On the Globalization of the QAnon Conspiracy Theory Through Telegram

Hoseini, Mohamad; Melo, Philipe; Benevenuto, Fabricio; Feldmann, Anja; Zannettou, Savvas

DOI
10.1145/3578503.3583603

Publication date
2023

Document Version
Final published version

Published in
WebSci 2023 - Proceedings of the 15th ACM Web Science Conference

On the Globalization of the QAnon Conspiracy Theory Through Telegram

Mohamad Hoseini
Max Planck Institute for Informatics
Saarbrücken, Germany
mhoseini@mpi-inf.mpg.de

Philipe Melo
Federal University of Minas Gerais
Belo Horizonte, Brazil
philipe@dcc.ufmg.br

Fabrício Benevenuto
Federal University of Minas Gerais
Belo Horizonte, Brazil
fabricio@dcc.ufmg.br

Anja Feldmann
Max Planck Institute for Informatics
Saarbrücken, Germany
anja@mpi-inf.mpg.de

Savvas Zannettou
Delft University of Technology
Delft, Netherlands
s.zannettou@tudelft.nl

ABSTRACT

QAnon is a far-right conspiracy theory that has implications in the real world, with supporters of the theory participating in real-world violent acts like the US capitol attack in 2021. At the same time, the QAnon theory started evolving into a global phenomenon by attracting followers across the globe and, in particular, in Europe, hence it is imperative to understand how QAnon has become a worldwide phenomenon and how this dissemination has been happening in the online space. This paper performs a large-scale data analysis of QAnon through Telegram by collecting 4.4M messages posted in 161 QAnon groups/channels. Using Google's Perspective API, we analyze the toxicity of QAnon content across languages and over time. Also, using a BERT-based topic modeling approach, we analyze the QAnon discourse across multiple languages. Among other things, we find that the German language is prevalent in our QAnon dataset, even overshadowing English after 2020. Also, we find that content posted in German and Portuguese tends to be more toxic compared to English. Our topic modeling indicates that QAnon supporters discuss various topics of interest within far-right movements, including world politics, conspiracy theories, COVID-19, and the anti-vaccination movement. Taken all together, we perform the first multilingual study on QAnon through Telegram and paint a nuanced overview of the globalization of QAnon.

CCS CONCEPTS

• **Information systems** → **Social networks**; **Data mining**; **Chat**; • **General and reference** → **Measurement**; • **Security and privacy** → **Social network security and privacy**.

KEYWORDS

QAnon, Telegram, social media, topic modeling, toxicity analysis

ACM Reference Format:
Mohamad Hoseini, Philipe Melo, Fabrício Benevenuto, Anja Feldmann, and Savvas Zannettou. 2023. On the Globalization of the QAnon Conspiracy Theory Through Telegram. In *15th ACM Web Science Conference 2023 (WebSci '23), April 30–May 01, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3578503.3583603

WebSci '23, April 30–May 01, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0089-7/23/04.
https://doi.org/10.1145/3578503.3583603

1 INTRODUCTION

We witness an explosion in the spread and popularity of conspiracy theories on the Web. People might get affected by continuous exposure to conspiratorial content, and this exposure may influence them into perpetrating violent acts in the real world. For instance, the Pizzagate conspiracy theory was the driving factor for a shooting at a pizzeria in Washington DC in 2016 [43]. Taken all together, there is a pressing need to understand how these conspiracy theories spread online and how users are radicalized from exposure to conspiratorial content.

One conspiracy theory that attracts high engagement from people is QAnon, which is a conspiracy theory alleging that a secret group of people (i.e., a cabal consisting of Democratic politicians, government officials, and Hollywood actors) were running a global child sex trafficking ring and were plotting against former US President Donald Trump[51]. Between 2017 and 2021, this conspiracy attracted many new followers across the globe and essentially evolved into a cult. Worryingly, the followers of the QAnon conspiracy theory have begun making threats or participating in violent real-world incidents (e.g., Capitol attack in 2021 [18]), hence highlighting the impact that the conspiracy theory has on the real world [6].

Motivated by the negative impact that QAnon has in the real world, mainstream platforms like Facebook, Twitter, and YouTube, started moderating and removing QAnon-related content [4, 5, 52, 53]. Then, QAnon supporters sought new online "homes" in less-moderated platforms and migrated to other platforms like Parler and Telegram [12]. Also, QAnon became a global phenomenon; the QAnon conspiracy theory has accumulated new followers worldwide, particularly in European countries like Germany and Spain [41]. Overall, it is crucial to understand how QAnon evolved and became a global phenomenon that has not yet been investigated on a large scale by any other work. To do this, we use Telegram as the source of our study for two reasons. First, anecdotal evidence suggests that QAnon followers migrated to Telegram after bans on other platforms [12]. Second, Telegram is a rapidly growing platform with worldwide coverage [45], hence it is the ideal platform for studying QAnon across the globe.

Hypotheses. We focus on testing the following hypotheses:

- **H1 - Activity:** We hypothesize that QAnon activity in Telegram increases in volume over time (due to moderation actions on other platforms) to a larger extent compared to groups focusing on other topics. Also, we hypothesize that there are substantial changes in the popularity of the used languages over time due to anecdotal evidence suggesting QAnon is popular in Europe [41].
- **H2 - Toxicity:** QAnon Content in Telegram is more toxic compared to the content on groups/channels focusing on other topics.
- **H3 - Topics:** We hypothesize that QAnon followers discuss various topics related to Politics, they are sharing false information, and that the popularity of these topics changes over time.

We argue that these three hypotheses are equally important and need to be studied together. **H1** allows us to understand how active the QAnon movement is on Telegram and especially how this activity has evolved. In **H2** and **H3** we focus on what content is shared in QAnon groups/channels, how these discussions differ over time, how the discourse differs from previous work or other platforms, and how toxic is content; this is equally important as it allows us to understand what the topics of discussions are and whether the QAnon discourse is becoming more toxic, which is of paramount importance given previous participation of QAnon followers in real-world violent acts. For instance, if QAnon followers have an anti-vax ideology, QAnon followers will likely participate in real-world protests related to the anti-vax movement; such findings can help us be better prepared for dealing with such protests and potentially mitigating real-world violence.

To test the above-mentioned hypotheses, we perform a large-scale data collection and analysis of QAnon-related groups/channels on Telegram. Overall, we collect 4.4M messages shared in 161 Telegram groups/channels between September 2019 and March 2021. Using Google's Perspective API [32], we investigate the toxicity of QAnon content on Telegram and assess whether the movement is becoming more toxic over time and whether there are substantial differences across languages. Also, using a multilingual BERT-based topic modeling approach [3], we study the QAnon discourse across multiple countries/languages.

Main findings. Our study provides some key findings:

- We find that QAnon activity in our dataset increased substantially during 2021 with an increase of almost 5x in terms of the number of messages and senders, while our baseline dataset has an increase of only 2x. Furthermore, by comparing content across languages, we find that German QAnon content overshadowed English (on average 54.7% for German and 28.4% for English) in popularity after June 2020. Our findings support our first hypothesis **(H1)**.
- By analyzing the toxicity of QAnon-related messages in our dataset, we find that content shared in Portuguese and German is more toxic compared to English (8.6% of the Portuguese messages and 2.8% of the German messages are toxic, while for English we only have 1% of QAnon messages being toxic). At the same time, we find that QAnon content posted in English and Portuguese is more toxic compared to our baseline dataset (3.6x and 1.2x, respectively). Our results partly support our second hypothesis (**H2**), since for German we find that the baseline actually had 1.15x more toxic messages compared to our QAnon dataset.
- Our topic modeling analysis highlights that QAnon has evolved into discussing various topics of interest within far-right movements across the globe. We find several topics of discussion like world politics, conspiracy theories, COVID-19, and the anti-vaccination movement (**H3**).

