Next generation internet technology topics on social media based on human values

D1.8: Final social media analysis report & visualisations

Work package	WP1: Topic Identification
Task	1.5 Public debate mapping
Due date	31/08/2021
Submission date	31/08/2021
Deliverable lead	DATALAB, Aarhus University
Dissemination level	Public
Nature	Report
Authors	Ida A. Nissen, Marie D. Mortensen, Maris Sala, Jessica G. Walter, Marina Charquero-Ballester, Mathias H. Sørensen, Kristoffer L. Nielbo (CHCAA), Anja Bechmann (all DATALAB)
Version	1
Reviewers	Egle Juospaityte, Katja Bego, Alberto Cottica, Kristóf Gyódi, Michał Paliński
Status	Final

Disclaimer: The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained herein.

Acknowledgement: This Report is part of a project that has received funding from the **European Union's Horizon 2020 research and innovation programme under grant agreement Nº825652.** We thank Peter B. Vahlstrup and Henrik Pedersen for the help with retrieving data and training the neural network.

Executive Summary

This report focuses on the intersection of internet technology and social issues. We aim to understand how the general public perceives this subject. We explore this question through social media analysis, looking at data gathered from Reddit, Facebook and Twitter and apply a selection of methods. The main goal was to analyze discussions on several social media platforms to identify trends and topics relevant for the next generation internet. A secondary goal was to examine social issues that accompany internet technology.

Human rights on the internet

At the core of the report are human rights on the internet. We used a UN-based document of ten statements of human rights and internet principles[1], which serve as the starting point for our different analyses. For example, the first statement is about ‘universality and equality’, which states the human right of being free and equal on the internet. Using a social starting point allows us to extend our analysis beyond purely tech-based discussions and enables us to map the areas where internet technology affects people’s lives. To achieve the vision of NGI for a more human-centric and democratic internet, we need to take the effects of internet technology on society into account. We performed five separate analyses, where the first analysis constitutes the main part of the report. This main part includes all ten human rights statements and identifies trends in discussions originating from the statements. The four smaller stand-alone parts of the report contain deep dives. The first deep dive analyzes topics relevant for the next generation internet, and maps the public perception of those discussions. The last three deep dives delve into social issues related to internet technology, and are related to one of the ten human rights statements. The five different analyses of the report are summarized below.

Trend detection in internet technology

The main analysis is about **detecting upcoming trends** in internet technology. We analyzed technological and social discussions with a basis in human rights. The research question for the main part was which topics are trending that are related to human rights on the internet? We found that the trending topics expressed concern for privacy and internet security, contained a movement against censorship and the power of mainstream internet services, and consisted of a technology-interested group discussing blockchain technology and hacking. Those topics can guide the choices to be made towards the next generation internet, which prioritizes citizens and their needs.

Mapping NGI-related discussions

The four deep dives analyze NGI-related topics and societal issues. The first deep dive analyzes **NGI-related discussions** on social media. This time the starting point were selected Twitter hashtags relevant for NGI. The goal was to identify the discussed topics, describe the emotions in the discussions, and map out the larger network of related topics. We found that most hashtags contained topics that were discussed as business opportunities and topics that raised concerns about data privacy and security. Particularly

[1] http://internetrightsandprinciples.org/ site/campaign

the discussions about privacy expressed negative emotions, with fearful and angry tweets. The hashtag 'AI' was the most mentioned by its co-hashtags, and many hashtags were closely related to the hashtags #artificialintelligence, #bigdata and #machinelearning. These findings corroborate the outcome of the main part in that privacy and security are key topics and a cause of concern and anxiety in the public. Furthermore, artificial intelligence is broadly present and will probably be a key player in the future internet.

Gender bias in algorithms

The last three deep dives delve into social issues related to internet technology. The first examined issue was about **gender inequality** and algorithmic bias. This deep dive relates to several human right statements: 'universality and equality' (the right of equality online), 'diversity' (cultural and linguistic diversity online), and 'network equality' (internet access, no content discrimination or filtering or traffic control). As such, internet technology such as algorithms should also support equality online. In this analysis, our goal was to train an algorithm to classify Facebook images based on whether they were uploaded by women or men. Gender equality was chosen as an example of inequality, as there is much focus on this topic and gender discrimination affects large groups of the population. We found that generally, images uploaded by women were related to social behavior and images uploaded by men were oriented towards action and objects. However, some images uploaded by women contained similar subjects as those uploaded by men (and vice versa), and were consequently falsely attributed to the other gender category. These results show that although gender patterns are discernable, a part of the users fall outside the general patterns, which might result in a discrimination bias. Such discrimination biases should be accounted for during the development of classification algorithms to decrease gender inequality at the hand of algorithms. Since algorithms are developed constantly and will be part of the future internet structure, the NGI project should have a focus on decreasing discrimination biases in algorithms to ensure an internet that is inclusive for minorities or discriminated groups.

Disinformation and evoked emotions

The next deep dive was into the societal issue of **disinformation** on the internet. It relates to the human right statement 'expression and association', which stresses the right to receive and communicate information online. It also touches on another human right statement, namely 'diversity', which advocates cultural and linguistic diversity online. Here, our goal was to analyze disinformation on social media and the evoked sentiments. We focused on disinformation around COVID-19, and analyzed the emotions of the discussions on Twitter. The results showed that disinformation in general did not lean towards a positive or negative emotional valence, but that certain types of disinformation did. Those related to conspiracy and virus characteristics had a stronger negative valence in their discussions than other types of disinformation. Emotional context affects the spreading of disinformation, and knowledge about the sentiments of disinformation types will help to better understand and possibly curb the spreading and social consequences of disinformation.

Privacy on social media

Lastly, we took a deep dive into **privacy** in relation to social media. This analysis is connected to the human right statement ‘privacy and data protection’, which states the right to privacy online. In this deep dive, we analyzed discussions in Facebook groups for their relation between privacy settings, gender and discussed topics. Our research question was whether privacy settings of groups (open, closed or secret) could be predicted from the topics discussed in the groups. We found that the topics could not predict the privacy settings, meaning that a group set to be private does not necessarily contain private topics. This has potential implications for data privacy and indicates that the content of Facebook groups should be better protected.

