

Codes, Patterns and Shapes of Contemporary Online Antisemitism and Conspiracy Narratives - an Annotation Guide and Labeled German-Language Dataset in the Context of COVID-19

Elisabeth Steffen¹, Helena Mihaljević¹, Milena Pustet¹, Nyco Bischoff¹, Maria do Mar Castro Varela², Yener Bayramoglu², Bahar Oghalai²

¹Hochschule für Technik und Wirtschaft Berlin

²Alice-Salomon-Hochschule Berlin

helena.mihaljevic@htw-berlin.de

Abstract

Over the course of the COVID-19 pandemic, existing conspiracy theories were refreshed and new ones were created, often interwoven with antisemitic narratives, stereotypes and codes. The sheer volume of antisemitic and conspiracy theory content on the Internet makes data-driven algorithmic approaches essential for anti-discrimination organizations and researchers alike. However, the manifestation and dissemination of these two interrelated phenomena is still quite under-researched in scholarly empirical research of large text corpora. Algorithmic approaches for the detection and classification of specific contents usually require labeled datasets, annotated based on conceptually sound guidelines. While there is a growing number of datasets for the more general phenomenon of hate speech, the development of corpora and annotation guidelines for antisemitic and conspiracy content is still in its infancy, especially for languages other than English. To address this gap, we have developed an annotation guide for antisemitic and conspiracy theory online content in the context of the COVID-19 pandemic that includes working definitions, e.g. of specific forms of antisemitism such as encoded and post-Holocaust antisemitism. We use the guide to annotate a German-language dataset consisting of ~3,700 Telegram messages sent between 03/2020 and 12/2021.

Introduction

Since its beginning, the COVID-19 pandemic is accompanied by an ‘infodemic’, in the course of which vast amounts of misinformation, hate speech, rumors and conspiracy theories are being spread, in particular through social and online media (Depoux et al. 2020; The Lancet Infectious Diseases 2020, 19). Due to their oftentimes racist or antisemitic content, many of the conspiracy theories contribute to an increase in discrimination and even violence against the targeted groups (RIAS e.V. 2020; Gover, Harper, and Langton 2020; Meisner and Wiedemann 2021).

The unprecedented role of digital technologies and social media in the spread of conspiracy theories and hate speech presents a key difference to previous pandemics. Respective narratives are being shared on video platforms, social networks, and messenger services, and with them racism, antisemitism, and calls for violence, which sometimes translate

into violent attacks in the real world.¹

The sheer volume and the rapidly evolving online dissemination of antisemitism and conspiracy theory content make data-driven algorithmic approaches indispensable (Marcellino et al. 2021). Current publicly available services for automated detection of related phenomena such as toxic language, however, do not adequately cover antisemitism, in particular when it is communicated using codes and metaphors (Mihaljević and Steffen 2022). In order to improve underlying machine learning models, comprehensive labeled data from online and social media are required.

Existing datasets related to conspiracy theories or antisemitism are typically generated by filtering texts using explicit keywords such as ‘5G’, ‘Bill Gates’ or ‘jew*’. However, such approaches introduce a keyword bias to the generated corpora that makes it difficult to detect lesser-known or new conspiracy narratives or to identify intentionally obfuscated or coded terms, the latter being increasingly used e.g. to evade regulation by platform operators. Especially antisemitic content is often conveyed in an encoded way, using metaphors and codes that work without explicit reference to Jews or Israel (Zannettou et al. 2020; Becker, Troschke, and Allington 2021). In addition, newer forms of antisemitism are on the rise (Schmalenberger and Hübscher 2022; Schneider 2020) which are not sufficiently covered by standard working definitions, and difficult to discover through commonly used keywords. This can also be observed in the context of the COVID-19 pandemic, when some anti-restriction protesters compare themselves with victims of the Shoah or equate the mandatory use of face masks with the obligation for Jewish citizens to wear the ‘Yellow Star’ in Nazi-Germany (Schneider 2020).

In this paper, we draw on extensive research to develop adequate working definitions and an annotation guide for antisemitism and conspiracy theories in online content in the context of the COVID-19 pandemic. Regarding antisemitism, we focus on encoded forms of antisemitism and post-Holocaust antisemitism. We develop our annotation scheme as an interdisciplinary team to ensure a compre-

¹The assaults in Christchurch, Halle, or Hanau were impelled by racist and antisemitic conspiracy narratives disseminated via different online platforms, and the killers used online platforms to stage their killing in live streams (Musyal and Stegemann 2020)

hensive conceptual approach. We provide real-world examples with our working definition to allow for its further development and its adaptation to other cultural, historical, and linguistic contexts and additional data sources. Furthermore, we use our working definition to annotate a German-language dataset *TelCovACT* consisting of 3,663 Telegram messages posted between March 11, 2020 and December 19, 2021 and thus promote research in a less studied language. We chose Telegram because of its popularity among opponents of the government’s measures to combat the coronavirus and the frequent spread of conspiracy theories and antisemitic statements (Comerford et al. 2021; Winter et al. 2021). The dataset is made available to foster further research, especially on automated detection of antisemitic and conspiracy-theory content.

Readers should be aware that this paper includes instances of explicit antisemitism, racism, and other forms of hateful language.

Related Work

Our literature review focuses on studies in which large amounts of data have been collected and analyzed, typically in conjunction with annotation efforts, in order to provide an overview of existing datasets and associated definitions and categorization schemes.

Conspiracy Theories in Social Media Some recent works provide openly accessible annotated datasets and use them, often as part of challenges, to develop models for automated classification (Alam et al. 2021; Golbeck et al. 2018). Here, conspiracy theories tend to be considered rather a subcategory and are often used synonymously with rumors or misinformation (Serrano, Papakyriakopoulos, and Hegelich 2020), so that existing datasets and annotation schemes are not based on a common theoretical foundation. According to a systematic literature review, only around a third of considered works provided a definition, thus analyzing “conspiracy theories online without explicitly defining the main object of their research” (Mahl, Schäfer, and Zeng 2022).

A frequently applied approach is to address specific known conspiracy theories and gather data by searching for selected keywords (Marcellino et al. 2021; Memon and Carley 2020; Moffitt, King, and Carley 2021; Serrano, Papakyriakopoulos, and Hegelich 2020) without discussing the labeling process in detail (Gerts et al. 2021; Pogorelov et al. 2020) or by resorting to examples in order to provide a definition (Moffitt, King, and Carley 2021). In most cases, a few thousand records are labeled manually. Marcellino et al. (2021) use a larger basis of 150,000 texts by refining the combination of search terms that refer to four well-known conspiracy theories, but the process is not fully clear. Classification models trained on such keyword-based datasets yield moderate to high accuracy and typically employ language models such as BERT. Part of the research makes their datasets and codebooks openly available.