2 BACKGROUND & RELATED WORK

The conspiracy theory emerged in October 2017 with a post on 4chan by a user named "Q," who claimed that he was an American government official with classified information about plots against then-President Donald Trump. "Q" continued disseminating cryptic messages about the QAnon conspiracy theory (called "Q drops") mainly on 8chan. The QAnon conspiracy theory has amassed a following in fringe Web communities like 4chan/8chan and mainstream ones like Facebook [42] and Twitter, especially after then-president Donald Trump retweeted QAnon-related content [28]. QAnon followers use their motto "Where We Go One, We Go All" (or simply wwg1wga) to tag content related to QAnon.

Over the past years, followers of the QAnon conspiracy theory have made violent threats or been linked to several incidents of real-world violence [6], with the Federal Bureau of Investigation (FBI) labeling it as a potential domestic terrorist threat [50]. In particular, on January 6th, 2021, supporters of the QAnon conspiracy theory attacked the US capitol in an attempt to overturn Donald Trump's defeat in the 2020 US elections by disrupting the Congress that was in the process of formalizing Joe Biden's victory [18]. Due to these threats and violent incidents, mainstream platforms like Facebook [4], Twitter [5], Reddit [52], and YouTube [53] started monitoring and removing QAnon-related groups, subreddits, and users. Naturally, following these content moderation interventions, supporters of the QAnon conspiracy theory flocked to other fringe Web communities with lax moderation, like Parler [2] and Gab [23, 54], or messaging platforms like Telegram [12].

Even though the idea of the QAnon conspiracy theory is US-centric, QAnon became a global phenomenon, in particular among people with far-right ideology. In 2020, the QAnon theory spread to Europe [41]. The conspiracy theory is nowadays shared among people from Spain, Italy, the United Kingdom, and Germany, one of the most popular "representatives" in Europe [7].

Previous work investigates several aspects of the QAnon conspiracy theory. Papasavva et al. [30] analyze content toxicity and narratives in a QAnon community on Voat, finding that discussions in popular communities on Voat are more toxic than in QAnon communities. Aliapoulios et al. [2] provide a dataset of 183M Parler posts, and they highlight that QAnon is one of the dominant topics on Parler. Miller [27] investigates a sample of QAnon-related comments on YouTube, highlighting the international nature of the movement. Garry et al. [16] explore QAnon supporters' behavior in spreading disinformation on Gab and Telegram, finding that the dissemination of disinformation is one of the main reasons for the growth of QAnon conspiracy. Hannah [19] also investigate the reasons for the growth of QAnon, finding that sharing and discussing

Q drops is one of the main reasons. Chandler [9] investigates how QAnon followers are influenced by Q drops, finding that Q drops focus on the perceived allies or enemies of QAnon. Planck [34] compares the QAnon community's rhetoric with a mainstream conservative community on Twitter, finding that tweets posted by QAnon supporters are more violent. Papasavva et al. [29] investigate a dataset of 4.9K canonical Q drops from six aggregation sites, finding inconsistencies among the drops and demonstrating that the drops have multiple authors. Ferrara et al. [14] investigate 240M election-related tweets finding that 13% of users spreading political conspiracies (including QAnon) are bots. Sipka et al. [44] compared the language and narratives of QAnon-related content on Parler, Gab, and Twitter on a dataset of about 100k posts with the #QAnon hashtag and they find a prevalence of anti-social language on Parler, while Gab has the most conspiratorial and toxic content. Phadke et al. [33] characterize 2K posts from 4chan and 8chan and 1.2M comments from 12 subreddits to understand the social imaginary within QAnon online communities and identify how their members express their belief and dissonance towards the conspiracy. Engel et al. [13] collected over 12M of posts from early QAnon users on Reddit and characterized how users engage in the QAnon conspiracy, showing they were dedicated and committed to the movement even after a massive ban of the QAnon from Reddit. Pasquetto et al. [31] examined the disinformation infrastructure of QAnon built on Italian digital media by a digital ethnography over a period of eleven months of QAnon activities on Facebook, Twitter, and Telegram communities. They observed a top-down design in the Qanon structure online in which decisions are made and imposed on the community while the followers are expected to participate and share but they are not allowed to directly contribute to how information is organized or curated.

Telegram. Telegram is a popular messaging platform, with 400M monthly active users, as of April 2020 [45]. Users can create public and private chat rooms called *channels* or *groups*. Channels support few-to-many communication, where only the creator and a few administrators can post messages, while groups support many-to-many communication (all members can post). Groups and channels can have a large number of members, with a limit of 200K members for groups and an unlimited number of members in channels. Telegram users can share messages in groups/channels, with Telegram-supporting text, images, videos, audio, stickers, etc. Users can forward messages between groups/channels, with Telegram showing an indication in its user interface that the message is forwarded and the source group/channel. Due to its privacy policy and encrypted nature (i.e., "all data is stored heavily encrypted"), Telegram attracted the interest of dangerous organizations like terrorists [47] and far-right groups [1]. Given this history and use of Telegram, in this work, we study the QAnon conspiracy theory through the lens of the Telegram platform. Also, we select Telegram as it is popular across the globe, hence assisting us in studying the globalization of the QAnon conspiracy theory.

3 DATASET

An inherent challenge when studying phenomena through platforms like Telegram is to discover groups/channels related to the topic of interest.

To discover groups/channels related to QAnon, we follow the methodology by Hoseini et al. [20]: 1) search on Twitter and Facebook for URLs to Telegram groups/channels; 2) collect metadata for each group/channel; 3) select groups/channels based on QAnon-related keywords. 4) validate the selected groups/channels; 5) join and collect all messages from all QAnon groups/channels; and 6) expand our QAnon groups/channels based on forwarded messages shared in already discovered QAnon groups/channels and repeat Step 5.

1. Discovering groups/channels. We use Twitter and Facebook to discover Telegram groups and channels. For Twitter, we use the Search and Streaming API to collect tweets that include Telegram URLs, following the methodology by Hoseini et al. [20], while for Facebook, we use the Crowdtangle API to obtain posts including Telegram URLs [10]. For both data sources, we perform queries with three URL patterns obtained from Hoseini et al. [20]: *t.me*, *telegram.me*, and *telegram.org*. Note that the list of these patterns is not exhaustive; there is also the *tg://join?invite* pattern, however, we did not include it in our collection since our initial experiments showed that they are rarely shared on Twitter/Facebook (less than 0.1% more URLs discovered by including this specific pattern). We collect Twitter and Facebook posts, including Telegram URLs between April 8, 2020, and October 10, 2020, ultimately collecting a set of 5,488,596 tweets and 14,004,394 Facebook posts that include a set of 922,289 unique Telegram URLs. Note that the Crowdtangle API tracks and provides data only from publicly available Groups and Pages (i.e., does not include user timeline posts).

2. Collecting group/channel metadata. Having discovered a set of Telegram URLs, we then use Telegram's Web client and obtain basic group/channel metadata from the URLs. These include: a) Name of the group/channel; b) Description of the group/channel; c) Number of members; and d) the URL type (i.e., channel or group).