Table of Content

1 Introduction

- 1.1 Purpose and scope

2 Trend detection of internet technology based on human value related discussions on Reddit

- 2.1 Background, aim and research question
- 2.2 Human value seeding list
- 2.3 Reddit dataset
  - 2.3.1 Identifying subreddits through the human value seeding list
  - 2.3.2 Discussions on Reddit
- 2.4 Analysis
  - 2.4.1 Trend reservoir model
  - 2.4.2 Text preprocessing
  - 2.4.3 Topic modeling
  - 2.4.4 Novelty, transience and resonance
  - 2.4.5 Identifying trending topics
- 2.5 Results - Trending topics
- 2.6 Discussion and conclusion

3 Deep dive into Twitter dataset of several NGI-related hashtags

- 3. 1 Background, research questions and hypotheses
- 3.2 Twitter dataset
- 3.3 Analysis
  - 3.3.1 Identifying topics relevant for NGI - Topic analysis
  - 3.3.2 Text preprocessing
  - 3.3.3 Selection of amount of topics and LDA
  - 3.3.5 Emotionally loaded discussions
  - 3.3.8 The network of topics surrounding NGI relevant topics - Co-hashtag networks
- 3.4 Results
  - 3.4.1 Topic analysis
  - 3.4.2 Emotion analysis
  - 3.4.3 Network analysis
- 3.5 Discussion
- 3.6 Conclusion

4 Deep dive into gender inequality

- 4.1 Background, research questions and hypotheses
- 4.2 Facebook dataset
- 4.3 Analysis
  - 4.3.1 Convolutional neural networks
  - 4.3.2 Data preprocessing

4.3.3 Data analysis
4.4 Results
4.4.1 Prediction of gender based on uploaded images
4.4.2 Detected gender patterns in the uploaded images
4.4.3 Falsely assumed gender patterns
4.5 Discussion and conclusion

5 Deep dive into disinformation – sentiments of COVID-19 misinformation on Twitter
5.1 Background, research questions and hypotheses
5.2 Misinformation dataset
5.2.1 Misinformation stories on Google Fact Check Explorer
5.2.2 Twitter sample
5.2.3 Selecting tweets related to misinformation stories
5.3 Analysis
5.3.1 Types of misinformation
5.3.2 Sentiment analysis
5.4 Results
5.4.1 Overview of tweets related to misinformation
5.4.2 Is the discussion about misinformation related to emotion?
5.4.3 What are the sentiments in the different topics?
5.5 Discussion and conclusion

6 Deep dive into privacy
6.1 Background, research questions and hypotheses
6.2 Theory
6.2 Methods
6.2.1 Dataset
6.2.2 Analysis
6.3 Results
6.3 Discussion and conclusion

7 Conclusion

8 References

9 Appendices
9.1 Appendix for Trend detection
9.2 Appendix for deep dive into NGI-related hashtags
9.3 Appendix for deep dive into disinformation

1 Introduction

Human rights on the internet are valuable guidelines for online behavior, internet technology, and governmental regulations. They are also an integral part of the goals of next generation internet (NGI), envisioning a more human-centric internet of the future. We therefore aimed to find an official document describing these values and principles to use as the basis that connects the European values to the development of technologies and, specifically, to the internet. We found that the most appropriate document for our goal was the 'Charter of Human Rights and Internet Principles'[2] (launched in 2011 and revisited in 2013), published by the Internet Rights & Principles Coalition (IRPC), based at the UN Internet Governance Forum (see Appendix 9.1.1). We chose this document given its specificity and inclusivity of the above-mentioned values. Furthermore, its focus on Human Rights and therefore, its universality, makes it appropriate for extension into other contexts outside of the European Union. The IRPC coalition is an international network of individuals and organizations working to promote human rights in the online environment and across the spectrum of internet policy-making, aligning with the NGI goals. Its participants come from varied backgrounds, including individuals from grassroots groups, international NGOs, researchers, activists, lawyers, businesses, internet and mobile phone service providers, technical communities, government representatives, and intergovernmental organizations. Those ten key rights and principles of human rights online form the basis of our different analysis of social media datasets.

Social media platforms provide an online space for people to connect and communicate. They enable people to discuss a diverse range of topics from everyday life to particular hobbies to global issues. Often, the latest news and trends are discussed without time-delay on such platforms, which makes them ideal for analyzing up-to-date discussions of topics of interest. Our topic of interest is emerging internet technology and the accompanying social and economic issues. By analyzing the discussions on different social media platforms, we can identify key issues and technologies and map the discussions around those topics.

For our analysis, we focused on three social media platforms: Reddit, Twitter, and Facebook, to research different aspects of human rights on the internet. By looking at discussions on different platforms, we include a larger group of users. For the different specific analyses, we chose the most suited platform. Reddit originated as a technology-based discussion forum, but it encompasses broader discussions today even though it remains a base for technology-interested people. It consists of topic-based communities named 'subreddits', which often attract members that are experts in that particular topic (Horne et al., 2017). This makes Reddit ideal for topical analysis of online discussions, and it is an optimal platform for assessing the latest discussions on internet technology. The research question with the Reddit data is to find which topics are trending that are related to human rights on the internet.

[2] See http://internetrightsandprinciples.org/ site/campaign

Twitter is a social networking platform based on microblogging in the form of short messages called 'tweets'. Users can post and retweet tweets, and assign specific topic labels to them called 'hashtags'. The hashtag system allows for a specific selection of topics to analyze. The focus of Twitter is the dissemination of information (Kwak et al., 2010) and news are often discussed fast, which allows for an analysis of up-to-date discussions.Twitter data was used to map discussions on topics that are relevant for NGI.

Facebook is a social networking service aimed at connecting people online. Users can send private messages to each other, be part of a group, or post messages viewable to friends or the public. People can also post images on their profile site, thereby selecting and showing pictures to others. Groups are centered around a topic or a common connection, and can be open or closed and only for members. The profile sites can be analyzed for the content posted, and the groups can be selected for different topics or settings. We analyzed Facebook images uploaded by users as well as privacy settings of groups to investigate discrimination bias by algorithms and the protection of privacy online.

Internet technology is a fast-changing field, with new technologies continuously coming up but only a few sticking. We are interested in the technologies that will be trending in the future (ten years from now), which might be small now but upcoming. It is challenging to identify emerging trends using data from the present, and instead of using standard trend detection methods, we apply a model that detects trend reservoirs (topics with trend potential). This model does not assume a 'spiky' behavior of trends, but identifies trends based on their novel content in relation to how sticky this content is. With this model, smaller and upcoming trends can also be identified.

Social media sources offer a vast material of discussions, emotions and opinions. A suitable analysis of those sources is topic modeling, which analyzes the semantics of the text and groups the discussions into topics. Those topics provide an overview of the content of discussions, which is advantageous when the data is large and manual analysis would be too time consuming. Here, we map out the topics of various discussions to gain an insight into which topics are discussed most frequently. Those topics provide a basis for gaining insights into the current discussions on social media.

