilization on a large scale, it is reasonable to assume that FS also contributes to dynamics between online communication and offline action. Future work should thus investigate the detailed dynamics of FS use, as it is currently an open question whether FS indicates a more latent climate behind slower communication dynamics and, therefore, drives anxiety rather than fear or whether it represents a more direct, affective component that is directly relatable to short-scale dynamics of communication, for instance weeks or even days.

Finally, a better understanding of FS types employed may allow for more nuanced interventions and support the creation of tailored counter-narratives, effectively addressing the underlying narratives and fears driving radicalization. As revealed in our study, the nuanced understanding of FS dynamics among different extremist groups on Telegram has significant policy implications and suggests several countermeasures. The differentiation between indirect and direct FS, coupled with the platform-specific prevalence of these communications, underscores the need for targeted and sophisticated approaches to counteract online extremism. Foremost, policymakers and platform administrators must

recognize the unique challenges Telegram's encrypted and semi-private nature poses, which facilitates FS dissemination. This calls for enhanced collaboration between tech companies and law enforcement to develop strategies that respect privacy while addressing the propagation of extremist content. The findings advocate for the development of policies that support the detection and moderation of FS, including advanced AI and machine learning technologies to identify nuanced forms of FS without infringing on legitimate free speech. Fear speech, particularly indirect FS, which is more subtle and thus potentially more insidious, may be among the great challenges in detecting and counteracting radical online spheres.

Limitations

Several limitations have to be noted. Considering the high volatility and ephemerality of radical/extremist online communication, retrospective data collection is not able to fully represent the entire German language Telegram communication of the three different actor types for the investigation period. While this arguably would be an impossible or near impossible task, we aimed to mitigate this limitation to the best of our capacity by including several datasets collected at different times. Additionally, since the most extreme content is often deleted either by radical/extremist actors themselves, for example, to avoid prosecution, or by security authorities (Buehling, 2024), we assume that the missing posts would result in an under- rather than an overestimation of FS. Further, since our research interest has focused on crisis-related fear speech, a large part of the collected posts had to be discarded for manual annotation and topic-specific classification, which decreased the sample size from 2.99 million to 311,939 posts. Further, as Saha et al. (2021 & 2023) noted, FS classification is a complex classification task. Although we specifically followed the rationale of constructing a rather conservative classifier by distinguishing between indirect (threat-based) and direct (threat-focused) communication, the difficulty of FS extraction is still high. Concerning the automatic classification, we have to point out that the ratio from manually to automatically coded posts is not ideal, and a larger amount of manually annotated data would be preferable to enhance the quality of

the automation. However, considering the complexity of fear speech coding, we are still content with the F-Score of .79. We further acknowledge that exploring FS across a broader political spectrum, including left-wing groups,[7] presents an intriguing prospect for future research, especially since FS should be universally applicable. Finally, while our study offers a foundational exploration into FS, we precluded a detailed comparison with hate speech. While we believe this comparison to be informative, the focus of our study was initially more on refining the conceptualization of FS and contributing to the (automatic) detection of FS. This limitation underscores the need for future research to explore the intricate relationship between FS and hate speech, enhancing our understanding of their roles in online discourse.

Conclusion

In this study, an extensive examination of fear speech prevalence of 2.9 million Telegram posts across three radical/extremist movements was conducted over a span of five years (2017-2023). The analysis revealed a high prevalence of FS, with a significant proportion of the posts (21%) being classified as direct FS, which increased to 34% when focusing on selected crisis-related topics. The manual and automated coding methods demonstrated consistency, especially in identifying COVID-19 and RioU as topics with the highest and lowest FS proportions. The data suggest a potential threat-proximity relation influencing FS levels. The findings also revealed group-specific tendencies in FS use, with the far-right actors exhibiting a higher inclination toward indirect FS, often aligning with fear-mongering or identity-priming antagonisms. The significant rise in FS from 10% to almost 26% toward the end of 2022, coinciding with a steep increase in posting volume, indicates that the radical Telegram sphere may be marked by fear, especially amidst crises like the COVID-19 pandemic. This underscores FS's potential as an indicator of radicalization dynamics and crisis escalation potential. However, the