3. Selecting QAnon groups/channels. The next step is to narrow down the set of groups/channels to the ones that mention QAnon. To do this, we search for the appearance of QAnon-related keywords on Twitter/Facebook posts that shared Telegram URLs or on the group/channel metadata obtained from Step 2. We use two QAnon-related keywords: *qanon* and *wwg1wga*. The former refers to the conspiracy theory itself, while the latter is the QAnon movement's motto that refers to "Where We Go One We Go All." We select these specific keywords mainly because they are prevalent and used extensively by members of the QAnon movement. Overall, we find 204 Telegram groups/channels that include the above keywords in their group/channel metadata or any posts collected from Twitter/Facebook.

4. Validating QAnon groups/channels. Then, we validate that the selected groups/channels are related to QAnon and remove any groups/channels that are not directly related (e.g., mentioning QAnon only once because of mentions in the news). To do this, an author of this study, who has previous experience with the QAnon conspiracy theory, manually annotated the 204 groups/channels obtained from Step 3. The annotator viewed each group/channel via Telegram's Web client and spent 5-10 minutes reading the content shared in the group/channel and checking the group/channel metadata to decide whether the group/channel is related and supports the QAnon conspiracy theory. The annotator focused only on

Table 1: Overview of our Telegram dataset.

Dataset	Source	#Groups	#Senders	#Messages
QAnon	Twitter/FB	78	92,322	3,503,381
	Forwarded	84	84	903,611
	Total	161	92,406	4,406,992
Baseline	Twitter/FB	869	195,499	7,983,230

selecting groups/channels that were promoting QAnon or were discussing theories related to QAnon and avoided selecting groups/channels that simply mentioned some news about QAnon but their primary focus was on another topic. Note that since many groups/channels are in languages other than English, the annotator used Google's translate functionality to translate content into English. Overall, we annotate all 204 groups/channels and find 77 QAnon groups/channels.

5. Joining and collecting messages in QAnon groups/channels. The next step in our data collection methodology is to join the QAnon groups/channels and collect all their messages. We join all QAnon groups/channels, and then we use the Telethon library [49], which uses Telegram's API [48] to collect all the messages shared within these groups. Note that we only join and collect data from public groups/channels. Initially, we collect 3.5M messages shared in 77 QAnon groups/channels between September 1, 2019, and March 9, 2021 (see Table 1).

6. Expanding QAnon groups/channels. During our manual validation of the QAnon groups and channels, we observed many messages shared in QAnon groups/channels that are forwarded messages from other groups/channels. Aiming to expand our set of QAnon groups/channels, we extract all groups/channels that forwarded messages in the 77 already discovered QAnon groups/channels and manually validate (see Step 4) the top 200 groups/channels in terms of the number of forwarded messages. Note that we only validate the top 200, as manually checking and validating the groups/channels is time-consuming. Using this approach, we discover an additional 84 QAnon groups/channels. Then, we repeat Step 5 for the newly discovered groups and collect all of their messages. Overall, by combining the initial dataset and the one after expanding the QAnon groups/channels, we obtain a set of 4.4M messages shared in 161 QAnon groups/channels between September 1, 2019, and March 9, 2021 (see Table 1).

Baseline dataset. We collect a baseline dataset to compare it with our QAnon dataset. To collect our baseline dataset, we follow Steps 1, 2, 3, and 5, with the only difference that we use a different set of keywords for selecting the groups/channels (note that we do not validate and manually check the groups/channels because they are not focusing on a specific topic). Specifically, we use a set of keywords obtained from First Draft [15], an organization that aims to fight disinformation on the Web. First Draft provided us with a list that includes 133 keywords/phrases[1] about important events in 2020 (e.g., the US election and the COVID-19 pandemic). Overall, we joined 869 groups/channels and collected 7.9M messages shared between September 1, 2019, and March 9, 2021 (see Table 1).

[1]Available in https://telegra.ph/Keywords-08-03.

Limitations. Our data collection and dataset have some limitations. First, as with all studies focusing on messaging platforms like Telegram and WhatsApp [37, 38], we cannot assess how representative our collected dataset is. This is because there is no single vantage point to discover all Telegram groups/channels; due to this, we focus only on groups/channels shared on Twitter and Facebook. Therefore, we likely miss QAnon groups/channels simply because they were not shared on Twitter or Facebook. Second, our dataset is biased toward more recent groups/channels active in 2020. Hence, we likely miss some groups/channels that were created before 2020 and eventually became inactive. Finally, our keyword filtering is based on just two keywords, which indicates that we initially miss QAnon groups/channels that do not use these keywords (see Step 3 above). We mitigated this by expanding our dataset based on forwarded messages (Step 6).

Ethical considerations. Before collecting any data, we obtained approval from our institution's ethical review board. Also, we stress that: a) we work entirely with publicly available data; b) we do not make any attempt to de-anonymize users; and c) we do not track users across platforms. Overall, we follow standard ethical guidelines [40] throughout our data collection and analysis.

4 RESULTS

Here, we present our analysis for investigating our three hypotheses related to the activity, toxicity, and topics in our QAnon dataset.

4.1 H1: Activity

We start our analysis by looking into the general activity across the QAnon groups/channels and how it differs from our baseline dataset. Fig. 1 shows the percentage of active groups, messages, and senders per week in our dataset. When looking at the activity of groups over time (see Fig. 1(a)), we observe that for both QAnon and baseline datasets, we have an increasing number of active groups over time; for QAnon, we have 11.8% active groups by September 2019, and by March 2021 the active groups/channels increase to 86%. For the baseline dataset, we find 12% and 58% active groups for September 2019 and March 2021, respectively. These increases in the overall activity for both datasets are likely due to Telegram becoming more popular over time [46] and the platform is onboarding more users that create more groups/channels on various topics of interest.

When looking at the activity of messages and senders (Fig. 1(b) and Fig. 1(c)), we again observe an increase in activity for both datasets over time. Specifically, by April 2020, we have 1% of all messages for QAnon and 1.5% for the baseline dataset. These percentages increase later on and by 2021, we observe an activity of 2% for the baseline, while for QAnon, we have an activity of over 3% with specific weeks increasing even over 5%. Importantly, we observe that the QAnon activity surpasses the baseline activity by October 2020, which likely indicates that the QAnon movement on Telegram substantially increased by that time, even surpassing other topics of interest.

The larger increase in QAnon compared to the baseline is likely because Facebook [4] removed accounts and groups related to QAnon from their platforms during October 2020, hence users likely migrated to alternative platforms like Telegram. Also, for the QAnon dataset, we observe a peak in activity during early 2021

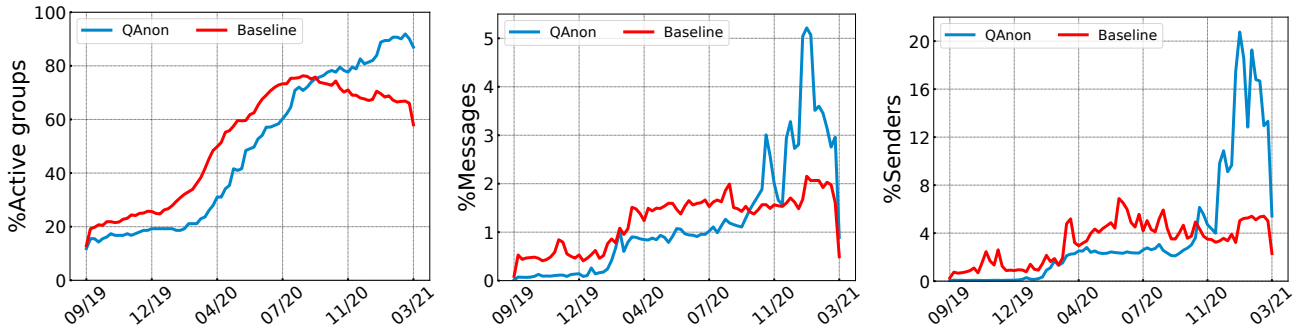


Figure 1: Activity within QAnon and baseline groups/channels over time. We report the percentage (over the entire dataset) of active groups/channels, messages, and senders per week.