Trend detection allows for an understanding of which topics of discussion have a higher importance over time. It goes beyond simple topic modeling of texts, which does not allow a ranking of which topics people find most relevant and important to discuss. Further, simple topic models lack time-dimensionality. By taking the time of the discussions into account, we can identify which topics are coming, staying or leaving. To inform NGI about the topics that will be relevant for the internet in the future, we do not simply want to find the topics discussed today but the topics that are growing and staying. This task can be fulfilled with the model we applied for identifying trends, which takes the unfolding of topics over time into account.

1.1 Purpose and scope

This report encompasses the final analysis on social media datasets with a focus on internet technology and human rights. We extend on the intermediary analysis of the deliverable D1.7, which applied a trend detection model on discussions on Reddit around artificial intelligence. This extended analysis with a new basis on human rights constitutes the main part of the report, and several deep dives into different focus areas constitute stand-alone smaller parts. The deep dives were chosen to highlight several of the ten key rights and principles of human rights online. They each study a specific case at the interface of human rights and internet technology.

The main aim of this report was to analyze discussions on several social media platforms to identify trends and topics relevant for the next generation internet. A secondary goal was to examine social issues that accompany internet technology. We have analyzed various data sources from the platforms Reddit, Twitter, and Facebook. They have a large number of users and encompass discussions on internet technology, societal issues, news, and everyday life.

In this report, we present our main analysis about detecting upcoming trends of internet technology. The analysis centers on discussions based on human values to incorporate societal aspects instead of a purely technological aspect. Additionally, we describe several deep dives into internet technology related topics and societal issues. The first deep dive is into NGI-related topics, where we identify discussed topics, investigate the emotions of those discussions, and map out the related topics. The second deep dive investigates the societal issue of gender inequality in relation to the discriminatory bias of algorithms. For the third deep dive, we look into the issue of disinformation on social media and the attached sentiments. The fourth deep dive is into privacy and whether topics determine privacy settings. Those deep dives give insight into both current discussions as well as societal issues accompanying internet technology.

2 Trend detection of internet technology based on human value related discussions on Reddit

This section constitutes the main part of the report and is in preparation for being submitted for publication in an academic journal.

2.1 Background, aim and research question

The overall mission of the Next Generation Internet (NGI) initiative is to shape the future internet as an ecosystem that embodies European values. In a broad sense, these include the protection of human dignity, freedom, democracy, equality, rule of law and human rights. Other values more specific to the internet ecosystem refer to openness, inclusivity, transparency, privacy, cooperation, and protection of data. Our lives are becoming increasingly digital, especially in the midst of the pandemic where societies rely heavily on interactions through the internet. Therefore, from a societal point of view, the key European values are closely tied to the operationalization of them technologically. We used the ten key rights and principles for the internet (see Introduction) as a starting point to go beyond a tech-centered approach and include social sustainability aspects into internet technology.

Our next aim is to identify the trends in those technological and social discussions. We apply a trend reservoir model that potentially detects not only the main trends but also weaker signals that might turn into future trends. In this way, we avoid identifying only the trends that are already big now, but also map out smaller topics that show the potential to be trending in ten years from now.

As we want to analyze debates on large social media platforms, we have chosen to analyze data over a two-year span from Reddit. Reddit is one of the large social media platforms and started as an online discussion forum for technologies. Today its scope is much broader and includes many social and cultural topics, but it is still one of the main places for technology-related discussions. This combination of social and technological debates makes it the ideal platform to explore for finding topics about human rights and internet technology.

To conclude, our aim was to identify upcoming trends in internet technology with a focus on human rights on the internet. The overarching research question is which topics are trending that are related to human rights on the internet? This is an explorative analysis and we will discuss the identified trends and set them into perspective.

2.2 Human value seeding list

We focus on the ten key rights and principles of human rights online (see introduction), launched in 2011 and rooted in international human rights standards. We proceed to distil those into shorter statements containing the right or principle itself accompanied by a specification of the environment (e.g., Freedom and equality online, access to a secure and open internet). Each statement is then divided into keywords, making it as concise as possible (see appendix A9.1.1 for the rights and extracted keywords).

2.3 Reddit dataset

2.3.1 Identifying subreddits through the human value seeding list

Here we used the Human Rights and Internet Principles Seeding List to test its ability to return meaningful subreddits for further analysis. We entered each set of keywords in Reddit's search engine under the section 'Communities and users' including posts from all time. Most recent data (i.e., December 2020) shows that Reddit contains a total of 2.561.393 subreddits or online communities (https://frontpagemetrics.com/history/month). Our search returned a total of 304 subreddits, from which we selected 49 as relevant to the project by examining the subreddit description. The subreddit description ('About Community') is a short explanation on the topic covered by the subreddit. The selection was done manually, excluding subreddits that would fall in any of the following categories: 1) Service provider and/or technical support 2) Gaming online or manga series 3) NSFW 4) Politics and/or news 5) Private or single member 6) Public person. In cases in which the subreddit could not be assigned to any of these categories, the decision was made based on whether the subreddit would fulfil both of the following criteria: 7) discussion concerning digital or internet technologies and 8) discussion related to human rights and values. From the selected 49 subreddits, 12 were re-occurrences, leaving us with a total of 37 subreddits discussing issues related to human rights on the internet. See appendix A9.1.2 for an itemized list of these selected subreddits.

2.3.2 Discussions on Reddit

We collected data from Reddit through the Reddit API, where we collected the posts and corresponding comments of the 37 selected subreddits over a two years period, from 1st of February 2019 until 1st of February 2021. The subreddits with less than 120 posts and comments were excluded from the further analysis for accurate estimation of the topic distribution, and two subreddits were excluded due to analysis errors, which resulted in 17 analyzed subreddits.

2.4 Analysis

2.4.1 Trend reservoir model

It is inherently difficult to predict trends in the future using data of today. Typical standard approaches view trends as a spike in the occurrence of a word (Madani et al., 2014). However, trends might not display spiky behavior, and trends might not be captured by looking at atomic words. The model we applied does not assume the shape of trends and it considers the distribution of topics instead of looking at singular words. To avoid only capturing trends that are already large, we did not require a minimum number of members of the subreddits and only a small minimum number of posts due to estimation of the topic distribution. This allowed smaller topics and communities to be included in the analysis, as they might be large in ten years from now when showing trend potential. The trend reservoir model we used is described below and in more detail in the report D1.7 and in (Nielbo et al., 2019)[3].

2.4.2 Text preprocessing

As the Reddit API scraper collects posts and comments separately, the subreddit's submissions and comments were joined together based on temporal data, such that the comments following a post are following the post in the dataset in the correct order.