[7] We retrospectively gathered data from German language left-wing Telegram channels from movements like Extinction Rebellion, “Die letzte Generation” (translates to Last Generation), and left-leaning news outlets such as the “TAZ” (total N ~ 58K posts for 2019-2023). In a sample of 200 posts, we found fewer than ten instances of direct FS. Thus, the German language left-wing telegram milieu may not only be far less active on Telegram, it also seems much less inclined to use FS.

nuanced variations in FS across different topics and groups highlight the necessity for further investigations to decipher the underlying dynamics of FS use and its implications for communication within radical online spheres.

Funding

This study was supported by grants from the German Federal Ministry of Education and Research within the framework of the program "Research for Civil Security" of the Federal Government and the German Federal Ministry of the Interior. (grant no. MOTRA-13N15223).

References

Abay Gaspar, H., Daase, C., Deitelhoff, N., Junk, J., & Sold, M. (2020). Radicalization and Political Violence – Challenges of Conceptualizing and Researching Origins, Processes and Politics of Illiberal Beliefs. *International Journal of Conflict and Violence (IJCV)*, *14*, 1–18. https://doi.org/10.4119/ijcv-3802

Albertson, B., & Gadarian, S. K. (2015). *Anxious Politics: Democratic Citizenship in a Threatening World.* Cambridge University Press. https://doi.org/10.1017/CBO9781139963107

Aristotle, & Roberts, W. R. (2004). *Rhetoric*. Courier Corporation.

Bar-Tal, D. (2013). *Intractable Conflicts: Socio-Psychological Foundations and Dynamics.* Cambridge University Press. https://doi.org/10.1017/CBO9781139025195

Benesch, S. (2023). Dangerous Speech. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech research* (pp. 185-197). Berlin https://doi.org/10.48541/DCR.V12.11

Benikova, D., Wojatzki, M., & Zesch, T. (2018). What Does This Imply? Examining the

Impact of Implicitness on the Perception of Hate Speech. In G. Rehm & T. Declerck (Eds.), *Language Technologies for the Challenges of the Digital Age* (Vol. 10713, pp. 171–179). Springer International Publishing. https://doi.org/10.1007/978-3-319-73706-5_14

Berkowitz, L. (1989). Frustration-aggression hypothesis: examination and reformulation. *Psychological bulletin, 106*(1), 59.

Buehling, K. (2023). Message Deletion on Telegram: Affected Data Types and Implications for Computational Analysis. *Communication Methods and Measures*, *0*(0), 1–23. https://doi.org/10.1080/19312458.2023.2183188

Buyse, A. (2014). Words of Violence: "Fear Speech," or How Violent Conflict Escalation Relates to the Freedom of Expression. *Human Rights Quarterly*, *36*(4), 779–797.

Canetti, D., Halperin, E., Hobfoll, S. E., Shapira, O., & Hirsch-Hoefler, S. (2009). Authoritarianism, perceived threat and exclusionism on the eve of the Disengagement: Evidence from Gaza. *International Journal of Intercultural Relations*, *33*(6), 463–474. https://doi.org/10.1016/j.ijintrel.2008.12.007

Carter, E. (2018). Right-wing extremism/radicalism: Reconstructing the concept. *Journal of Political Ideologies*, *23*(2), 157–182. https://doi.org/10.1080/13569317.2018.1451227

Chitrakar, R. (2020). Threat perception in online anti-migrant speech: A Slovene case study. *Slovenščina 2.0: Empirical, Applied and Interdisciplinary Research*, *8*, 66–91. https://doi.org/10.4312/slo2.0.2020.1.66-91

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (2nd ed., pp. 323–417). Lawrence Erlbaum Associates, Inc.