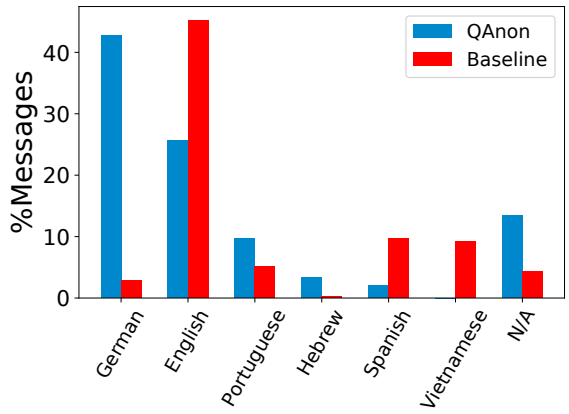


Figure 2: Percentage of messages for the top languages in our datasets. We report the union of the top five languages in our QAnon and baseline dataset. N/A refers to messages that we were unable to infer a language (e.g., messages consisting entirely of URLs).

(over 5% of all messages and over 20% of the users were actively sharing messages), which coincides with the attack in the US capitol by QAnon supporters. This initial analysis indicates that the QAnon conspiracy theory is growing rapidly on Telegram in terms of the number of groups/channels (almost 7x increase while baseline has 4.8x increase), the number of messages (over 5x increase while baseline has 2x), and the number of users sharing messages (over 5x increase while baseline has 2x).

Next, we analyze the languages that appear in our QAnon and baseline datasets. Fig. 2 shows the percentage of messages for the top five languages in our QAnon and baseline datasets (the figure includes the union of the top five languages on both datasets). We observe substantial differences in the popularity of languages across the two datasets; German is the most popular language in our QAnon dataset, with 42.8% of all messages (only 2.9% in the baseline). The most popular language is English for the baseline dataset, with 45.2% of all messages (25.7% for the QAnon dataset). Other popular languages in our QAnon dataset are Portuguese (9.7%), Hebrew (3.3%), and Spanish (2.1%).

Next, we look into how the popularity of the five most popular languages changed over time to understand how QAnon became a global phenomenon on Telegram. Fig. 3 shows the popularity of the languages over time in our QAnon dataset (we omit the figure for the baseline since there are no substantial differences in the popularity of languages in the baseline dataset). We observe that English was the most popular language between September 2019 and December 2019, with over half of the QAnon-related messages posted in English (55%), with German having a substantial percentage (39.3% of the QAnon-related messages). Furthermore, between February and April 2020, we observe a substantial increase in the popularity of the Portuguese language, which became the most popular language with 48.4% of the messages of this period, overshadowing both English and German. This period coincides with the beginning of the COVID-19 pandemic in Brazil, when the virus was first confirmed to have spread to Brazil in February 2020 [8]. Finally, after June 2020, we find that German is consistently the most popular language in our dataset, reaching 54.7% of the messages, followed by English (28.4%) and Portuguese (6.3%) having stable popularity.

Remarks. Our results confirm our first hypothesis. QAnon's popularity in our dataset is rapidly increasing and surpassing the baseline dataset (almost a 5x increase in messages and senders in 2021, whereas for the baseline dataset, we only find a 2x increase). Also, we observe substantial shifts in language popularity in our QAnon dataset, with English being the most popular between September 2019 and February 2020 (55%), Portuguese being the most popular between February 2020 and April 2020 (48.4%), while German is the most popular language after June 2020 (54.7%). These findings prompt the need to further investigate the multilingual aspect of conspiracy theories that become a global phenomenon like QAnon.

4.2 H2: Toxicity

Here, we investigate the toxicity of content shared in our QAnon and baseline datasets. The QAnon movement has links with events of real-world violence, hence it is important to analyze the toxicity of QAnon discussions on Telegram. We aim to uncover whether

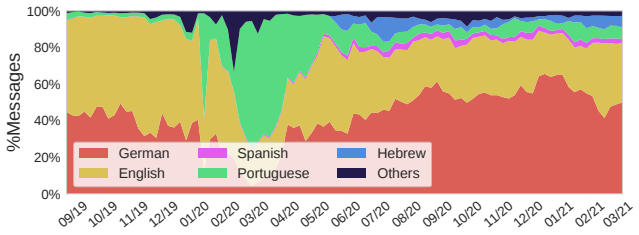


Figure 3: Distribution of messages across languages. We report the percentage of messages in each language per week.

Table 2: Percentage of toxic messages.

English		German		Portuguese	
QAnon	1.01%	**QAnon**	2.86%	**QAnon**	8.62%
Baseline	0.28%	**Baseline**	3.30%	**Baseline**	6.99%
Voat	6.51%	**Voat**	N/A	**Voat**	N/A

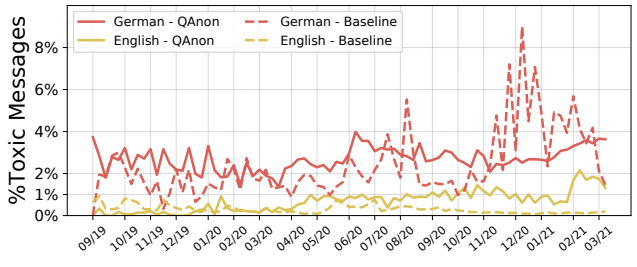
QAnon discussions in our dataset are more toxic than other discussions and how toxicity changes over time (i.e., are QAnon discussions in our dataset becoming more toxic over time).

Toxicity Assessment. To quantify how toxic the content in our datasets is and whether there are changes over time, we use Google's Perspective API [32] to annotate each message in our dataset with a score that reflects how rude or disrespectful a comment is. Following Ribeiro et al. [39], we use the SEVERE_TOXICITY model provided by the Perspective API, mainly because it is robust to positive uses of curse words. We use Perspective API for annotating content mainly because it offers production-ready models that support multiple languages; as of May 2021, the Perspective API supports English, Spanish, French, German, Portuguese, Italian, and Russian. The Perspective API allows us to assess the toxicity of messages posted in any of the seven languages above, which corresponds to 65% of the messages in our dataset. The rest of the messages do not include any text (20% are sharing only audio, video, or images) or are in other languages (15%) that the Perspective API does not support. Note that the use of the Perspective API to assess the toxicity of content is likely to introduce some false positives or biases [11]. Previous work [17], has validated the performance of the Perspective API, however, it focuses mainly on the English annotations.

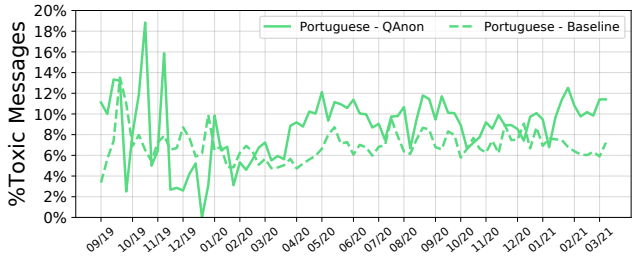
Given that, likely, the Perspective API performs differently across languages, we make a manual validation of the performance of the Perspective API in the three most popular languages in our dataset: English, German, and Portuguese (see Appendix for details). Based on our annotation, we treat a message as toxic if it scores over 0.7 for English, 0.75 for German, and 0.65 for Portuguese. We use these specific thresholds because our validation procedure demonstrates that we achieve the highest performance in terms of F1 score when using them. Also, we limit our analyses to the three aforementioned languages mainly because we did not validate the performance in other languages as this task is outside the scope of this work and requires the recruitment of native speakers for each language.