The texts were cleaned of URLs, numbers, non-English characters, and underscores from usernames and subsequently lemmatized using SpaCy's (Honnibal & Montani, 2017) large

[3] The model is available at https://github.com/AU-DATALAB/newsFluxus

English model. English stop words were removed, together with pronouns and excess whitespaces, and the lemmas were tokenized using SpaCy. The texts that were empty after cleaning, lemmatization, and tokenizing, were also removed.

There were seven subreddits which were far larger than the rest: they had over 64,000 data points, compared to the rest which had less than 6000 data points. These large subreddits are conspiracy, technology, Bitcoin, privacy, privacytoolsIO, Stellar, and netsec. To ensure these subreddits would give a meaningful signal without resulting in processing errors, they were downsampled based on their timestamp. The downsampling consisted of sorting the posts into time bins of 30 min. The posts belonging to the same time bin were concatenated into the same continuous post. The trend reservoir model is a temporal model and thus downsampling based on timestamp enables to keep all of the textual data without losing much info in terms of the time dimension. Concatenating subreddits together with the comments would have yielded better results for the topic modeling as the concatenated posts would be around the same topics, but since it would mean losing the time dimension of the comments, it was not done. However, the subreddits conspiracy and technology still suffered from processing errors due to their large size.

2.4.3 Topic modeling

The trend estimation is based on the distributions of topics of each document (posts and comments). Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a popular method for topic modeling and it was used to obtain the topic distributions of each post. To choose the optimal number of topics for each subreddit, we tested the topic numbers 20, 30, 50 and 80 and chose the topic number with the largest coherence score. A large coherence score indicates that the top words associated with the topics are semantically close to each other. Consequently, the optimal number of topics varied per subreddit.

The downsampling resulted in longer posts per time point, and thus more topics could be recovered in each post. For the downsampled subreddits, we therefore tested the smaller range of topics together with a larger range - the coherence scores were consistently higher for the higher numbers of topics. The topic tune range used for the seven large subreddits was 100, 500, 1000, 1500 and 3000. As more text was used for the topic modeling in the downsampled subreddits compared to the other subreddits, this might have improved the topic model results for the large subreddits. However, since we extracted the topics for each subreddit separately and did not compare the performance among the subreddits, this did not influence our results negatively. Besides, the posts and comments also varied in length, regardless of downsampling.

2.4.4 Novelty, transience and resonance

We use a model that predicts trends reservoirs, which are texts that have the potential to become trends in the future. The trend potential of each subreddit is based on the **novelty**, **transience** and **resonance** over time, where the subreddit's posts are divided into time windows spanning three consecutive posts. **Novelty** reflects how new the content is at a given time window compared to earlier time windows. For the calculation, the topic distribution of a given time window is compared to the topic distribution of the previous time window. The difference is measured using a concept from information theory called Jensen-Shannon divergence, which is a measure for how much distributions differ. A large difference means that many new topics have been introduced, which results in a high novelty. Similarly, **transience** reflects how new the content is at a given time window, but compared to later time windows. Here, the topic distribution of a time window is compared to the topic distribution of the subsequent time window. A high transience means that topics change and do not stick, they are only transient. Finally, **resonance** indicates to which extent topics are both novel and sticky. It is the difference between novelty and transience, and a high resonance is achieved by a high novelty and low transience.

We estimate the trend potential of a subreddit using the slope of resonance upon novelty, which is an indicator of trend reservoirs. For this, resonance is plotted as a function of novelty and the linear slope coefficient is estimated, hereafter called '**novelty-resonance slope**'. It has a positive value when topics with a high novelty have a low transience, and topics with a low novelty have a high transience. The larger the novelty-resonance slope, the more novel and sticky the content of the subreddit.

2.4.5 Identifying trending topics

We aimed to identify the trending topics within each subreddit. For this we looked at the novelty-resonance slope over time and further analyzed the posts with a high slope value. First, we calculated the novelty and resonance over the entire length of the dataset to generate a novelty and a resonance time series. We then calculated the novelty-resonance slope over smaller segments using sliding windows of size 21 (inhouse calculations showed that 21 datapoints was the minimum for reliable results) over the previously calculated novelty and resonance time series to get a time series of the novelty-resonance slope. We selected the posts in the windows with the highest slope values using a threshold of 10%. We then analyzed those posts using topic modelling to retrieve the topics of the most trending posts. The topic modeling was done following the same pipeline as described in 2.4.3.

2.5 Results - Trending topics

Of the 17 analyzed subreddits, the seven largest subreddits were conspiracy, technology, Bitcoin, privacy, privacytoolsIO, Stellar, and netsec. For an overview of the analyzed and excluded subreddits and their size, see Appendix Table A9.1.3. To identify which topics were trending in the different subreddits, we investigated the time segments with high novelty-resonance slope values. Specifically, we selected the top 10% posts within a subreddit with the highest novelty-resonance slope values over time, and performed a topic model on those posts. The results are given below for each of the subreddits together with the number of members (asseded 29th July 2021). The number of topics and top keywords for the three most representative topics for each subreddit are in appendix A9.1.4. For each topic, LDA allows for extracting the top words describing the topic, and the post that has the highest probability to match the topic the best. These posts were used to aid understanding the meaning of the topic words in context.

InformationPolicy, 'Information policy and culture', 1010 members
This subreddit is about information propagation through social, digital, cultural and economic networks. It covers topics such as open access to information, privacy, censorship, intellectual property rights, and culture wars (information community conflicts). The subreddit's content is broad according to the description, but the topic modeling of the most trending posts gives more insight. The first topic contains several words related to being critical about the source or intention. The post that contributed most to the first topic talks about 'cancel culture', which is a culture of excluding a user from social or professional circles. In the other topics, 'twitter' is an often-occurring word alongside other internet services like 'youtube' and 'google'. The second topic relates to the internet in China, with the most contributing post being about 'control societies'. The third topic is about power and large players, namely politics, society, and industry. Overall, the most trending topics within the subreddit express concern about information validity, censorship, and powerful actors on the internet.

Antisocialmedia, 'AntisocialMedia', 870 members
Antisocialmedia is about antisocial behavior on social media, censorship, control of media, cyberbullying, filter bubble, privacy and the psychological effects of social media. Of the trending posts within this subreddit, the first topic is in general about social media platforms. The most related post is about Clout, a social platform that does not censor content, use search manipulation, track or sell user data. The second topic is similarly about social media platforms, games and further unspecific keywords. The most related post is about being censored on Halo Waypoint, which is a portal for the online game Halo. The third topic is about snapchat and private life. The top post expresses concern that the use of snapchat excludes privacy. In summary, the trends are about alternative social media platforms and about issues (censorship and privacy) of the mainstream social media platforms.