Cohen-Chen, S., Halperin, E., Porat, R., & Bar-Tal, D. (2014). The Differential Effects of

Hope and Fear on Information Processing in Intractable Conflict. *Journal of Social and Political Psychology*, *2*(1), 11–30. https://doi.org/10.5964/jspp.v2i1.230

Daniel-Watanabe, L., & Fletcher, P. C. (2022). Are Fear and Anxiety Truly Distinct? Biological Psychiatry Global Open Science, 2(4), 341–349. https://doi.org/10.1016/j.bpsgos.2021.09.006

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. http://arxiv.org/abs/1810.04805

Doosje, B., Loseman, A., & Van Den Bos, K. (2013). Determinants of Radicalization of Islamic Youth in the Netherlands: Personal Uncertainty, Perceived Injustice, and Perceived Group Threat. *Journal of Social Issues*, *69*(3), 586–604. https://doi.org/10.1111/josi.12030

Eibl-Eibesfeldt, I., & Sütterlin, C. (1990). Fear, defence and aggression in animals and man: Some ethological perspectives. In P. F. Brain, S. Parmigiani, R. J. Blanchard, & D. Mainardi (Eds.), *Fear and Defense* (pp. 381–408). Harwood.

Ellis, D., & Tucker, I. (2022). *Emotion in the digital age: Technologies, data and psychosocial life*. Routledge.

Freiheit, M. & Zick, A. (2022). Die Rolle von islamistischen Gruppen und Milieus in der Hinwendung und Radikalisierung von jungen Menschen. In B. Milbradt, A. Frank, F. Greuel, & M. Herding (Hrsg.), *Handbuch Radikalisierung im Jugendalter. Phänomene, Herausforderungen, Prävention* (pp. 247-262). Opladen: Verlag Barbara Budrich.

Gagliardone, I., Pohjonen, M., Beyene, Z., Zerai, A., Aynekulu, G., Bekalu, M., Bright, J., Moges, M. A., Seifu, M., Stremlau, N., Taflan, P., Gebrewolde, T. M., & Teferra, Z. (2016). *Mechachal: Online Debates and Elections in Ethiopia - From Hate Speech to Engagement in Social Media* (SSRN Scholarly Paper 2831369). https://doi.org/10.2139/ssrn.2831369

Goldenberg, A., & Gross, J. J. (2020). Digital Emotion Contagion. *Trends in Cognitive Sciences*, *24*(4), 316–328. https://doi.org/10.1016/j.tics.2020.01.009

Golder, M. (2016). "Far-Right Parties in Europe."*Annual Review of Political Science, 19*(3):477–97.

Greipl, S., Hohner, J., Schulze, H., & Rieger, D. (2022). Radikalisierung im Internet: Ansätze zur Differenzierung, empirische Befunde und Perspektiven zu Online-Gruppendynamiken. MOTRA-Monitor, 42–71.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. http://arxiv.org/abs/2203.05794

Gross, J. J. (Ed.). (2013). *Handbook of emotion regulation*. Guilford publications.

Guhl, J., Ebner, J., & Rau, J. (2022). *The Online Ecosystem of the German Far-Right*. Institute for Strategic Dialogue (ISD).

Hawkins, K. A., Carlin, R. E., Littvay, L., & Kaltwasser, C. R. (Hrsg.). (2018). *The Ideational Approach to Populism: Concept, Theory, and Analysis* (1. Aufl.). Routledge. https://doi.org/10.4324/9781315196923

Helminen, V., Elovainio, M., & Jokela, M. (2022). Clinical symptoms of anxiety disorders as predictors of political attitudes: A prospective cohort study. *International Journal of Psychology, 57*(2), 181–189. https://doi.org/10.1002/ijop.12796

Heitmeyer, W. (2002). Gruppenbezogene Menschenfeindlichkeit. Die theoretische Konzeption und erste empirische Ergebnisse [Group-focused enmity. Theoretical conception and first empirical results]. In W. Heitmeyer (Ed.), *Deutsche Zustände*, Folge 1 Vol. 1, (pp. 15–36). Frankfurt. Suhrkamp

Huddy, L., Feldman, S., & Weber, C. (2007). The Political Consequences of Perceived Threat and Felt Insecurity. *The ANNALS of the American Academy of Political and Social Science*, *614*(1), 131–153. https://doi.org/10.1177/0002716207305951

Hochschild, Arlie R. (2016): *Strangers in Their Own Land: Anger and Mourning on the American Right*. New York/London: New Press.