Other works provide a solid theoretical foundation of the subject, discussing the relations among concepts such as conspiracy theories, rumors or misinformation (Kou et al. 2017; Samory and Mitra 2018; Wood 2018). While the ap-

plied definitions differ, they share a number of conceptual elements, in particular the assumption of a “secret plot between powerful people or organizations” (Mahl, Schäfer, and Zeng 2022) that work deliberately for their own sake and against the common good (Uscinski et al. 2020). An extensive literature review by Samory and Mitra (2018) reveals that the majority of relevant research “relies on agents, actions, and goals as key elements in defining conspiracy theories or conspiracy beliefs,” making paradigmatic examples dispensable. Kou et al. (2017) provide an operational definition of a conspiracy theory about public health crises containing the following three criteria: 1) the theory includes an explanation of the causality behind an event, 2) the explanation refers to primary actors (individuals or organizations, ‘the Other’) whose actions are being kept secret from the public, 3) the actions have a malicious purpose, harming the greater good in favor of the actor’s own agenda.

Conspiracy Theories and Antisemitism We could not identify any publicly available datasets connecting conspiracy theories and antisemitism, and it seems that antisemitism has only recently attracted the attention of research on conspiracy theories in social media: Mahl, Schäfer, and Zeng (2022) found that only 2.1% of recent empirical studies addressing single conspiracy narratives focus on antisemitic narratives.

However, antisemitic stereotypes play an important role in COVID-19 related conspiracy theories: An alarming prevalence of antisemitism, for example in the context of deep state conspiracy narratives, has been found on Twitter, together with a prevailing Nazi-Germany rhetoric in numerous German tweets debating coronavirus health measures (Karakoulaki and Dessì 2021).

The pandemic has led to new antisemitic conspiracy theories (Cohen, Kaati, and Pelzer 2021), while recycling old stereotypes. The narrative of “Jews ruling international financial, political and media institutions” is identified as most dominant antisemitic conspiracy theory element across different European countries and social media platforms (Comerford et al. 2021, 9).

A recent comprehensive report examined the links between COVID-19 anti-vaccination conspiracy theories and antisemitism in Twitter and Facebook in seven European countries between March and August 2021 (Karakoulaki and Dessì 2021). Their key findings include that 1) antivaxxers typically perceive themselves as victims and resort to Holocaust comparisons and the self-labeling as ‘the new Jews’; 2) references to and variations of established antisemitic conspiracy theories such as ‘The Great Reset’ and ‘New World Order’ play a significant role; and 3) antisemitic codes such as ‘globalists’ are frequently used throughout Europe. This showcases why an awareness of antisemitic codes, the structure of antisemitic argumentations, and a consideration of specific forms such as post-Holocaust antisemitism is relevant for a classification of antisemitic conspiracy content.

Antisemitic Online Content Almost all scientific studies known to us addressing the large-scale annotation of antisemitic content in text data rely on the working defini-

tion by the International Holocaust Remembrance Alliance (IHRA) as the main basis for their coding schemes (Becker and Allington 2021; Chandra et al. 2021; Guhl, Ebner, and Rau 2020; Jikeli, Cavar, and Miehlung 2019; Jikeli et al. 2022; Schwarz-Friesel 2019).² As shown by Jikeli, Cavar, and Miehlung (2019), using an English-language corpus containing the word (stems) ‘Jew*’ or ‘Israel’, the IHRA definition is well suited in such a setting to generate a gold standard corpus for antisemitic content. However, many parts of the IHRA definition need further elaboration and refinement in order to serve as an annotation basis for automatic detection systems. Accordingly, Jikeli et al. (2022) extend their previous work and build an annotation scheme with many examples based on a close reading of the definition. Similarly, the project ‘Decoding Antisemitism’ states to use the IHRA definition as a conceptual framework but extend it with further categories (Becker, Troschke, and Allington 2021) in order to annotate comments of major media outlets.

Among the few studies addressing algorithmic detection of antisemitic online content, we identified the paper by Warner and Hirschberg (2012) as the earliest work. The authors used a template-based strategy to extract features from text and then trained an SVM, obtaining an F1 score of ~ 0.6 on a custom dataset consisting of 9,000 paragraphs. Guhl, Ebner, and Rau (2020) used a commercial software to train a classifier for online content from the imageboard 4chan. The model achieved an F1-score of ~ 0.76 on a dataset containing 2,907 posts. Ozalp et al. (2020) trained a supervised machine learning model using 853 manually annotated tweets to detect online “antagonistic content related to Jewish identity” in tweets containing certain keywords by UK-based users. Smedt (2021) created a machine learning-based system for scoring English and German language texts regarding the level of antisemitic toxicity. The model is based on a self-developed lexicon consisting of over 2,000 relevant words and phrases containing Nazi-Germany rhetoric, dehumanizing adjectives and violence-inciting verbs, far-right terminology, alt-right neologisms, coded language, and revived conspiracy theories. In Chandra et al. (2021), a multimodal deep learning classification model is trained on text and images, with an F1-score of 0.71 for Twitter and 0.9 on Gab.

Existing works in this context including all previously mentioned, use corpora created with keyword filters, which often relate to Jewishness or Judaism (e.g. ‘jew*’, ‘hebrew’) (Chandra et al. 2021; Jikeli, Cavar, and Miehlung 2019; Jikeli et al. 2022; Ozalp et al. 2020), the state Israel, or reflect known and sometimes platform-specific antisemitic slurs (e.g. ‘kike’, ‘ZioNazi’, ‘(((jew)))’) (Zannettou et al. 2020), or are associated with antisemitic stereotypes and narratives (e.g. ‘happy merchant’, ‘6 million’) (Guhl, Ebner, and Rau 2020). The limitations resulting from this approach are partly reflected by the respective authors: Ozalp et al. (2020), for instance, underline that “much antisemitic hate speech comes in the form of conspiracy theories (or allu-

sions to such theories) and image-based hate speech—such as memes—that would not be captured by these keywords”. Jikeli et al. (2022) justify the keyword-based approach with the need to obtain a sufficient number of positive examples in the annotated dataset.

While most of existing research focusses on English-language content, German-language texts are covered by Monika-Schwarz Friesel’s large-scale empirical study on antisemitic online content based on a large variety of text corpora, and the project ‘Decoding Antisemitism’ that analyzes user comments on the webpages of mainstream media outlets in English, German and French (Becker, Troschke, and Allington 2021). Except for Chandra et al. (2021); Jikeli et al. (2022), the manually annotated datasets are not announced as publicly available.

Elaboration of Adequate Working Definitions

Antisemitism and conspiracy theories are inherently complex phenomena that can be difficult to annotate, especially when expressed as short texts in messenger services or social media platforms (Ozalp et al. 2020). Thus, a careful elaboration of underlying concepts is necessary for a reliable annotation process (Roß et al. 2016). In this section we present our working definitions, which represent the conceptual basis of our annotation scheme.

Antisemitism

The basis for our working definition of antisemitism is the working definition of the International Holocaust Remembrance Alliance (IHRA):

Antisemitism is a certain perception of Jews, which may be expressed as hatred toward Jews. Rhetorical and physical manifestations of antisemitism are directed toward Jewish or non-Jewish individuals and/or their property, toward Jewish community institutions and religious facilities. (International Holocaust Remembrance Alliance 2016)

This definition has been recognized and implemented by numerous countries, cities, governmental and non-governmental institutions in various political and social fields. It has also shown to be a viable ground for the annotation of corpora containing explicit references to Jews, Israel or antisemitic stereotypes (Jikeli, Cavar, and Miehlung 2019; Jikeli et al. 2022).