CyberSec101, 'Cyber Security - Interviews - hack - Privacy', 2499 members
The subreddit offers a place for cyber security videos with explanations and advice. The topics are hacking, privacy, anonymity, whistleblowing and interviews with industry experts. Looking into the topics within the most trending posts, the first topic is about communication platforms and data collection. It hints at the worry about companies overhearing private conversations online. The second topic discusses the risk of hackers when using devices in China. The most contributing post asks about risks of spyware after using laptops and phones in China, and whether to wipe the devices before going back to the US. The third topic is about getting help from a hacker or computer guru. Hacking or getting access to information is the suffusing topic in this subreddit, be it from companies disregarding privacy laws, spyware from China, or the persons themselves.

Rad_Decentralization, 'Radical Decentralization - The Nature of the Future', 16511 members
The subreddit's scope is to subvert traditional hierarchical systems to achieve a world that is more resilient, innovative, networked, transparent and sustainable. The first topic contains

few specific words, the most specific being 'bitcoin', which is the most well-known decentralized cryptocurrency. The second and third topics are both about decentralized blockchains, which is the technology behind bitcoin. The most contributing post to the third topic mentions advantages and disadvantages of decentralized platforms. To conclude, the most trending topics within decentralization are bitcoin and blockchains.

SmashingSecurity, 'Smashing Security podcast', 1602 members
This subreddit is for discussions of listeners of the podcast 'Smashing Security'. Their topics are cybersecurity, cybercrime, hacking and online privacy. Of the most trending posts, the first topic relates to security of accounts and passwords. The second topic is both about podcast episodes about scams and hackers. The third topic mentions Google and contains words related to innovation and inspiration. Not surprisingly, the keywords 'podcast' and 'episode' re-occurred in the other topics as well. To conclude, hacking in general and hacking of accounts are the main trending topics here.

Privacytools, 'PrivacyTools', 2904 members
The focus of this subreddit is on online privacy and mass surveillance. This subreddit stopped in September 2020 due to overlap with other subreddits and refers to the subreddit privacytoolsIO (also analyzed here). We have 1,5 years of posts of this subreddit in the analysis. Of the most trending posts, the first topic is about accounts, logins and possibly security concerns related to them. The most related post recommends a tool for filling in login information on websites, thereby bypassing to copy paste passwords. The second topic relates to privacy and devices and the third topic is about the use of phones (storage, files, advertisement guards). The trending topics of this subreddit are mostly regarding privacy on online accounts and phones.

FreeAsInFreedom, 'Personal and digital freedom for all', 872 members
The subreddit centers on issues related to privacy and personal freedom, online as well as offline. The first topic of the most trending posts is about facial recognition, police and speech. The top post talks about censorship in speech. The second topic regards everyday life and the top post mentions mainstream appeal and finds centralized social media services undemocratic. The third topic relates to smartphones and forum discussions. This subreddit relates privacy issues to the everyday life of users, where the most trending topics are facial recognition, freedom and censorship of speech, and privacy on smartphone use.

Degoogleyourlife, 'Resources For Minimizing Google and Other Intrusive Conglomerates From Your Life', 2041 members
This subreddit is about freeing oneself of intrusive conglomerates to protect one's online privacy and security. The first topic discusses open source, access, encryption and management. Based on the subreddit's title, it might be about alternatives to Google apps (like its email service). The second topic is about companies like Facebook and Google and opting-out. The top post recommends people to opt-out of those services. The third topic does not contain many specific keywords, but seems to discuss complex issues with pro and con arguments. The writer of the top post expresses no worries about companies

recording his/her mundane communications. Overall, the trending topics here are about alternatives for services of the large companies on the internet.

Snowden, 'Edward Snowden, a grand chap', 16,520 members
The subreddit's title reveals a positive sentiment about Edward Snowden and his role as a whistleblower. Its focus is on discussions about Edward Snowden, the NSA (national security agency), and PRISM (the clandestine surveillance program that the NSA uses for mass data collection of internet communication). Of the trending posts, the first topic is unspecific and about people and probably Edward Snowden (guy, someone). The second topic relates to the US and the third topic is about beliefs, rights, and countries. Those topics are rather unspecific, but the title underlines the importance of not accepting unconsented surveillance.

Netsec, '/r/netsec – Information Security News & Discussion', *419,789 members*
This subreddit provides a forum for technical news and discussions of information security. The first topic is about information and security and mentions two portuguese-speaking countries (Portugal and Brazil). The second topic clearly relates to the job market and changing jobs. A top post thanks another user for a response and appreciates that an unspecified project is free to use. Lastly, the third topic is about attacks and the linux operating system. In summary, online security is the broader topic and it entails knowledge, challenges and job opportunities.

Bitcoin, 'Bitcoin – The Currency of the Internet', *3,225,803 members*
The subreddit is dedicated to bitcoin, which is a decentralized digital currency. For all 20 topics of the trending posts, the word 'transaction' is a keyword and mostly the primary one. The first topic relates to discussions about transferring money, probably bitcoins. Several keywords relate to the conventional monetary system ('cash', 'bank') and to considerations about beginning ('lose', 'sure', 'start'). The second topic, while also about transactions, seems to be more about transactions that took place instead of consideration before making transfers (the first topic). Having similar keywords to the first topic, the third topic is also about transferring money (Coinbase is a platform for exchanging cryptocurrency) and the topic relates to quantities and value. To conclude, all topics discuss the transaction of Bitcoins.

Privacy, 'Privacy & Freedom in the Information Age', *1,194,534 members*
The subreddit focuses on the intersection of technology, privacy and freedom in a digital setting. The first topic of the trending posts relates to work and phone, and the top post asks for recommendation for a subreddit about help for booting the operating system. The second topic centers around google. The top post talks about passwords and playing games. The third topic contains the words 'datum', 'good', and 'user', but is further unspecific. Overall, the topics contain mostly unspecific verbs and do not allow for specific conclusions, but the large number of members underscore the perceived importance of privacy.

Stellar, 'Stellar', *196,682 members*

Stellar is a decentralized network for trading several types of cryptocurrency. The subreddit is for news, announcements and discussions about Stellar. Stellar has its own cryptocurrency called XLM (also called Lumens). The first topic of the most trending posts mentions Warren Buffett, who advocated buying Stellar's cryptocurrency. The topic further contains emotionally loaded words ('meltdown', 'maximalists', 'suckers', 'fearful'). This points in the direction that buying XLM is discussed with emotions and subjective opinions. The second topic contains keywords relating to facebook, cryptocurrency, government, privacy, service and tokens. Even though those words seem unrelated, they show how broad the impact of Stellar is as it is linked to social media platforms, the government and privacy discussions. A top post states the opinion that it will still take five to ten years before most people pay with cryptocurrency in stores. The third topic revolves around the Stellar network, XLM and blockchain. Here, a top post mentions mobile coins in a collaboration between Stellar and the messaging app Signal. Overall, the trending topics are about using cryptocurrency and its economic and societal impact.