Isen, A. M. (1990). The Influence of Positive and Negative Affect on Cognitive Organization: Some Implications for Development. In N. L. Stein, B. L. Leventhal, & T. Trabasso (Eds.), *Psychological and biological approaches to emotion* (pp. 75–94). Erlbaum. https://api.semanticscholar.org/CorpusID:149753957

Jarymowicz, M., & Bar-Tal, D. (2006). The dominance of fear over hope in the life of individuals and collectives. *European Journal of Social Psychology*, *36*(3), 367–392. https://doi.org/10.1002/ejsp.302

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Jost, P., & Dogruel, L. (2023). Radical Mobilization in Times of Crisis: Use and Effects of Appeals and Populist Communication Features in Telegram Channels. *Social Media + Society*, *9*(3), 20563051231186372. https://doi.org/10.1177/20563051231186372

Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication, 71*(6), 922-946.

Klein, A. (2021). Social Networks and the Challenge of Hate Disguised as Fear and Politics. *Journal for Deradicalization*, *26*, Article 26.

Kopytowska, M., & Chilton, P. (2018). “Rivers of blood”: Migration, fear and threat construction. *Lodz Papers in Pragmatics, 14*(1), 133–161. https://doi.org/10.1515/lpp-2018-0007

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the*

National Academy of Sciences of the United States of America, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Kruglanski, A. W. (2004). *The psychology of closed mindedness* (pp. xi, 192). Psychology Press.

Kümpel, A. S., & Unkel, J. (2023). Differential perceptions of and reactions to incivil and intolerant user comments. *Journal of Computer-Mediated Communication*, *28*(4), zmad018. https://doi.org/10.1093/jcmc/zmad018

Laurer, M., Atteveldt, W. van, Casas, A., & Welbers, K. (2023). Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis*, 1–17. https://doi.org/10.1017/pan.2023.20

Lăzăroiu, G., & Adams, C. (2020). Viral Panic and Contagious Fear in Scary Times: The Proliferation of COVID-19 Misinformation and Fake News. *Analysis and Metaphysics*, *19*(0), 80. https://doi.org/10.22381/AM1920209

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Touchstone.

LeDoux, J. E. (1995). Emotion: Clues from the Brain. *Annual Review of Psychology*, *46*(1), 209–235. https://doi.org/10.1146/annurev.ps.46.020195.001233

Lischer, S. K. (1999). Causes of Communal War: Fear and Feasibility. *Studies in Conflict & Terrorism*, *22*(4), 331–355. https://doi.org/10.1080/105761099265676

Lonami (2019) *Telethon Revision 7325718f*. Available at: https://docs.telethon.dev/en/stable/# (accessed 06 October 2021).

Marcks, H., & Pawelz, J. (2022). From Myths of Victimhood to Fantasies of Violence: How Far-Right Narratives of Imperilment Work. *Terrorism and Political*

Violence, *34*(7), 1415–1432. https://doi.org/10.1080/09546553.2020.1788544

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective Intelligence and Political Judgment*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/A/bo3636531.html

Marcus, G. E., Valentino, N. A., Vasilopoulos, P., & Foucault, M. (2019). Applying the Theory of Affective Intelligence to Support for Authoritarian Policies and Parties. *Political Psychology, 40*(S1), 109–139. https://doi.org/10.1111/pops.12571

Meiering, D., Dziri, A., Foroutan, N., Teune, S., Lehnert, E., & Abou Taam, M. (2018). *Brückennarrative: Verbindende Elemente in der Radikalisierung von Gruppen*. Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK).