Results. First, we look into the prevalence of toxic messages in our QAnon dataset by comparing it with our baseline dataset, and the Voat dataset obtained from [30]. Voat was a social network that hosted many QAnon followers that migrated from other platforms (Voat was shut down in December 2020). We use Voat as a baseline because its another platform where QAnon followers migrated to after bans from mainstream platforms, the time period of the Voat dataset is a subset of the time period in our dataset, and because QAnon was very popular on Voat before the platform's shut down [30]. Table 2 reports the percentage of messages that are toxic in our QAnon/baseline datasets and the above-mentioned Voat dataset. First, we observe that QAnon discussions in our dataset shared in German and Portuguese tend to be more toxic than discussions in English (2.86% for German and 8.62% for Portuguese compared to 1.01% for English). These findings are particularly alarming when combined with the popularity results of these languages in our QAnon dataset. This is because Portuguese and German are overshadowing English during 2020 (see Fig. 3), hence less toxic discussions in English give way to more toxic discussions in Portuguese and German. Second, for English and Portuguese, we observe that QAnon discussions in our dataset are more toxic than our baseline dataset; 3.6x greater percentage for English and 1.2x greater percentage for Portuguese. On the other hand, we observe that the baseline dataset has a greater percentage of toxic messages (1.15x more) for German. Third, the QAnon Voat dataset (only available in English) has a substantially larger percentage of toxic messages than the Telegram one (Voat has a 6.4x larger percentage than Telegram). This difference in the toxicity levels between Voat and Telegram is likely due to the fundamental differences between the two platforms and the audience they attract. While Voat is a fringe Web community mainly discussing conspiracy theories, Telegram is a more general-purpose and mainstream platform. Nevertheless, Voat's toxicity levels are comparable with our QAnon dataset in other languages (i.e., Portuguese), which highlights the need to monitor and further study the QAnon movement across the globe, particularly on platforms like Telegram. These results indicate that platforms like Telegram, which allow users to create their own sub-communities, can be exploited to create fringe communities that can disseminate harmful and toxic content in such prevalence comparable with other notorious communities known for the dissemination of hateful content like Voat.

We also look into how the toxicity in our QAnon and baseline datasets changes over time. Fig. 4 shows the weekly percentage of toxic messages. We observe that for English, we have a steady increase of toxic messages over time in our QAnon dataset; before April 2020, the percentage of toxic messages is below 1%, between April 2020 and December 2020 is stable at 1%, while during 2021, we find 2x more toxic messages (2% of all messages are toxic). For German, we observe that our QAnon dataset has a larger percentage of toxic messages between September 2019 and July 2020 (on average 2.7% for QAnon and 1.7% for baseline), while the baseline has a substantially larger percentage after November 2020 (on average 2.8% for QAnon and 4.6% for baseline). For Portuguese, we observe some big peaks in toxicity before January 2020, however, these peaks are likely because we only have a small number of messages during that period (see Fig. 3). Looking at the rest of the figure, we can observe a big increase from 6% to 12% between early 2020 and



(a) English & German



(b) Portuguese

Figure 4: Percentage of toxic messages per week.

May 2020. We manually examined some of these toxic messages, finding that they are related to the COVID-19 pandemic in Brazil, including anti-vaccine conspiracies. Also, we find politics-related messages that attack two Brazilian ex-ministers that left the government during this period. Overall, similarly to English, we observe an increasing trend of toxic messages posted in Portuguese over time in our QAnon dataset.

Remarks. Our analysis partly confirms our second hypothesis; we find that QAnon discussions in our dataset are more toxic than our baseline for English and Portuguese (1.2x and 3.6x more toxic messages for English and Portuguese, respectively). Our German QAnon dataset does not support our hypothesis since we find a higher percentage of toxic messages in our baseline (1.6x more toxic messages in the baseline). Alarmingly, our results show an increase in QAnon content toxicity over time in our Telegram dataset. These findings emphasize the importance of monitoring such groups within the Telegram platform and taking moderation actions in cases where communities orchestrate campaigns that might have a negative impact in the real world (e.g., real-world violence).

4.3 H3: Topics

Thus far, we have analyzed our datasets' activity and toxicity aspects without analyzing the discussion topics. Here, we analyze the content of the messages shared within QAnon groups/channels using a BERT topic modeling approach.

BERT Topic Modeling. To analyze QAnon discourse across multiple languages, we use a Bidirectional Encoder Representations from

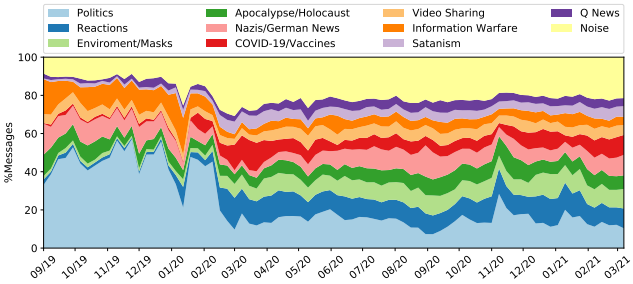


Figure 5: Percentage of messages across topics per week.

Transformers (BERT)-based topic modeling methodology by Angelov [3]. We use a pre-trained multilingual BERT model (distiluse-base-multilingual-cased) from Reimers and Gurevych [36] to embed documents from multiple languages to the same high-dimensional vector space. We select this specific model mainly because it supports 50 languages and performs well in semantic similarity tasks. Then, we use Uniform Manifold Approximation and Projection (UMAP) proposed by McInnes et al. [25] to reduce the dimensionality of the extracted embeddings. This is an important step, as it allows us to increase the performance and scalability of the next step (i.e., clustering). Then, we group the reduced embeddings using the HDBSCAN algorithm [24]. We treat each cluster as a separate topic and then we use hierarchical reduction (i.e., iteratively combining the most similar clusters) to obtain a small number of high-level topics/clusters. Finally, to generate topic representations, we calculate the centroid of each cluster based on the embeddings of all documents in the cluster and then select the most similar words (based on the BERT embeddings of the words that appear in the documents of each cluster) that are closer to the centroid.

We apply this topic methodology after preprocessing all messages by removing emojis and URLs from the text and filtering out messages with an empty body (i.e., messages sharing only URLs, emojis, videos, or images). We focus only on messages posted in the top six languages in our QAnon dataset and we remove very short messages (less than 5 words). After our preprocessing steps, we end up with a set of 2.2M messages, which is the input to our topic modeling approach. Since our topic modeling approach relies on UMAP, a stochastic technique, our approach can yield varying results on different runs. To alleviate this, we train five separate topic models and select the one that provides the highest average coherence score. For each model, we hierarchically reduce the number of topics to N (we experimented with numbers between 10 and 20) by iteratively combining the most similar clusters until we end up with N clusters (each cluster represents a high-level topic). For each model, we calculate the coherence scores for $N \in \{10, 15, 20\}$ and then select the model with the largest average coherence score. To select the number of topics to present, we again select N based on the coherence scores; we obtain the largest coherence score when $N = 10$ (0.58 vs. 0.53 and 0.52 for 15 and 20, respectively). Below, we report our analysis using the best-performing model in terms of the coherence scores.

Table 3 reports the ten extracted high-level topics along with the number of messages that are mapped to each topic (note that

Table 3: Topics extracted from our multilingual BERT topic modeling. We report the main theme of the topic, example terms that describe the topic, and the number of messages that are mapped to each topic.