ComputerSecurity, 'Computer Security – IT security news, articles and tools', *27,094 members*

The subreddit focuses on IT security. In the topics of the trending posts, the word 'use' occurs often, but also the word 'remove'. The first topic seems to discuss the security of connections regarding passwords. The second topic is about hardware (drive, router). The third topic relates to connections (vpn, server) but it is also about passwords. Overall, the trending posts in this subreddit are concerned with the protection of data, be it in accounts or on hardware.

Iexec, 'Blockchain-Based Decentralized Cloud Computing', 7,730 members

The subreddit focuses on iExec, which is a decentralized cloud computing platform. RLC are the tokens of iExec and can be exchanged for computer resources. In the trending posts of this subreddit, the name of the platform itself ('iexec') occurs in most topics. The first topic centers on technological aspects of iExec and its cryptocurrency. The second topic is about iExec's token RLC, which also features as a keyword in the third topic, but this topic is more centered on the exchange of tokens (iExec's network can be utilized in exchange of the tokens, and btc bet is a website for bitcoin betting). The most related post is about corporations purchasing RLC. Overall, the trending posts are about iExec, it's token RLC, and cryptocurrency in general.

Cyberlaws, 'cyberlaws: Legal News Related To Technology And The Net', 34,156 members

The subreddit is about legal news about technology, for example computer crime, copyright, privacy, free speech, intellectual property, net neutrality, the RIAA (Recording Industry Association of America, which represents and protects the US music industry). Of the most trending posts, the first topic is about bots, work, and other words that are difficult to group together. RuneScape is a fantasy online game, but there is no clear link to the other words. The most associated post to this topic is about copyright infringements, which would fit with the keywords 'right' and 'owner', but are difficult to relate to 'bot' and 'work'. The second topic is associated with phone surveillance, with the keywords 'threat', 'nso' (which is a private spyware that enables remote smartphone surveillance), and 'phone'. The third topic is clearly related to legal issues and criminal cyber activity. The top post is of a computer security expert talking about giving advice to lawyers about technical matters. In conclusion, two clear topics emerged: phone surveillance and legal action against cyber crime.

privacytoolsIO, 'PrivacyTools', 190,751 members. Large subreddit
This subreddit warns about organizations that monitor and record people's online activities. The first sentence in the subreddit's description is 'You are being watched'. Their goal is to provide resources to protect user's privacy against global, mass surveillance. The topics are privacy and security. The first topic is about apps and privacy and contains verbs expressing wishes. Those wishes could be about apps that respect privacy. This fits with one of the top posts that describes the safety of password recovery. The second topic also expresses wishes and the keyword 'vpn'. The top posts relate to storing passwords, whether in memory, on USB, or through Bitwarden (a password manager). The third topic relates to privacy and work. The top posts talk about privacy policies of different apps and software, encryption, and file recovery after deletion. There is a lot of overlap in the keywords of the topics and the word 'use' is the top keyword in all three topics. Altogether, the topics of the trending posts in this subreddit are about a broad range of technology for protecting privacy. Taking into account the large number of members subscribed to this group, the protection of user's privacy seems a very important topic.

Summary of trending topics
The trending topics were qualitatively labelled into overarching topics with subtopics. To summarize the findings (see Figure 1), one of the most occurring trending topic was privacy. A lot of concern about privacy of personal data was expressed, with specific topics being privacy laws, facial recognition, privacy on smartphones, data protection, privacy violations of social media platforms and the integration of privacy in everyday use of digital services.
A related topic is censorship. Specified areas are freedom and censorship of speech, and censorship by mainstream social media platforms. Both of these issues with privacy and censorship (and also security) led to trends searching for alternatives and away from mainstream social media platforms and internet companies with virtual monopoly.
Another trending topic was internet security. Surveillance of smartphones, cybercrime and legal actions against it, spyware, mass surveillance and the whistleblower case of Edward Snowden fall under this topic.
Further, hacking occurred several times as a trending topic. It was trending in a podcast as well as a subreddit about security. Some topics were about getting access to information, hacking accounts, being hacked and spied upon, and hacking software.
A last topic relates to decentralisation, blockchains and cryptocurrency. The cryptocurrency Bitcoin was trending in two subreddits, and two other cryptocurrencies were discussed: RLC of iExec and XLM of Stellar. Discussions about cryptocurrencies showed up as trending as opposed to further discussions about decentralization.