As others have pointed out, a strength of the definition is that it covers most contemporary manifestations of antisemitism offering descriptive examples. However, it needs to be interpreted in context, and its examples need to be concretized in order to use it for annotation (Jikeli et al. 2022). For example, the definition does not explicitly address equations of current pandemic containment measures by democratic governments with Nazi crimes against Jews. Likewise, antisemitic narratives addressing non-Jews are only briefly mentioned but not further elaborated (Jikeli, Cavar, and Miehlung 2019). Conspiracy theories regarding vaccinations are a particularly vivid example that combine these manifestations of antisemitism (Karakoulaki and Dessì 2021). We

²Chandra et al. (2021) combine the IHRA definition with Brustein’s categorization of antisemitism into political, economic, religious, and racial antisemitism.

thus propose the following extensions and concretizations of the presented definition:

1. Post-Holocaust antisemitism (PHA) that refers to the instrumentalization of the victims of the Holocaust for a political agenda.
2. Linguistic encodings of antisemitic statements that do not mention Jews or the State of Israel.

Our conceptualization aims at being sufficiently context-sensitive for contemporary manifestations of antisemitic conspiracy narratives by being close enough to empirical data, and abstract and generic enough to be adaptable to future research.

Post-Holocaust Antisemitism The term post-Holocaust antisemitism (PHA) was coined by Marin (1980) to describe “antisemitism without antisemites”. Corresponding narratives explicitly name Jews as part of argumentation strategies which instrumentalize the victims of the Holocaust for a political agenda and at the same time shift the perpetrator-victim coordinates by undertaking relativizing Holocaust comparisons. According to Salzborn (2021), the instrumentalization is an essential component of this form of antisemitism. It fulfills a dual function: With regard to the past, it historically relativizes the Holocaust and infamously instrumentalizes the antisemitic policy of extermination. With regard to the present, it allows conspiracy theorists to demonize democratically legitimized rulings and measures by describing themselves as victims of a dictatorial state.

In times of the COVID-19 pandemic, we encounter forms of PHA in comparisons or equations of the state measures to combat the pandemic with the National Socialist persecution of Jews. In Germany, some participants of demonstrations against restriction policies wear a yellow star with the imprint ‘ungeimpft’ (unvaccinated) and thus symbolically compare themselves with Jews who under the National Socialist regime were forced to wear one. Leaflets and placards reading ‘Impfung macht frei’ (vaccination sets you free), a reference to the slogan ‘Arbeit macht frei’ (work sets you free) at the entrance to Auschwitz and other Nazi concentration camps, were distributed or shown at protests (Belghaus, am Orde, and Jakob 2020).

Encoded Antisemitism While the anonymity of online platforms on the one hand presents a fertile ground for explicit antisemitic hate speech, antisemitism is also often expressed via encoded, implicit manifestations. Findings from ongoing research indicate that “users use a variety of coded forms to communicate their antisemitic attributions” (Becker, Troschke, and Allington 2021, 7), including semiotic markers such as icons or emoticons, abbreviations, word plays, allusions, and metaphors (Becker, Troschke, and Allington 2021, 7). This also applies to manifestations of antisemitism in the context of COVID-19 conspiracy narratives: While explicitly antisemitic conspiracy narratives frame the pandemic as a smokescreen used by Zionists, the Rothschild family or George Soros to expand their power, others are more implicit and make references to individuals like Bill Gates, the ‘New World Order’ or generally ‘the’ (economic or political) elites (RIAS e.V. 2021;

Finkelstein et al. 2020; Comerford et al. 2021), thus operating encoded and completely without explicitly mentioning Jews.

Expressing antisemitism in encoded, implicit ways allows users to avoid social ostracism, the deletion of content from social media platforms, or even criminal consequences. Despite political and ideological differences, antisemitic discourses show a great uniformity and homogeneity regarding the stereotypes and codes used in them, highlighting their relevance for the transmission of antisemitism (Schwarz-Friesel 2019). Encoded manifestations of antisemitism can thus be assumed to play an important role for the online dissemination of antisemitism. Hence, it is of central importance to include them in respective annotation guidelines.

Conspiracy Theories

The term conspiracy theory was first coined by Popper, who argued that social sciences should not fall into the trap of providing simple explanations for unintended events, which he termed as ‘conspiracy theory of society’ (Popper and Kiesewetter 2003, 306). According to Popper, unlike scientific explanations, conspiracy theories provide simple answers for complex social and political events. Some conspiracy theories even refer to scientific studies as well as academic experts to support their arguments. At the same time, a defining feature of conspiracy theories is their “self-sealing quality”, meaning that they “are extremely resistant to correction, certainly through direct denials of counterspeech by government officials.” (Sunstein and Vermeule 2009).

Working Definition The concept of conspiracy theories is often used synonymously with similar forms of deceptive content such as disinformation (intentional dissemination of incorrect information) or misinformation (unintentional dissemination of incorrect information), rumors (unverified information), or fake news (fabricated news or a label used for delegitimizing news media) (Mahl, Schäfer, and Zeng 2022).

While conspiracy theories partially overlap with these concepts (e.g. a conspiracy theory might contain misinformation), they do have their own unique characteristics as attempts to create an alternative interpretation of events (Mahl, Schäfer, and Zeng 2022): Conspiracy theories formulate the strong belief that a secret group of people, who have the evil goal of taking over an institution, a country or even the entire world, intentionally cause complex, and in most cases unsolved, events and phenomena (Butter 2018). The exact intention (of a power-takeover) does not always have to be explicitly articulated; what is important, however, is the existence of a harmful intention and that the respective goal is of significant relevance to the public. A conspiracy theory can thus be considered an effort to explain some event or practice by reference to the machinations of powerful people, who have managed to conceal their role (Sunstein and Vermeule 2009). It needs actors (e.g. corrupt elites) who supposedly pursue a concrete malicious goal (e.g. control the population) using a strategy (e.g. by inserting a microchip via vaccinations) (see also Samory and Mitra (2018)).

Such narratives are based on a simple dualism between

good and evil, and leave no space for the unintentional or unforeseeable. The nature of social media and messenger services entails that more complex narratives are often incompletely rendered, especially when the counterpart can be assumed to be (partially) knowledgeable (Sadler 2021; Ernst et al. 2017). Accordingly, we believe that it is useful to annotate which of the components (actors, goal, strategy) actually appear in a given text. This will allow for post-annotation categorization of conspiracy theories in terms of completeness and fragmentation.

Conspiracy Theories and the COVID-19 Pandemic

Studies show that particularly times of crises such as pandemics are prone to the emergence and spread of conspiracy theories (Heller 2015; Kitta 2012; Starbird et al. 2014). Since the threat posed by a disease is not directly tangible, pandemics often foster a range of conspiracy theories (Hepfer 2015). As an effect, the identified ‘culprits’ can be named concretely and become tangible, which seemingly helps to structure an overwhelming situation. As was the case with other pandemics, the interpretations circulating on social networks lead to fatal mis- and disinformation about the origin and routes of infection or measures against the COVID-19 disease (Smallman 2015). A growing body of literature has observed how the outbreak of the pandemic not only led to a circulation of conspiracy theories but also how such theories led to the catalyzation or emergence of transnational movements such as QAnon and the so called Querdenken movement (Bodner, Welch, and Brodie 2020).