Topic	Terms	#Messages
Politics	trumppresidente, trumppresident, presidenciales, presidencial, senatswahlen, presidential, presidenciais, diabolsonaro, kongresswahlen, presidency, obama, presidenttrump, republicano, impeach, impeachment	353,696
Reactions	hahahahahaha, hahahaha, hahahahaha, hahahah, hahaha, ohhhh, ahhhh, mhhh, ahhh, hahah, ohhh, uhh, haha, ahh, ohh, dahingerafft, yhwh, mhh, hmmmm, oooh	257,039
Enviroment/Masks	wwf, stromaggregate, noah, kohlekraftwerke, atomkraftwerken, boooooooooom, maskenkontrolle, atomkraftwerke,mikroelektronik, kontrollgruppe	206,848
Nazis	nazideutschland, nazistas, neonazis, nazista, fascists, fascist, fascistas, polizeigesetz, fascism, nazis, fascismo, bundespolizei, faschistischen, kriminalpolizei, massenproteste	175,201
Apocalypse/Holocaust	wikileaks, killuminati, reichstagssturm, apocalisse, apocalipse, rechtsradikaler, apokalypse, johnfkennedyjr, apocalypse, weltkriegen, holocausto, doomsday, rechtsradikale, holocaust	169,319
COVID-19/Vaccines	impfenden, vacinacao, vaccinations, vaccines, impfen, impfens, vaccination, vacunarse, grippevirus, grippeviren, vacunado, vacinar, ungeimpft, virusnachweis, geimpften, impfgruppe	159,787
Video Sharing	videokanal, videobeitrag, youtubekanal, videonachricht, originalvideo, schockvideo, kurzvideos, video, videolink, beweisvideos, videointerview, youtubelink, videoschalte, videobotschaft, youtube	119,238
Information Warfare	staatsterror, infokrieg, cyberkrieg, patriotsfight, atomkrieg, terrorists, terroristas, militari, weltkrieges, weltkrieg, staatsfeind, militares, vietnamkrieg, weltkriegs, military	116,618
Satanists	satanists, satanismo, antichristen, satanisten, satanism, satanistas, antichrist, satanismus, hausdemokraten, satanist, satanistischen, anticristo, cristianismo, satanischer, satanic	105,151
Q News	wahrheitssuche, qnews, wahrheitskanal, halbwahrheiten, wahrheitssucher, justthenews, hoax, breakingnews, faktenchecker, wahrheiten, extremnews, telenews, conspiracies, freetruthmedia, q_for_you_news,	97,929

21% of the messages are not mapped into any topic and they are considered noise), while Fig. 5 shows the distribution of messages into these topics per week in our dataset.

The most popular topic in our QAnon dataset is Politics (353K messages); by examining the terms and some messages mapped to this topic, we find political messages in various countries like the USA, Germany, Brazil, and Italy. These results compound previous findings from Papasavva et al. [30] and Miller [27] that found political discussions and discussions of international topics in Voat's and YouTube's QAnon community. Other popular topics in our QAnon dataset are related to reacting to other messages during a discussion (257K messages), discussions about environmental issues and masks (206K messages), discussing topics related to Nazis/Neonazis (175K), as well as historical events (holocaust) or possible future events (apocalypse) (169K). By manually inspecting messages referring to the holocaust, we find that QAnon followers call the holocaust a hoax and have a holocaust denial approach to this specific topic. Another popular topic in our QAnon dataset is the COVID-19 pandemic and the debate around vaccines (159K messages). Again, we inspect some messages on this topic. We find that QAnon followers have a strong anti-vax ideology and share a lot of false information about this subject. Some examples include messages claiming that COVID-19 vaccines make people sterile, that vaccines are a lie and a fraud, and that people with medical professions are refusing to get vaccinated because they know vaccines do not work. Also, we find several messages pointing out that the COVID-19 pandemic is a plan of Bill Gates to reduce the earth's population. Also, we find a topic related to sharing videos to disseminate QAnon ideology (119K messages), highlighting that videos play an integral role in QAnon. The rest of the topics are related to cryptic messages about Information Warfare (116K), a topic that alleges that politicians are actually Satanists (105K), and a topic for disseminating news about QAnon or Q drops (97K). Our results confirm and reinforce anecdotal evidence presented by Scott [41] highlighting that QAnon follows an anti-vax ideology and that they treat world politicians as arch enemies (e.g., by claiming they are Satanists).

Looking at the popularity of these topics over time (see Fig. 5), we find that before February 2020, QAnon discussions are mainly related to Politics, with almost 50% of the messages being on that topic. After February 2020, we observed that the popularity of the Politics topic decreases (below 20% of all messages shared per week). We observe the insurgence and the popularity of other topics like the Reactions to COVID-19/vaccines topic and the Environment/Masks topic. The increase in popularity of the topic reaction likely indicates an increase in users' engagement with QAnon-related messages. Additionally, we observe that topics that emerged after February 2020 are long-lasting as they have a considerable percentage of all weekly messages during the whole time period until the end of our dataset. Overall, these results highlight that QAnon's discussions are evolving over time and that nowadays, QAnon is not only related to Politics, rather QAnon followers discuss a wide variety of topics that can be weaponized for spreading potentially false or harmful information (e.g., false information on vaccines and the COVID-19 pandemic).

Finally, we look into the languages of the messages in each topic to understand if topics are specific to one language and quantify how popular these topics are in each language. Fig. 6 shows the percentage of messages that are assigned to each topic and each language (e.g., 60% of the messages in the COVID-19/Vaccines topics are shared in German, see Fig. 6(b)). Unsurprisingly, the most popular languages in almost all topics are German and English, mainly because of their popularity in our QAnon dataset. In the Politics topic, we observe similar popularity between German and English, with 45% and 38% of all Politics messages. 7% of Politics messages

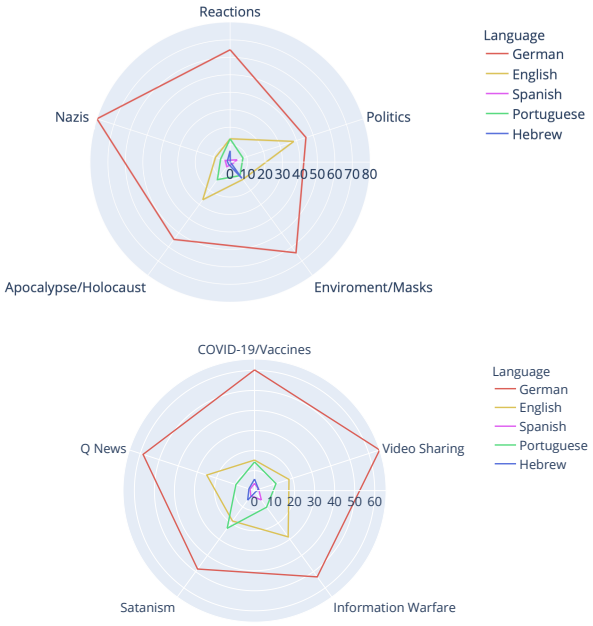


Figure 6: Distribution of topic-specific messages across languages. We report the percentage of topic-specific messages to each language.

are shared in Portuguese, while for the messages in Hebrew, we find that they rarely talk about Politics (only 0.1% of all Politics messages are in Hebrew). For the COVID-19/Vaccines topics, we find that Portuguese and English have similar popularity, highlighting that there is likely a lot of false information disseminated in Portuguese related to the pandemic in QAnon groups (based on our manual examinations, we find a lot of false information in that specific topic). In summary, our language-specific distributions in Fig. 6 indicate that most of the topics are not specific to one language. Rather they are discussed across many QAnon groups/channels and, more importantly, across many languages.