Mudde, C. (2002). *The Ideology of the Extreme Right*. Manchester University Press. https://doi.org/10.7228/manchester/9780719057939.001.0001

Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In *Handbook of emotions* (pp. 511–536). The Guilford Press.

Pires, T., Schlinger, E., & Garrette, D. (2019). *How multilingual is Multilingual BERT?* (arXiv:1906.01502). arXiv. http://arxiv.org/abs/1906.01502

Pirro, A. L. P. (2023). Far right: The significance of an umbrella concept. *Nations and Nationalism, 29* (1), 101–112. https://doi.org/10.1111/nana.12860

Popper KR (2003) *Die offene Gesellschaft und ihre Feinde: Band 2; Falsche Propheten: Hegel, Marx und die Folgen.* J.C.B. Mohr (Paul Siebeck).

Romero-Rodríguez, L. M., Castillo-Abdul, B., & Cuesta-Valiño, P. (2023). The Process of The Transfer of Hate Speech to Demonization and Social Polarization. *Politics and Governance*, *11*(2). https://doi.org/10.17645/pag.v11i2.6663

Rothschild, Z. K., Landau, M. J., Sullivan, D., & Keefer, L. A. (2012). A dual-motive

model of scapegoating: Displacing blame to reduce guilt or increase control. *Journal of Personality and Social Psychology*, *102*(6), 1148–1163. https://doi.org/10.1037/a0027413

Rothut, S., Schulze, H., Hohner, J., & Rieger, D. (2023). Ambassadors of ideology: A conceptualization and computational investigation of far-right influencers, their networking structures, and communication practices. *New Media & Society*, 14614448231164409. https://doi.org/10.1177/14614448231164409

Rothut, S., Schulze, H., Hohner, J., Greipl, S., Rieger, D., & Döring, M. (2022). Radikalisierung im Internet: ein systematischer Überblick über Forschungsstand, Wirkungsebenen sowie Implikationen für Wissenschaft und Praxis. (CoRE-NRW Kurzgutachten, 5). *Bonn: Bonn International Centre for Conflict Studies (BICC) gGmbH.* https://nbn-resolving.org/urn:nbn:de:0168-ssoar-88147-1

Sabucedo, J.-M., Mar Durán, M. A., & Rodríguez, M.-S. (2011). Emotional responses and attitudes to the peace talks with ETA. *Revista Latinoamericana de Psicología*, *43*(2), 289–296.

Saha, P., Garimella, K., Kalyan, N. K., Pandey, S. K., Meher, P. M., Mathew, B., & Mukherjee, A. (2023). On the rise of fear speech in online social media. *Proceedings of the National Academy of Sciences*, *120*(11), e2212270120. https://doi.org/10.1073/pnas.2212270120

Saha, P., Mathew, B., Garimella, K., & Mukherjee, A. (2021). *"Short is the Road that Leads from Fear to Hate": Fear Speech in Indian WhatsApp Groups* (arXiv:2102.03870). arXiv. http://arxiv.org/abs/2102.03870

Sanh, V., Wolf, T., & Rush, A. (2020). Movement Pruning: Adaptive Sparsity by Fine-Tuning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (pp. 20378–20389). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/eae15aabaa768ae4a5993a8a4f4fa6e4-Paper.pdf

Sari, Ö. L. (2007). Perceptions of Threat and Expressions of Prejudice Toward the New Minorities of Western Europe. *Journal of International Migration and Integration / Revue de l'integration et de La Migration Internationale*, *8*(3), 289–306. https://doi.org/10.1007/s12134-007-0023-y

Sauer, B., Dietze, G., & Roth, J. (2020). Authoritarian right-wing populism as masculinist identity politics. The role of affects. Right-Wing Populism and Gender. *European Perspectives and Beyond,* 23–40.

Sayimer, I., & Derman, M. R. (2017). Syrian refugees as victims of fear and danger discourse in social media: A YouTube analysis. *Global Media Journal TR Edition*, *8*(15), 384–403.