Compatibility with Antisemitism There is a high degree of compatibility between antisemitism and conspiracy theories that is largely due to the strong structural ties between these two phenomena. Haury (2019, 49-50) elaborates the following fundamental principles characterizing modern antisemitism which are also central for our understanding of conspiracy theories:

- A specific form of personification, which attributes subjectless societal processes to the conscious, intentional, and malevolent actions of individuals; this inevitably induces the construction of an omnipotent enemy who has secretly taken over crucial points of control.
- A Manichean worldview which radically divides the world into good and evil, a dualism based on an ontological construction of group identities. The enemies are represented as a foreign, external community with an immutable ‘nature’ and characteristics. The usually nationally or ethnically constructed in-group is imagined as inherently good, naturally rooted, and free of internal conflicts or contradictions.
- The enemy group is imagined as a corrosive and subversive threat for the in-group, potentially destroying its identity as well as its societal and political structures. Expulsion and extermination of the enemy group are seen as not only legitimate measures but as a last resort in face of the omnipotent, conspiratorial enemies who allegedly aim at destroying the collective.

Conspiracy theories are thus an ideal medium for the dissemination of antisemitic tropes, images and narratives.

Looking at the narratives of protesters against COVID-19 countermeasures in Germany that reference conspiracy theories, it is possible to identify facets of all the outlined characteristics Lelle and Balsam (2020): Conspiracy theorists accusing individuals like Bill Gates, the (meanwhile former) chancellor Angela Merkel or virologist Christian Drosten of having a stake in and making profit out of the pandemic are examples of personification. Furthermore, the named persons are often accused of being part of a global conspiracy. The protesters perceive themselves as an ‘awakened’ group which is fighting evil and spreading truth to disclose the lies of the global, malignant group of conspirators. Even though it must be noted that not all speakers at the demonstrations promote the idea of a nationally or ethnically defined community, most of them emphasize the notion of community, unity and ‘naturalness’. As a consequence they contribute to a homogenization of the group of protestors. Furthermore, the frequent participation of Nazis and so-called ‘Reichsbürger’ at the protests contributes to the spread of ethnic and nationalist ideologies in the movement.

Finally, the increasing aggression, violent fantasies on posters and in chat groups, as well as the first violent attacks attributed to the spectrum of this movement point to its partial radicalization (CeMAS – Center für Monitoring, Analyse und Strategie 2021). In some cases, this process is accompanied by a shift from structurally antisemitic attitudes to explicit and violent hatred of Jews (Rose 2021).

Corpus and Annotation Scheme

Our dataset was created based on two primary data sources: 1) A set of public Telegram channels identified as particularly relevant for the mobilization against COVID-19 measures in Germany through a network analysis in the early phase of the pandemic (Forschungsinstitut Gesellschaftlicher Zusammenhalt 2020), of which we selected those with more than 1,000 followers. 2) Telegram channels mentioned in a Twitter dataset containing selected keywords related to ‘Querdenken’ from three time periods around key demonstrations in 2020 and 2021. All channels were manually validated and ranked for relevance based on a random sample of 100 messages per channel. Thereby, a channel was considered relevant if it contained a high amount of conspiracy theories and/or antisemitic narratives with reference to the COVID-19 pandemic. From the initial 215 channels, 133 were considered as particularly relevant. We drew a random sample of messages from each of the pre-selected channels. The number of samples per channel ranged from 50 to 300, depending on the channel’s size, resulting in a subset of 28,000 messages. To increase the diversity among instances, highly similar messages were identified using a normalized Levenshtein distance and subsequently removed from the dataset. We also removed rather short texts. After this step, another random sample was drawn and annotated, resulting in the final dataset *TelCovACT* containing 3,663 messages from public German-language Telegram channels in the time period between March 11, 2020 and December 19, 2021. Further details can be found in the datasheet documenting the dataset (Bischoff, Pustet, and Mihaljević 2022).

Annotation Scheme

Our annotation scheme is based on the working definitions of antisemitism and conspiracy theories presented above, and extended by sub-labels to specify the stance and content of a given instance. The stance-related sublabels for conspiracy theory content are based on an adapted form of the Rumor Interaction Analysis System (RIAS) scheme (Wood 2018)). Table 1 shows a condensed outline of our final annotation scheme. The detailed annotation guide is available at (Steffen and Mihaljević 2022).

Final Dataset

Approximately 14% of the messages are labeled as antisemitic and 36% as communicating conspiracy theories. At least one conspiracy-theory (antisemitism) related message was identified in 101 (85) of 133 channels. For almost all texts containing antisemitic content, the stance was classified as affirmative (94%). Almost 60% were labeled as encoded³, making it the most frequent sub-form of antisemitism in the corpus. For conspiracy theories, belief was the most frequent stance (95%), followed by ‘authenticating’ (24%). The narratives most often included a strategy (72%) and an actor (64%). It should be noted that conspiracy theories were mostly communicated in a fragmented way, with only 26% containing all of actor, strategy and goal, while 13% communicated the respective content using a reference only, such as ‘#QAnon’.

More than 72% of all texts labeled as antisemitic also contained conspiracy theory content, while the majority of conspiracy theory messages (71%) were not labeled as antisemitic. This does not mean that the majority of the conspiracy theories were not antisemitic; our annotation scheme requires more than a mention of an antisemitic conspiracy theory to be labeled as such. Considering the level of fragmentation in the communication of conspiracy theoretical content, with 13% communicating via a reference only, it is not surprising that antisemitism is expressed in a minority of conspiracy theory messages. It is worth noting that the distribution of sub-forms of antisemitism differs significantly (chi-squared test with p-value < 0.01) depending on the presence of conspiracy theory content: In the group of texts communicating both, the PHA label is given to less than 9%, followed by ‘other’ with 27%, and encoded antisemitism being the leading label with a frequency of 72%. However, if no conspiracy content was present, other forms of antisemitism are found most frequently (56%), followed by PHA in 25% of all cases, and only 23% labeled as encoded.

We computed chi-square test scores in order to find words that are most significant for the two categories. As expected, in texts communicating conspiracy theories we find in particular references to well-known theories such as New World Order, the Great Reset, deep state or ‘plandemie’ (referring to a planned pandemic), actors such as Bill Gates, freemasons, Soros or Clinton, and words indicating strategies or goals such as lie, execute or dictatorship. The most significant words that are positively correlated with texts commu-

³Note that multiple labels were possible.

nicating antisemitism include references to Jewish identity such as jew or jewish as well as frequent codes such as illuminati, Soros, Rothschild, freemason or satanic.