Remarks. Our topic modeling analysis confirms our third hypothesis. QAnon followers on Telegram share and discuss various topics, and they disseminate conspiratorial or false information about Politics and the COVID-19 pandemic. Also, our analysis shows that the QAnon discourse is becoming more diverse, with the Politics topic losing popularity after February 2020 and other topics like the COVID-19 pandemic gaining a substantial share of the discussions. Also, most of the topics are not specific to one language, but rather they span across multiple languages.

5 DISCUSSION & CONCLUSION

In this work, we performed the first multilingual analysis of QAnon content on Telegram. We joined 161 groups/channels on Telegram and collected a total of 4.4M messages shared over 18 months. Using Perspective API and multilingual topic modeling, we shed light on how the QAnon conspiracy theory evolved and became a global phenomenon through Telegram. Our analysis shows that QAnon content on Telegram is increasing in volume (during 2021, a 5x increase in terms of messages). The number of active groups is increasing in both QAnon and baseline datasets during the time of collecting the group URLs from Twitter and Facebook. Unlike the baseline dataset, surprisingly, we observe that the number of active groups/channels in the QAnon dataset has been increasing until March 2021. This indicates that QAnon groups/channels are comparatively long-lasting and active. Also, a considerable increase in the percentage of messages and senders in early 2021 implies that QAnon groups/channels have a stronger reaction to events in the real world. An implication of this increased activity is the need for real-time monitoring tools that can help us track the spread of QAnon content in messaging platforms, similar to systems developed by Melo et al. [26]. This kind of system would at least allow journalists and public authorities to counter misinformation campaigns that are designed to target radical groups.

Our toxicity analysis compounds the findings from Planck [34], which indicates that QAnon is sharing a lot of toxic and violent messages. Our analysis paints a nuanced overview of the toxicity of QAnon across multiple languages and highlights that there are substantial differences across languages. Our results and toxicity validation have several implications for researchers focusing on QAnon or hate speech. First, our results show that QAnon content in languages like German and Portuguese are substantially more toxic than English content, emphasizing the need to study this problem through the lens of languages other than English. Second, in contrast with the findings from Papasavva et al. [30], we find QAnon content being more toxic compared to the baseline for English and Portuguese, which shows the differences that exist across platforms and time. In addition, QAnon followers are likely becoming more toxic over time, particularly after multiple moderation interventions (i.e., bans) from mainstream platforms like Reddit, Facebook, and YouTube. Indeed Ribeiro et al. [39] show that moderation interventions on Reddit may lead to increasing radicalization signals after users migrate to other platforms. Third, our toxicity validation highlights that models like the Perspective API perform differently across languages. This prompts the need to further study the performance of these models across languages and investigate ways to improve their multilingual aspect.

Our topic modeling analysis reinforces findings from previous work [27, 30] and complements these previous efforts by investigating the same phenomenon on Telegram. We showed that QAnon on Telegram is becoming more diverse in terms of their discussed topics. Also, we found messages that were sharing false information across multiple languages, particularly related to the COVID-19 pandemic and international politics. This emphasizes the emerging problem of spreading multilingual false information and the challenges in detecting and tackling it. Our work highlights the need to create organizations that aim to check facts and tackle the spread of QAnon-related false information across languages and countries (e.g., efforts similar to the #CoronaVirusFacts Alliance focusing on the COVID-19 pandemic [35]).

ACKNOWLEDGMENTS

This work was partially funded by FAPEMIG, FAPESP, and CNPq.

REFERENCES

[1] ADL. 2019. Telegram: The Latest Safe Haven for White Supremacists. https://www.adl.org/blog/telegram-the-latest-safe-haven-for-white-supremacists.

[2] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. 2021. An Early Look at the Parler Online Social Network. In *ICWSM*.

[3] Dimo Angelov. 2020. Top2vec: Distributed representations of topics. *arXiv:2008.09470* (2020).

[4] BBC. 2020. Facebook bans QAnon conspiracy theory accounts across all platforms. https://bbc.in/3hqARxH.

[5] BBC. 2021. Twitter suspends 70,000 accounts linked to QAnon. https://bbc.in/2RlDyGn.

[6] Lois Beckett. 2020. QAnon: a timeline of violence linked to the conspiracy theory. https://bit.ly/3Cb9a3i.

[7] Katrin Bennhold. 2020. QAnon Is Thriving in Germany. The Extreme Right Is Delighted. https://nyti.ms/3fh824m.

[8] Brazil Ministry of Health. 2020. Brasil confirma primeiro caso do novo coronavirus. https://bit.ly/2VFiFIh.

[9] Kylar J Chandler. 2020. Where We Go 1 We Go All: A Public Discourse Analysis of QAnon. *McNair Scholars Research Journal* (2020).

[10] Crowdtangle. 2020. Crowdtangle API. https://github.com/CrowdTangle/API/wiki.

[11] Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. 2019. Racial bias in hate speech and abusive language detection datasets. In *Third Workshop on Abusive Language Online*.

[12] Ej Dickson. 2021. The QAnon Community Is in Crisis - But On Telegram, It's Also Growing. https://bit.ly/3k6752e.

[13] Kristen Engel, Yiqing Hua, Taixiang Zeng, and Mor Naaman. 2022. Characterizing Reddit Participation of Users Who Engage in the QAnon Conspiracy Theories. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–22.

[14] Emilio Ferrara, Herbert Chang, Emily Chen, Goran Muric, and Jaimin Patel. 2020. Characterizing social media manipulation in the 2020 US presidential election. *First Monday* (2020).

[15] First Draft. 2020. https://firstdraftnews.org/.

[16] Amanda Garry, Samantha Walther, Rukaya Rukaya, and Ayan Mohammed. 2021. QAnon Conspiracy Theory: Examining its Evolution and Mechanisms of Radicalization. *Journal for Deradicalization* (2021).

[17] Sam Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A Smith. 2020. Realtoxicityprompts: Evaluating neural toxic degeneration in language models. In *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[18] David Gilbert. 2021. QAnon Led the Storming of the US Capitol. https://bit.ly/3k8I7iP.

[19] Matthew Hannah. 2021. QAnon and the information dark age. *First Monday* (2021).

[20] Mohamad Hoseini, Philipe Melo, Manoel Júnior, Fabrício Benevenuto, Balakrishnan Chandrasekaran, Anja Feldmann, and Savvas Zannettou. 2020. Demystifying the Messaging Platforms' Ecosystem Through the Lens of Twitter. In *The 2020 Internet Measurement Conference (IMC)*. 345–359.

[21] Klaus Krippendorff. 2011. Computing Krippendorff's alpha-reliability.

[22] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[23] Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. 2018. Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. In *The international conference series on Advances in Social Network Analysis and Mining (ASONAM)*.

[24] Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* (2017).

[25] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426* (2018).

[26] Philipe Melo, Johnnatan Messias, Gustavo Resende, Kiran Garimella, Jussara Almeida, and Fabrício Benevenuto. 2019. Whatsapp monitor: A fact-checking system for whatsapp. In *The International AAAI Conference on Web and Social Media (ICWSM)*, Vol. 13. 676–677.

[27] Daniel Taninecz Miller. 2021. Characterizing QAnon: Analysis of YouTube comments presents new conclusions about a popular conservative conspiracy. *First Monday* (2021).