Scheller, S. (2019). The Strategic Use of Fear Appeals in Political Communication. Political Communication, 36(4), 586–608. https://doi.org/10.1080/10584609.2019.1631918

Schmid, U. K. (2023). Humorous hate speech on social media: A mixed-methods investigation of users' perceptions and processing of hateful memes. *New Media & Society*, 14614448231198169. https://doi.org/10.1177/14614448231198169

Schulze, H., Greipl, S., Hohner, J., & Rieger, D. (2023). Zwischen Furcht und Feindseligkeit: Narrative Radikalisierungsangebote in Online-Gruppen. *MOTRA Monitor 2022*, 40-64.

Schulze, H., Hohner, J., Greipl, S., Girgnhuber, M., Desta, I., & Rieger, D. (2022). Far-right conspiracy groups on fringe platforms: A longitudinal analysis of radicalization dynamics on Telegram. Convergence, 13548565221104977. https://doi.org/10.1177/13548565221104977

Schulze, H. (2021). Zur Bedeutung von Dark Social & Deplatforming. Eine quantitative Exploration der deutschsprachigen Rechtsaußenszene auf Telegram. [Dark Social & Deplatforming. A quantitative exploration of the German-speaking far-right scene on Telegram.] Zeitschrift für Semiotik, 42 (3–4).

Stegmann, Y., Andreatta, M., & Wieser, M. J. (2023). The effect of inherently threatening contexts on visuocortical engagement to conditioned threat. Psychophysiology, 60(4), e14208. https://doi.org/10.1111/psyp.14208

Tausch, N., Bode, S., & Halperin, E. (in press). Emotions in violent extremism. In: M. Obaidi & J. Kunst (Eds.), Handbook of the Psychology of Violent Extremism, Cambridge University Press.

Tunstall, L., Werra, L. von, Wolf, T., & Géron, A. (2022). *Natural language processing with transformers: Building language applications with hugging face* (Revised edition). O'Reilly.

van der Brug, W., & Fennema, M. (2007). Forum: Causes of Voting for the Radical Right. *International Journal of Public Opinion Research - INT J PUBLIC OPIN RES*, *19*, 474–487. https://doi.org/10.1093/ijpor/edm031

Van Prooijen, J.-W., Spadaro, G., & Wang, H. (2022). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. *Current Opinion in Psychology*, *43*, 65–69. https://doi.org/10.1016/j.copsyc.2021.06.013

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Witte, K. (1996). Chapter 16 - Fear as motivator, fear as inhibitor: Using the extended parallel process model to explain fear appeal successes and failures. In P. A. Andersen & L. K. Guerrero (Eds.), *Handbook of Communication and Emotion* (pp. 423–450). Academic Press. https://doi.org/10.1016/B978-012057770-5/50018-7

Wright, L. (2019, May 7). *The Radicalizing Language of Fear and Threat*. Dangerous Speech Project. https://dangerousspeech.org/the-radicalizing-language-of-fear-and-threat/

Zehring, M., & Domahidi, E. (2023). German Corona Protest Mobilizers on Telegram

and Their Relations to the Far Right: A Network and Topic Analysis. *Social Media+ Society, 9*(1), 20563051231155106.

Zhang, Z., & Luo, L. (2019). Hate speech detection: A solved problem? The challenging case of long tail on Twitter. *Semantic Web*, *10*(5), 925–945. https://doi.org/10.3233/SW-180338

Ziolkowski, B., Lehmann, C., & Blum, F. (2022). *Fürchtet Euch!: Funktionen von Untergangsszenarien im extremistischen Kontext.* Landesamt für Verfassungsschutz Baden-Württemberg.

Appendix

The online Appendix can be found at https://osf.io/t3yx8/.

Supplemental Information

The trained classifier as well as the domain adaptation model can be found at: https://huggingface.co/PatrickSchwabl/distilbert_fearspeech_classifier.