Figure 1 presents a comparison of the 100 most frequent verbs and nouns in texts labeled as expressing conspiracy theories versus those that were not. These present first insights into the actors, strategies and goals that are associated with conspiracy theories in TelCovACT. The wordclouds in the left-hand column (positive label) show especially verbs such as ‘versklaven’ (enslave), ‘hinrichten’ (put someone to death) or ‘unterwandern’ (infiltrate) that are associated with the imagination of a malicious secret takeover, while nouns such as ‘Weltordnung’ (world order) or ‘Diktatur’ (dictatorship) indicate the supposedly pursued malicious goals by actors such as ‘Illuminat’ or ‘Hochfinanz’ (high finance).

Data Ethics and Privacy

Our approach to handle data ethics, privacy and protection follows best practices as documented in (Rivers and Lewis 2014). This includes exclusively collecting publicly available data and preventing data from being used to identify authors: Even though Telegram’s Terms of Service states that user names and ids cannot be linked to a user’s phone number as the only personal data collected by Telegram, we chose to additionally anonymize the dataset by replacing user names, user-ids as well as links to such by USER. Furthermore we decided only to provide our annotated dataset on personal request for research purposes approved for our ethical standards, thereby preventing any attempt of abuse. We also note that the annotator team comprised nine individuals with diverse socio-demographic backgrounds, working in various disciplines (five in political science or sociology and four in data science) with different levels of academic training.

Annotation Process and Evaluation

The annotators reflected on their annotation experiences and discussed examples of conflicting annotations in a workshop to gain insights into factors which had affected their annotation decisions. During the workshop, almost all discussed conflicts could be resolved. After the joint workshop, 445⁴ texts were labeled by two annotators per message and used to compute the inter-annotator reliability. Our results indicate solid agreement among annotators, with Cohen’s kappa being 0.7 for conspiracy theory and 0.84 for antisemitism. Based on the workshop insights, two annotators resolved annotation disagreements for the final dataset in a common session. All instances in the dataset are annotated by two annotators.

The main insights from the workshop (and resulting modifications to the initial annotation scheme) are the following:

Positively Biased Corpus The annotated dataset was generated exclusively from channels known to spread conspiracy theories and antisemitic content, and all annotators were

⁴Of the 500 texts originally selected at random, 55 were marked as unsuitable and thus excluded from the evaluation.

Antisemitism

Post-Holocaust	The MASK now becomes the Yellow Star of the unvaccinated! A year ago it was a #conspiracy theory that #unvaccinated are marked separately. Today #Lindner demands a #mask obligation for all who are not #vaccinated. Is the #mask really becoming the new #Jewish star?
Encoded	The long-term plan cooked up by the satanic zionists to kill billions of people is blowing up after the failure of their riots and fake pandemic.
Other	What has the Jew done to us? All “vaccines” are gene poison injections and come from Jewish corporations!
<i>Stance labels: Affirmative; critical; neutral/uncertain</i>	
Conspiracy theory	
Actor	System failure - who trusts the globalists? Get connected with each other. From the 2nd lockdown the way to forced vaccination and expropriation it is not far.
Strategy	The sanity of some people can only be doubted. In Germany, no one is treated in a hospital without undergoing a compulsory test! I call it health dictatorship and omitted assistance, because our medical staff knows exactly what is going on in hospitals. No dialysis without compulsory test ! Maybe the test is the true vaccination !!?
Goal	Here is a small and incomplete list of publications that more or less openly admit that they want to eradicate and/or replace the white race. Your normie friends will still call you “conspiracy theorists” if you bring it up.
Reference	QAnons for Berlin The whole World is with us. Nothing can stop what is coming. Nothing. Make Earth Great Again. WWG1WGA
<i>Stance labels: Belief; authenticating; directive; rhetorical question; disbelief; neutral/uncertain</i>	

Table 1: A condensed outline of the final annotation scheme for antisemitic and conspiracy theory narratives in online communication, with English translations of examples from the corpus for reader comprehension.

aware of this ‘positive bias’. This contextual knowledge influenced how some interpreted a message. Furthermore, the dataset mainly consisted of messages affirming antisemitism or conspiracy theories, making the selection of stance labels appear rather obsolete. However, for more heterogeneous data sources, the differentiation by stance is generally a useful additional information that can be utilized for training classifiers (Marcellino et al. 2021).

Antisemitism Annotators expressed discomfort with classifying a text as ‘no antisemitism’, a label that was provided in our initial version of the annotation scheme, arguing that this could be interpreted as confirming a given text to be antisemitism-free. Furthermore they explained that sometimes a text itself could not be classified as antisemitic following the provided definition, but nonetheless would contain certain antisemitic undertones. For the final annotation scheme, we thus removed the choices ‘no’ and ‘uncertain’ for both main categories and instead added a label ‘review required’ for cases requiring exchange with others.

It was discussed how to classify texts that do not match our definition of antisemitism but include references to antisemitic conspiracy theories such as QAnon. We had only provided a ‘reference’ label for the conspiracy theory category, but not for antisemitism. Annotators discussed whether these messages should be classified as ‘encoded antisemitism’; however, it was argued that the appearance of a single term or code is not sufficient, even when it is often used as an antisemitic code, making it difficult to introduce such a label for antisemitism.

Annotators perceived the sub-label ‘other’ as potentially trivializing because it invokes the impression of being used for ‘secondary’ forms of antisemitism, while being too coarse-grained as it subsumes a variety of antisemitic con-

tent. Since our aim was to focus on post-Holocaust antisemitism and encoded antisemitism as previously unexposed manifestations of antisemitism in existing annotation tasks, we nonetheless consider the use of the sub-label ‘other’ as adequate. Depending on the research focus, however, a differentiation of it should be considered, e.g. using existing annotation schemes as in Jikeli et al. (2022).

Other discussions evolved around messages mentioning Israeli politics. While the content itself mostly could not be classified as antisemitic, some annotators saw the mere focus on Israel in the context of our channel selection as a clear indication of antisemitic bias. While such consideration of focus and agenda-setting is common for qualitative approaches like critical discourse analysis, we doubt that it can be transferred to the training of classifiers that typically work on single-message level and have no knowledge of the ‘overall tendency’ of a channel (in fact, finding channels communicating a certain amount of respective content is a plausible application scenario of a classifier). Another example for such a controversy around the consideration of context was a message which described COVID-19 prevention measures as systematic discrimination and fascist. For some annotators, the use of the term ‘Faschismus’ in a German-language COVID 19-context indicated a clear relativization of the Shoah. They argued that in a German context, the term fascism is widely used synonymously with the German national socialist regime, and thus interpreted this message as a manifestation of post-Holocaust antisemitism. Other annotators doubted this interpretation, arguing that the term fascism potentially describes different kinds of phenomena.

Conspiracy Theories The provided definition with its division into the elements actor, strategy, and goal was overall perceived as helpful and comparatively easy to apply. At the



Figure 1: The wordclouds show the 100 most frequent verbs (first row) and nouns (second row) in texts from TelCovACT that were labeled as expressing conspiracy theories (left-hand column) versus those that were not (right-hand column).

same time, annotators stated they sometimes found it difficult to clearly separate strategy and goal. Furthermore, if they could not identify a goal, they were more hesitant to label a text as conspiracy theory. Moreover, several messages were observed in which actors were only implicitly mentioned, e.g. as ‘they’ or ‘our enemies’, which is why we consider it an important feature of our annotation scheme to include implicit mentions.