[28] Tina Nguyen. 2020. Trump isn't secretly winking at QAnon. He's retweeting its followers. https://politi.co/3z8NYJo.

[29] Antonis Papasavva, Max Aliapoulios, Cameron Ballard, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Jeremy Blackburn. 2022. The Gospel According to Q: Understanding the QAnon Conspiracy from the Perspective of Canonical Information. In *ICWSM*.

[30] Antonis Papasavva, Jeremy Blackburn, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro. 2021. "Is it a Qoincidence?: An Exploratory Study of QAnon on Voat. In *The Web Conference*.

[31] Irene V Pasquetto, Alberto F Olivieri, Luca Tacchetti, Gianni Riotta, and Alessandra Spada. 2022. Disinformation as Infrastructure: Making and maintaining the QAnon conspiracy on Italian digital media. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–31.

[32] Perspective API. 2018. https://www.perspectiveapi.com/.

[33] Shruti Phadke, Mattia Samory, and Tanushree Mitra. 2021. Characterizing social imaginaries and self-disclosures of dissonance in online conspiracy discussion communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–35.

[34] Samuel Planck. 2020. Where We Go One, We Go All: QAnon and Violent Rhetoric on Twitter. *Locus: The Seton Hall Journal of Undergraduate Research* 3, 1 (2020), 11.

[35] Poynter. 2021. Fighting the Infodemic: The #CoronaVirusFacts Alliance. https://www.poynter.org/coronavirusfactsalliance/.

[36] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Conference on Empirical Methods in Natural Language Processing*.

[37] Julio CS Reis, Philipe Melo, Kiran Garimella, Jussara M Almeida, Dean Eckles, and Fabrício Benevenuto. 2020. A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *The International AAAI Conference on Web and Social Media (ICWSM)*.

[38] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis) information dissemination in WhatsApp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference (WWW)*. 818–828.

[39] Manoel Horta Ribeiro, Shagun Jhaver, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Robert West. 2021. Does Platform Migration Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. In *The ACM Conference on Computer Supported Cooperative Work (CSCW)*.

[40] Caitlin M Rivers and Bryan L Lewis. 2014. Ethical research standards in a world of big data. *F1000Research* 3 (2014).

[41] Mark Scott. 2020. QAnon goes European. https://www.politico.eu/article/qanon-europe-coronavirus-protests/.

[42] Ari Sen and Brandy Zadrozny. 2020. QAnon groups have millions of members on Facebook, documents show. https://nbcnews.to/33D9GI0.

[43] Faiz Siddiqui and Susan Svrluga. 2016. N.C. man told police he went to D.C. pizzeria with gun to investigate conspiracy theory. https://wapo.st/3bsMws5.

[44] Andrea Sipka, Aniko Hannak, and Aleksandra Urman. 2022. Comparing the Language of QAnon-related content on Parler, Gab, and Twitter. In *14th ACM Web Science Conference 2022*. 411–421.

[45] Statista. 2020. Number of monthly active Telegram users worldwide from March 2014 to April 2020. https://www.statista.com/statistics/234038/telegram-messenger-mau-users/.

[46] Statista. 2021. Downloads of Telegram. https://www.statista.com/statistics/1260684/telegram-global-downloads-by-region/.

[47] Rebecca Tan. 2017. Terrorists' love for Telegram, explained. https://bit.ly/2XmxOz6.

[48] Telegram. 2020. Telegram API. https://core.telegram.org/method/channels.joinChannel.

[49] Telethon. 2017. Telethon's Documentation. https://docs.telethon.dev/en/latest/.

[50] The Philadelphia Inquirer. 2020. FBI calls QAnon a domestic terrorist threat. https://bit.ly/3nx7vkf.

[51] Julia Carrie Wong. 2018. What is QAnon? Explaining the bizarre rightwing conspiracy theory. https://bit.ly/3tCCzA8.

[52] Brandy Zadrozny and Ben Collins. 2018. Reddit bans Qanon subreddits after months of violent threats. https://nbcnews.to/3wgGOBR.

[53] Brandy Zadrozny and Ben Collins. 2020. YouTube bans QAnon. https://nbcnews.to/3ybJC4H.

[54] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. 2018. What is gab: A bastion of free speech or an alt-right echo chamber. In *The Web Conference 2018*. 1007–1014.

A VALIDATION OF THE PERSPECTIVE API

Given that the Perspective API is essentially a black box, it is important to assess its performance in our dataset, and more importantly, how well it performs across multiple languages. To do this, we extracted random samples of messages from our QAnon dataset in English, German, and Portuguese. Then we performed annotation on each message to determine whether it was toxic or not. We focus on these three languages as they are the most popular in our dataset. We extracted a random sample of 500 messages for each

Table 4: Performance evaluation metrics for the Perspective API's Severe Toxicity model in English, German, and Portuguese. We report the Precision, Recall, and F1 Scores for varying Perspective Severe Toxicity thresholds.

	English			German			Portuguese		
Thresh.	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
0.50	0.475	0.906	0.623	0.317	0.917	0.471	0.506	0.799	0.620
0.55	0.522	0.858	0.649	0.326	0.881	0.476	0.577	0.753	0.654
0.60	0.562	0.858	0.679	0.363	0.821	0.504	0.580	0.753	0.655
0.65	0.617	0.811	0.701	0.400	0.762	0.525	**0.606**	**0.740**	**0.667**
0.70	**0.686**	**0.740**	**0.712**	0.474	0.643	0.545	0.716	0.506	0.593
0.75	0.717	0.677	0.696	**0.491**	**0.643**	**0.557**	0.735	0.487	0.586
0.80	0.753	0.551	0.636	0.510	0.583	0.544	0.740	0.481	0.583
0.85	0.833	0.394	0.535	0.597	0.440	0.507	0.845	0.318	0.462
0.90	0.920	0.181	0.303	0.684	0.310	0.426	0.919	0.221	0.356

language while ensuring that our random sample covers the entire score range from Perspective API. We extracted 50 random messages that had a score between 0 and 0.1, 50 messages from 0.1 and 0.2, and so on. Then, we recruited three annotators (Ph.D. students or researchers) for each language; for English, the annotators were fluent in English, while for Portuguese and German, we recruited native speakers. The annotators were provided with the following definition of toxicity: "We define toxicity as a rude, disrespectful, or unreasonable comment that is likely to make someone leave a discussion" (obtained from Perspective API's website), and were asked to independently annotate each message as toxic or not (the annotators were unable to see the actual Perspective score, they only had access to the comment itself). Then, to obtain our ground truth, we annotated each message as toxic or not based on the majority agreement of the three annotators. We also calculated the inter-annotator agreement using Krippendorff's alpha coefficient [21]; we find 0.41, 0.43, and 0.44 for English, Portuguese, and German, respectively. The coefficient values ranging from 0.41 to 0.60 indicate that the annotators had a moderate agreement [22] across languages and highlight the subjectivity when people annotate content as toxic or not.

Then, to assess the performance of the Perspective API and select an appropriate threshold for each language (i.e., any message that has a Perspective score above the threshold is considered toxic), we varied the threshold and calculated standard performance metrics like precision, recall, and F1 score (see Table 4). Based on our validation results and performance metrics, we treat a message as toxic if it has a score over 0.7 for English, over 0.75 for German, and over 0.65 for Portuguese (thresholds with the largest F1 score, see Table 4). Also, our validation results show that the Perspective API does not perform the same across languages; English is the best-performing language (0.712 F1 score), followed by Portuguese (0.667 F1 score), and German (0.557 F1 score). Future work should further validate the performance of the Perspective API on a larger scale and across multiple languages/datasets.