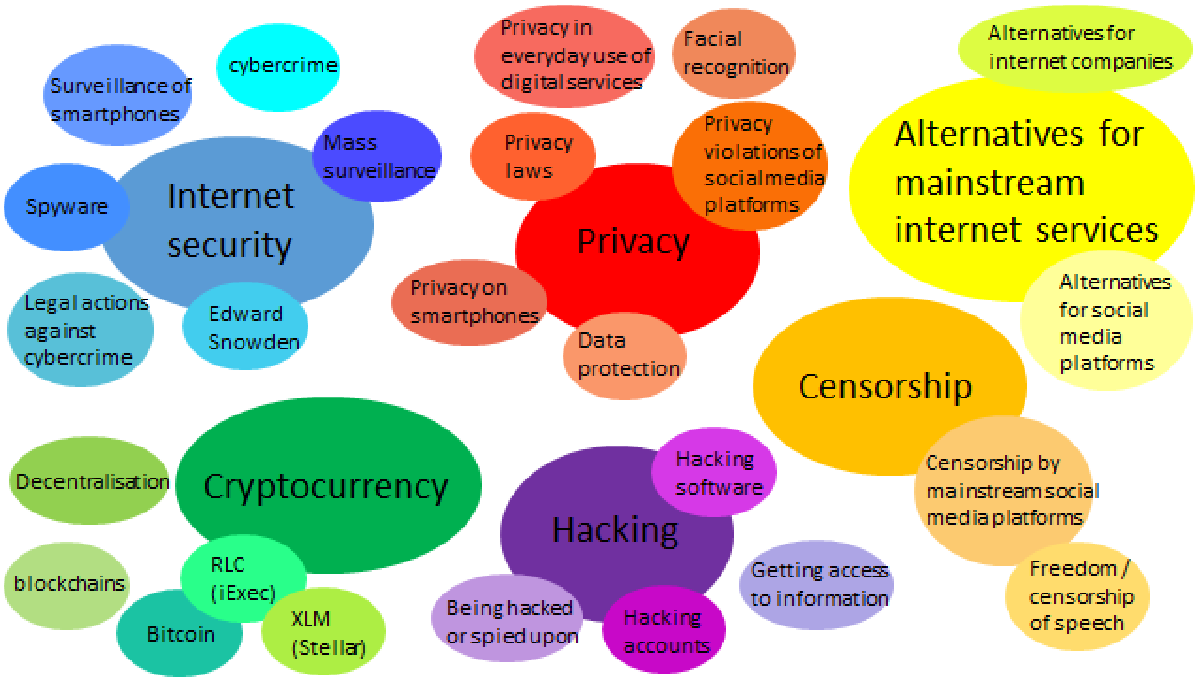

Surveillance of smartphones
cybercrime
Internet security
Mass surveillance
Spyware
Legal actions against cybercrime
Edward Snowden
Privacy in everyday use of digital services
Facial recognition
Privacy laws
Privacy violations of social media platforms
Privacy
Privacy on smartphones
Data protection
Alternatives for internet companies
Alternatives for mainstream internet services
Alternatives for social media platforms
Censorship
Censorship by mainstream social media platforms
Freedom / censorship of speech
Getting access to information
Decentralisation
Cryptocurrency
blockchains
RLC (iExec)
Bitcoin
XLM (Stellar)
Hacking
Hacking software
Being hacked or spied upon
Hacking accounts

Figure 1: A summary of the identified trending topics with subtopics. The topics and subtopics were qualitatively summarised from the topics discovered with the topic model. The sizes of the topics do not convey popularity of topics.

2.6 Discussion and conclusion

We analyzed 17 subreddits related to human rights on the internet. The applied trend reservoir model identifies trends based on the introduction of novel content that subsequently stays. We analyzed the topics of the most trending posts within each subreddit. The main topics that were qualitatively labelled based on the topic modeling results were ‘privacy’, ‘censorship’, ‘alternatives to mainstream internet services’, ‘internet security’, ‘hacking’, and ‘cryptocurrency’.

Some of these topics are addressed in the ten key rights and principles of human rights online (‘privacy’, ‘censorship’, and ‘internet security’). Those topics were discussed with concern, which underlines the need for protection of those rights and that they were not fully implemented in the online world yet. The topic ‘alternatives to mainstream internet services’ is not expressed as a right in the ten key rights and principles, but the wish for alternatives is grounded in unfulfilled needs (such as privacy protection). Some of the trending topics were technical discussions (‘hacking’ and ‘cryptocurrency’), but they also relate to human rights. Hacking is tied to the right for security, and cryptocurrency is based on blockchain technology, which relates to standards of the internet architecture. Those topics point to trending discussions that are both related to internet technology and human rights and also have accompanying societal issues.

In a previous report (D1.9), we selected eight key topics based on computational approaches and expert opinions from workshops. Comparing the identified trending topics to those eight key topics, we find that several topics re-occur and other topics become more prominent.

The largest topic in the current analysis ('privacy') was also prominent in the earlier report and constituted one of the key topics with the title 'personal data control'. That key topic stated privacy and data control to be basic digital rights. Similar subtopics were mentioned, such as facial recognition, tech giants and data protection concerns, European laws concerning privacy (GDPR), and privacy concerns about Chinese technology (Huawei). The last subtopic did not concern Huawei in the current analysis, but regarded visits to China and Chinese spyware. This supports the vision of Next Generation Internet to shape a future internet without the level of control exerted by the Chinese government, as they invoke mistrust and concern.

The identified topic 'censorship' was also part of a key topic called 'trustworthy information flow', the other part being fake news. The key topic mentions especially content moderation by online platforms and governments, which correspond to the currently identified topics of censorship by mainstream social media platforms and freedom of speech. It should be noted that although censorship might seem to limit freedom of speech, it is also a tool to combat hate speech and cyberviolence. Another identified topic, namely 'cryptocurrency', was also part of a key topic in the previous report ('decentralized power on the internet'). However, the previous report focused not on cryptocurrency itself, but more on the power of a few giant tech companies and how to decentralize internet power using, for example, blockchain technology.

Cryptocurrency showed up in the topic model on news media coverage in that key topic as well as in two other key topics. The current analysis showed 'cryptocurrency' to be a trendier topic than 'decentralization of power', with a lot of interest towards new technologies. This difference might be a result of using different methods, as the key topic 'decentralization of power' was identified by expert opinion to accommodate a broader picture regarding social issues in internet technology, and cryptocurrency was detected in discussions on the technology-centered discussion forum Reddit.

The identified topic 'alternatives for mainstream internet services' does not constitute a previously identified key topic. It is only slightly similar to the key topic 'decentralized power on the internet' as that topic includes searching for alternatives to the current few internet companies in order to disperse power. As such, the reasons to look for alternative digital services are privacy concerns, data control, censorship, content moderation and decentralizing power.

The two last identified topics of 'internet security' and 'hacking' did not constitute a key topic in the previous report, they were only slightly overlapping with the key topic 'safer online environments'. However, this key topic was more about inclusion, cyberviolence, hate speech and racism. The currently identified topics were both ethical (cybercrime, Edward Snowden, mass surveillance) and technical (hacking, software, spyware), potentially reflecting our choices using human values as a starting point and analyzing discussions on the tech-oriented Reddit network. The topics 'internet security' and 'hacking' are expressing concern about online safety and given their presence in this analysis, these concerns should be addressed in a future internet to make users feel safe and secure.

To conclude, we identified six main trending topics at the intersection of human values and internet technology. The overarching topics were 'the government of data' (data privacy, internet security), 'internet technologies' (cryptocurrency, hacking), 'ethical issues' (censorship), and 'a movement away from mainstream platforms'. Our findings justify the goal of Next Generation Internet to establish a democratic European model of the Internet. On the one hand, privacy concerns were very apparent and worries about Chinese spyware and control even reaching across national borders. On the other hand, the American model of a few tech companies virtually having monopoly on the internet is also not desirable, as the users consider data control very important, search for alternative internet services, and discuss censorship and content moderation. Additionally, human values were influencing the discussions with freedom of speech, freedom of choice (alternative internet services), personal data ownership and data protection (privacy). The topics presented here constitute important topics that are currently trending, and they facilitate the choices to be made on the way to the Next Generation Internet.

3 Deep dive into Twitter dataset of several NGI-related hashtags

3. 1 Background, research questions and hypotheses

The online community has grown to become an environment where millions of users engage in discussions concerning social, governmental, and environmental issues, as well as technological, political and personal news. It has become a direct channel for news coverage as well a place to publish one's own opinions and feelings. Utilizing the information from many social media users can assist in predicting the future content in social media, more data-driven research can be facilitated and lastly, it can help create more responsive policy-making in real life and on the internet, maintaining the functioning of a democratic environment on the platform.