Some texts were classified as conspiracy theory even though neither actor nor strategy nor a goal could be identified. This applies for example to texts describing the great majority of society ignorant of the conspiracy, e.g. by referring to them as ‘Schlafschafe’ (sheeples), or calling for an awakening of the masses. It was suggested to include the element ‘target’ or ‘victim’ of a conspiracy to our definition to include these kinds of texts. Other examples missing the defined triple were texts suggesting that ‘the truth’ was generally disguised. It was argued that both types of messages should be labeled as conspiracy theory.

Additionally, some annotators decided to apply the sub-label ‘reference’ if they interpreted a message as conspiracy theory but perceived it as too implicit and fragmented to apply the actual definition. In various cases, it became evident that background knowledge had influenced the decision, for example if a message contained links to platforms known to be disseminating conspiracy theory content.

A lot of texts turned out to contain fake news, dis- or misinformation. For the sake of feasibility, however, we had deliberately decided against providing respective labels, since this would require thorough fact checking, in some cases even scientific expert knowledge. Nevertheless, such text

fragments were perceived as important discursive elements of conspiracy theories by some annotators who chose to label them as ‘reference’ as a workaround. These decisions contributed to a partly inconsistent application of this sub-label.

Limitations and Potential Applications of the Dataset

Our aim was not to generate a dataset representative of the German-language anti-COVID-19 movement, which we consider as highly challenging given its informal and dynamic nature. Instead, we focused on open Telegram channels with a significant amount of antisemitism and conspiracy theories to create a diverse dataset that is not strongly biased towards a small set of explicit keywords.

The dataset can be used for classification tasks. It is imbalanced, with a significantly higher frequency of negative labels (reflecting the reality), a common issue in similar datasets created for e.g. the automated detection of hate speech that requires adjustments during model training and evaluation. Additionally, the corpus size is relatively small compared to widely used English-language corpora for hate speech detection, but similar to other annotated German-language datasets with a more specific focus. Initial experiments training classifiers based on SOTA approaches to detect conspiracy theories have yielded satisfactory results, similar to existing literature. For antisemitism, we suggest to combine the TelCovACT corpus with other existing data sources, since the percentage of positive labels in this category is considerably smaller. In general, we suggest to use our corpus as one subset to combine with other sources into

larger and more heterogeneous training corpora.

Moreover, we recommend to use TelCovACT to validate existing approaches to detecting conspiracy theories, antisemitic speech, as well as related phenomena such as offensive and hateful language. As an example of this, we used the dataset to explore the potential and limitations of the toxicity scorer provided by the Perspective API for detecting antisemitic language, showing its weaknesses with regard to encoded forms of antisemitism as well as a false positive bias towards texts containing words such as Jew or Israel (Mihaljević and Steffen 2022).

The corpus has also been requested as source for qualitative research, and we encourage other researchers to apply it for this cause.

The dataset has a topic bias towards the COVID-19 pandemic, with approximately half of the texts explicitly related to it. Additionally, it is limited to a specific time period and does not capture earlier or later stages of German-language conspiracy narratives and antisemitic language on Telegram channels. Furthermore, the dataset is biased towards the sources from which we selected the channels. To address this, we recommend that future data collection efforts include Telegram channels that are mentioned on other platforms such as Facebook, YouTube, and Reddit, to provide a more comprehensive picture of the range and types of narratives present within the German-language anti-COVID-19 movement.

Discussion and Future Work

Due to time constraints, we applied labels to the entire message. In future work, it would be valuable to explore the possibility of annotating specific segments of a text. Such an approach could provide valuable insights into the use of certain antisemitic codes and how they relate to other parts of the message. For example, it could help to identify patterns in the ways certain codes are used and how they interact with other elements of the text. This, in turn, could inform more nuanced classification approaches and contribute to a deeper understanding of the dynamics of antisemitic and conspiracy discourse.

The fraction of post-Holocaust antisemitism (PHA) in our dataset was lower than expected from qualitative analyses of Twitter data and public protests. We assume that the lack of regulation and content moderation on Telegram allows to uncover more directly one's antisemitic views, while PHA occurs more frequently in regulated contexts.

The high proportion of encoded antisemitism, in particular in connection with conspiracy theory content, confirms that the antisemitic codes of a 'global elite' in control of global political and economic processes can easily be adapted for the expression of belief in conspiracy theories. The large association of the two phenomena underlines the importance of approaching antisemitism in large online corpora without the restriction to keywords explicitly referring to Jews, Jewishness or Jewish collectives or institutions.

It is also worth noting that classification models trained on a dataset like this with high overall toxicity are more likely to actually learn aspects specific to antisemitism or conspiracy

theory and not to related but differing phenomena such as offensive language or hate speech.

With respect to the annotation process, we found that a major factor for different annotators' assessments was the handling of the context of a message or the consideration of related concepts such as misinformation. In this context, it is worth noting that some research explicitly includes contextual information such as images or external links into the annotation decision (Jikeli et al. 2022), which is particularly helpful when labeling short texts. However, this places additional demands on the training of classification models. An alternative might be to treat threads or sequences of texts as entities instead of single messages.

It also became clear that for annotators with a stronger affiliation to qualitative disciplines, it feels unfamiliar, not to say problematic, if they are asked to take a binary yes/no decision when interpreting a text. On the other hand, the group discussion showed that it is possible to reach a shared understanding and interpretation based on the predefined categories provided in our annotation guide in most of the cases; however, direct exchange between annotators needs to be assured in the labeling process.

We find it important to make these differences and difficulties transparent, because we consider them relevant for other interdisciplinary research as well. After all, the divergences demonstrate the complexity of annotating human-written artifacts, a task which inevitably reduces complex social phenomena to a simplified classification. With this, it will hardly ever be possible to dissolve all conflicts emerging among annotators. These conflicts could also be made productive to foster explicit and careful choices of how to resolve annotator disagreements: As Gordon et al. (2022) have pointed out, the question whose voices are being heard when providing data for machine learning algorithms is often still left implicit and typically resolved by a majority vote. We think that especially for complex social phenomena, this process should gain more attention in future research – last but not least because power relations and discrimination affect people differently and are thus received with more or less sensitivity by them.

References

- Alam, F.; Dalvi, F.; Shaar, S.; Durrani, N.; Mubarak, H.; Nikolov, A.; Da San Martino, G.; Abdelali, A.; Sajjad, H.; Darwish, K.; and Nakov, P. 2021. Fighting the COVID-19 Infodemic in Social Media: A Holistic Perspective and a Call to Arms. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 913–922.
- Becker, M. J.; and Allington, D. 2021. Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online. Discourse Report No. 2. Technical report, TU Berlin, Centre for Research on Antisemitism (ZfA).
- Becker, M. J.; Troschke, H.; and Allington, D. 2021. Decoding Antisemitism: An AI-driven Study on Hate Speech & Imagery Online. Discourse Report No. 1. Technical report, TU Berlin, Centre for Research on Antisemitism (ZfA).
- Belghaus, N.; am Orde, S.; and Jakob, C. 2020. Impfgegner und die Coronapandemie: Gegen den Stich. <https://taz.de/Impfgegner-und-die-Coronapandemie/!5735702/>. Accessed: 2023-04-12.

- Bischoff, N.; Pustet, M.; and Mihaljević, H. 2022. Datasheet for the dataset TelCovACT. <https://zenodo.org/record/7093870>. Dataset.
- Bodner, J.; Welch, W.; and Brodie, I. 2020. *Covid-19 conspiracy theories: QAnon, 5G, the New World Order and other viral ideas*. Jefferson, NC: MacFarland & Co Inc.
- Butter, M. 2018. “Nichts ist, wie es scheint.” *Über Verschwörungstheorien*. Berlin: Suhrkamp Verlag.
- CeMAS – Center für Monitoring, Analyse und Strategie. 2021. Einschätzung zur verschwörungsideologischen Mobilisierung. <https://cemas.io/blog/verschwörungsideologische-mobilisierung/>. Accessed: 2023-04-10.
- Chandra, M.; Pailla, D.; Bhatia, H.; Sanchawala, A.; Gupta, M.; Shrivastava, M.; and Kumaraguru, P. 2021. “Subverting the Jewtocracy”: Online Antisemitism Detection Using Multimodal Deep Learning. arXiv:2104.05947.
- Cohen, K.; Kaati, L.; and Pelzer, B. 2021. Antisemitism in Social Media. Conspiracies, Stereotypes, and Holocaust Denial. Technical Report FOI-R-5198-SE, FOI (Swedish Defence Research Agency).
- Comerford, M.; Gerster, L.; European Commission. Directorate General for Justice and Consumers; and Institute for Strategic Dialogue. 2021. The rise of antisemitism online during the pandemic: a study of French and German content. Technical report, European Commission, Luxembourg.
- Depoux, A.; Martin, S.; Karafillakis, E.; Preet, R.; Wilder-Smith, A.; and Larson, H. 2020. The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of Travel Medicine*, 27(3).
- Ernst, N.; Engesser, S.; Büchel, F.; Blassnig, S.; and Esser, F. 2017. Extreme parties and populism: an analysis of Facebook and Twitter across six countries. *Information, Communication & Society*, 20(9): 1347–1364.
- Finkelstein, J.; Paresky, P.; Goldenberg, A.; Zannettou, S.; Jussim, L.; Riggleman, D.; Farmer, J.; Goldenberg, P.; Donohue, J.; and Modi, M. H. 2020. Antisemitic Disinformation: A Study of the Online Dissemination of Anti-Jewish Conspiracy Theories. Technical report, The Network Contagion Research Institute.
- Forschungsinstitut Gesellschaftlicher Zusammenhalt. 2020. Fact-sheet: Proteste in der Corona-Pandemie: Gefahr für unsere Demokratie? Technical report, FGZ (Forschungsinstitut Gesellschaftlicher Zusammenhalt), Standort Jena.
- Gerts, D.; Shelley, C. D.; Parikh, N.; Pitts, T.; Ross, C. W.; Fairchild, G.; Chavez, N. Y. V.; and Daughton, A. R. 2021. “Thought I’d Share First” and Other Conspiracy Theory Tweets from the COVID-19 Infodemic: Exploratory Study. *JMIR Public Health and Surveillance*, 7(4).
- Golbeck, J.; Mauriello, M.; Auxier, B.; Bhanushali, K. H.; Bonk, C.; Bouzaghane, M. A.; Buntain, C.; Chanduka, R.; Cheakalos, P.; Everett, J. B.; Falak, W.; Gieringer, C.; Graney, J.; Hoffman, K. M.; Huth, L.; Ma, Z.; Jha, M.; Khan, M.; Kori, V.; Lewis, E.; Mirano, G.; Mohn IV, W. T.; Mussenden, S.; Nelson, T. M.; Mcwillie, S.; Pant, A.; Shetye, P.; Shrestha, R.; Steinheimer, A.; Subramanian, A.; and Visnansky, G. 2018. Fake News vs Satire: A Dataset and Analysis. In *Proceedings of the 10th ACM Conference on Web Science*, 17–21. ACM.
- Gordon, M. L.; Lam, M. S.; Park, J. S.; Patel, K.; Hancock, J.; Hashimoto, T.; and Bernstein, M. S. 2022. Jury Learning: Integrating Dissenting Voices into Machine Learning Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22.
- Gover, A. R.; Harper, S. B.; and Langton, L. 2020. Anti-Asian Hate Crime During the COVID-19 Pandemic: Exploring the Reproduction of Inequality. *American Journal of Criminal Justice*, 45(4): 647–667.
- Guhl, J.; Ebner, J.; and Rau, J. 2020. The Online Ecosystem of the German Far-Right. Technical report, Institute for Strategic Dialogue.
- Hauray, T. 2019. *Antisemitismus von Links: Facetten der Judenfeindschaft*. Berlin: Aktion Courage e.V.
- Heller, J. 2015. Rumors and Realities: Making Sense of HIV/AIDS Conspiracy Narratives and Contemporary Legends. *American Journal of Public Health*, 105(1): e43–e50.
- Hepfer, K. 2015. *Verschwörungstheorien. Eine philosophische Kritik der Unvernunft*. Bielefeld: transcript Verlag.
- International Holocaust Remembrance Alliance. 2016. The working definition of antisemitism. <https://www.holocaustremembrance.com/resources/working-definitions-charters/working-definition-antisemitism>. Accessed: 2222-01-03.
- Jikeli, G.; Cavar, D.; Jeong, W.; Miehling, D.; Wagh, P.; and Pak, D. 2022. Toward an AI Definition of Antisemitism? In Hübscher, M.; and von Mering, S., eds., *Antisemitism on Social Media*, 193–212. Oxon/New York: Routledge.
- Jikeli, G.; Cavar, D.; and Miehling, D. 2019. Annotating Antisemitic Online Content. Towards an Applicable Definition of Antisemitism.
- Karakoulaki, M.; and Dessì, G. 2021. Antisemitism and anti-vax discourse in Europe. A report on conspiracy ideologies and anti-Jewish hatred on Facebook and Twitter. Technical report, Media Diversity Institute.
- Kitta, A. 2012. *Vaccinations and public concern in history: legend, rumor, and risk perception*. Number 28 in Routledge studies in the history of science, technology, and medicine. New York: Routledge.
- Kou, Y.; Gui, X.; Chen, Y.; and Pine, K. 2017. Conspiracy Talk on Social Media: Collective Sensemaking during a Public Health Crisis. In *Proceedings of the ACM on Human-Computer Interaction*, volume 1. New York, NY, USA: Association for Computing Machinery.
- Lelle, N.; and Balsam, J. 2020. Struktureller Antisemitismus ist Antisemitismus (noch) ohne Juden. <https://www.amadeu-antonio-stiftung.de/tacheles-struktureller-antisemitismus-ist-antisemitismus-noch-ohne-juden-66257/>. Accessed: 2023-04-10.
- Mahl, D.; Schäfer, M. S.; and Zeng, J. 2022. Conspiracy theories in online environments: An interdisciplinary literature review and agenda for future research. *New Media & Society*, 1–21.
- Marcellino, W.; Helmus, T. C.; Kerrigan, J.; Reiningger, H.; Karimov, R. I.; and Lawrence, R. A. 2021. Detecting Conspiracy Theories on Social Media: Improving Machine Learning to Detect and Understand Online Conspiracy Theories. Technical report, RAND Corporation.
- Marin, B. 1980. A Post-Holocaust “Anti-Semitism without Anti-Semites”? Austria as a Case in Point. *Political Psychology*, 2(2): 57–74.
- Meisner, M.; and Wiedemann, C. 2021. Sündenböcke der Pandemie. Wie die Coronakrise Rassismus nährt. In *Fehlender Mindestabstand. Die Coronakrise und die Netzwerke der Demokratiefeinde*. Freiburg, Basel, Wien: Herder (Verlag).
- Memon, S. A.; and Carley, K. M. 2020. Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset. arXiv:2008.00791.

- Mihaljević, H.; and Steffen, E. 2022. How toxic is antisemitism? Potentials and limitations of automated toxicity scoring for antisemitic online content. In *Proceedings of the 2nd Workshop on Computational Linguistics for Political Text Analysis (CPSS-2022)*.
- Moffitt, J. D.; King, C.; and Carley, K. M. 2021. Hunting Conspiracy Theories During the COVID-19 Pandemic. *Social Media + Society*, 7(3): 1–17.
- Musyal, S.; and Stegemann, P. 2020. Christchurch, Halle, Hanau: Vom Online-Hass zum rechten Terror. *Blätter für deutsche und internationale Politik*.
- Ozalp, S.; Williams, M. L.; Burnap, P.; Liu, H.; and Mostafa, M. 2020. Antisemitism on Twitter: Collective Efficacy and the Role of Community Organisations in Challenging Online Hate Speech. *Social Media + Society*, 1–20.
- Pogorelov, K.; Schroeder, D. T.; Burchard, L.; Moe, J.; Brenner, S.; Filkukova, P.; and Langguth, J. 2020. FakeNews: Corona Virus and 5G Conspiracy Task at MediaEval 2020. In *Working Notes Proceedings of the MediaEval 2020 Workshop*, volume 2882. CEUR-WS.org.
- Popper, K. R.; and Kiesewetter, H. 2003. *Die offene Gesellschaft und ihre Feinde. Band II: Falsche Propheten: Hegel, Marx und die Folgen*, volume 6. Tübingen: Mohr Siebeck, 8 edition.
- RIAS e.V. 2020. Antisemitismus im Kontext der Covid19-Pandemie. Technical report, Bundesverband der Recherche- und Informationsstellen Antisemitismus e.V. (Bundesverband RIAS), Berlin, Germany.
- RIAS e.V. 2021. Antisemitische Verschwörungsmythen in Zeiten der Coronapandemie. Das Beispiel QAnon. Technical report, American Jewish Committee Berlin Ramer Institute, Bundesverband RIAS e.V.
- Rivers, C. M.; and Lewis, B. L. 2014. Ethical research standards in a world of big data. *F1000Research*, 3.
- Rose, H. 2021. Pandemic Hate: COVID-related Antisemitism and Islamophobia, and the Role of Social Media. Technical report, Institute for Freedom of Faith & Security in Europe (IFFSE).
- Roß, B.; Rist, M.; Carbonell, G.; Cabrera, B.; Kurowsky, N.; and Wojatzki, M. 2016. Measuring the Reliability of Hate Speech Annotations: the Case of the European Refugee Crisis. In *NLP4CMC III: 3rd Workshop on Natural Language Processing for Computer-Mediated Communication 2016*.
- Sadler, N. 2021. *Fragmented Narrative. Telling and Interpreting Stories in the Twitter Age*. London: Routledge.
- Salzborn, S. 2021. Verschwörungsmythen und Antisemitismus. *Aus Politik und Zeitgeschichte*, 35-36/2021: 41–47.
- Samory, M.; and Mitra, T. 2018. 'The Government Spies Using Our Webcams': The Language of Conspiracy Theories in Online Discussions. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 1–24.
- Schmalenberger, S.; and Hübscher, M. 2022. Tertiary Antisemitism in Social Media Posts of Germany's Alternative für Deutschland. In Hübscher, M.; and von Mering, S., eds., *Antisemitism on Social Media*, 35–54. Oxon/New York: Routledge.
- Schneider, J. 2020. "Querdenken"-Demo: Kruder Vergleich mit Sophie Scholl. <https://www.zdf.de/uri/b600a1a4-af16-4ca1-8e1d-436fa49fe8cc>. Accessed: 2022-01-08.
- Schwarz-Friesel, M. 2019. *Judenhass im Internet: Antisemitismus als kulturelle Konstante und kollektives Gefühl*. Leipzig: Hentrich und Hentrich.
- Serrano, J. C. M.; Papakyriakopoulos, O.; and Hegelich, S. 2020. NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*. Association for Computational Linguistics.
- Smallman, S. 2015. Whom do you trust? Doubt and conspiracy theories in the 2009 influenza pandemic. *Journal of International and Global Studies*, 6(2).
- Smedt, T. D. 2021. Online anti-Semitism across platforms. arXiv:2112.07783.
- Starbird, K.; Maddock, J.; Orand, M.; Achterman, P.; and Mason, R. M. 2014. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In *iConference 2014 Proceedings*, 654–662.
- Steffen, E.; and Mihaljević, H. 2022. Annotation scheme and guide for antisemitism and conspiracy theory in online texts. <https://zenodo.org/record/6411663>.
- Sunstein, C. R.; and Vermeule, A. 2009. Conspiracy Theories: Causes and Cures*. *Journal of Political Philosophy*, 17(2): 202–227.
- The Lancet Infectious Diseases. 2020. The COVID-19 infodemic. *The Lancet. Infectious Diseases*, 20(8).
- Uscinski, J. E.; Enders, A. M.; Klofstad, C.; Seelig, M.; Funchion, J.; Everett, C.; Wuchty, S.; Premaratne, K.; and Murthi, M. 2020. Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review*, 1(3).
- Warner, W.; and Hirschberg, J. 2012. Detecting Hate Speech on the World Wide Web. In *Proceedings of the Second Workshop on Language in Social Media*, 19–26. Montréal, Canada: Association for Computational Linguistics.
- Winter, H.; Gerster, L.; Helmer, J.; and Baaken, T. 2021. Überdosis Desinformation: Die Vertrauenskrise. Technical report, Institute for Strategic Dialogue.
- Wood, M. J. 2018. Propagating and Debunking Conspiracy Theories on Twitter During the 2015–2016 Zika Virus Outbreak. *Cyberpsychology, Behavior, and Social Networking*, 485–490.
- Zannettou, S.; Finkelstein, J.; Bradlyn, B.; and Blackburn, J. 2020. A Quantitative Approach to Understanding Online Antisemitism. In *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020*, volume 14, 786–797. AAAI.