Twitter is a highly popular social media platform, and with its 320 million monthly active users it is suitable for identifying the needs, considerations and fears of social media users (Desilver, 2016). Also, with the possibility of using hashtags, users are enabled to both target their tweets which gives a good opportunity to perform targeted analyses within specific hashtags.

This study is an exploratory study aimed at mapping discussions relevant for NGI by employing methods of topic modeling, emotion analysis, network analysis of hashtags. It attempts to create an in-depth inspection of the topics and their relation within and across the hashtags.

The hashtags chosen for this analysis were selected from a long list of trending keywords identified in deliverable D1.2 from a dataset of more than 213.000 technology news articles. The most trending keywords based on observing changes in frequencies of the keywords over time were identified and grouped into categories technologies and social issues. The keywords' trending score in terms of frequencies over time were then coupled with the keywords' popularity on Twitter (based on Hashtagify score, which assigns 100 to the most used hashtags and a value close to 0 for rarely used hashtags) to choose the five most popular technology keywords and the five most popular social issues keywords.

All in all, the hashtags are related to modern topics and fall into the three categories technology ('AI', 'quantumcomputing', 'blockchain', 'IoT', '5g'), communication ('hatespeech' and 'fakenews') and data privacy ('cybersecurity', 'privacy' and 'gdpr').

3.2 Twitter dataset

We use Twitter data collected between July, 2019 until January, 2020 using Twitter's public Streaming API. We collected tweets containing the following hashtags:

Hashtag	**Number of tweets collected**
IoT (Internet of Things)	765,397
blockchain	1,038,061
AI (Artificial Intelligence)	2,076,208
5g	1,313
quantumcomputing	27,983
fakenews	1,243,131
cybersecurity	1,075,170
privacy	225,606
gdpr	136,377
hatespeech	26,763

3.3 Analysis

3.3.1 Identifying topics relevant for NGI - Topic analysis

In this section, we bring forward which subjects and topics are dealt with in each of the datasets containing the 10 hashtags listed above. To identify the topics that are relevant in relation to NGI, we highlight topics that reoccur across hashtags and investigate central words.

3.3.2 Text preprocessing

Firstly, all datasets were cleaned. This involved removing hashtags, urls, mentions, punctuation and stopwords. Words that appeared in less than 5 documents and above 50% of the documents were removed as well. We performed lemmatization and 'part-of-speech'-tagging using SpaCy (Honnibal & Montani, 2017). With 'part-of-speech'-tagging, nouns, adjectives, verbs and adverbs were identified and only these were kept in the tweets.

3.3.3 Selection of amount of topics and LDA

To choose the optimal amount of topics to compute, coherence values from a large range of topics were tested. Coherence measures the semantic similarity between the words occurring in a topic. We tested between 10 and 50 topics with intervals of Using Latent Dirichlet Allocation (LDA) (Blei et al., 2003) each dataset containing a hashtag was categorized into the suitable amount of topics defined above. We kept an automatic eta and an asymmetric alpha parameter. As dealing with all optimal topics would be too comprehensive, we selected the top three dominant topics for further inspection. The 10 most relevant keywords for the top three topics across all documents in each dataset were extracted and using these, we labelled the topics and counted which labels reoccurred across topics and across hashtags.

3.3.5 Emotionally loaded discussions

This section attempts to identify which emotions dominate each hashtag. Based on a pre-trained emotion classifier, each tweet in every dataset is labeled with the most dominant emotion based on its text. The model used to classify the emotional categories of tweets is based on a BERT transformers model which has been fine-tuned on a Twitter emotion dataset provided by HuggingFace. It classifies each tweet as either 'sadness', 'joy', 'love', 'anger', 'fear' or 'surprise' and returns the probability or certainty of the classified emotion. The model is developed by Bhadresh Savani and was downloaded through HuggingFace at bhadresh-savani/distilbert-base-uncased-emotion (15/06/2021). To see whether there were a substantially larger or smaller amount of emotions, we compared every hashtag and every emotion within a hashtag against a baseline distribution. The baseline was created by collecting the emotional distribution of all hashtags into one. We then tested whether there was a significant difference between the hashtag in question and the baseline using Pearson's chi square test of independence. As the test was performed multiple times, p-values were corrected using Bonferroni. After understanding which distributions were significantly different, the individual emotion within every hashtag and the baseline were compared using a post-hoc residuals test which was, as well, corrected using Bonferroni.

3.3.8 The network of topics surrounding NGI relevant topics - Co-hashtag networks

To further inspect the semantic context and the structure of the 10 hashtags, we performed a network analysis using the hashtag itself and the most frequent hashtags it is mentioned with. From all tweets, we extracted all hashtags and listed every combination of hashtag pairs per tweet, counted their individual occurrence as well as their co-occurrence. With this information, the top 200 co-occurring hashtags were selected to create a network that would be simple but cohesive and present the most important hashtags. The total size of every network was collected as well. The network was visualized with the spring layout, using the Fruchterman-Reingold force-directed algorithm.

3.4 Results

3.4.1 Topic analysis

The topic analysis returned both tables with central keywords and word clouds from each topic in each of the hashtags which all can be found in the appendix, an example is provided below. We will highlight the key information from each hashtags and the recurring themes seen across the hashtags.

Within most of the hashtags we see that the top three topics dominating the field have different keywords and labels. We see for example that within the hashtag AI, both tech-related, healthcare related and solution related topics dominate the discussion. Within the hashtag 'quantumcomputing', topics related to finances, behaviour and research are dealt with. The outcome is, however, opposite for 'IoT' where topics overlap to some extent and are all related to the tech industry and developing a tech business. Despite this, most topics within the 10 hashtags reflect that the hashtags are used for different purposes and different discussions.

Additionally, there are topics that reoccur across the hashtags. Apart from the hashtags 'AI', 'hatespeech', 'fakenews' and '5g' we see that the topics are related to business and security. The topics are most likely concerning business or security within the specific field of the hashtag so that it is cybersecurity-related business when the hashtag is 'cybersecurity'. These results suggest that Twitter is a platform where users go to, in general, discuss such topics. When these topics are recurring, it additionally points to the possibility that these topics are relevant for NGI and will continue to be dealt with in the future.

Topic #n	Keywords	Label
1	cryptography, ibm, lead, day, key, breakthrough, so, paper, show, people	Research
2	quantumcompute, noise, researcher, computer, classical, old, limit, beat, random, fundamentally	Technical
3	world, change, problem, solve, technology, how, potential, business, company, quantumcomputer	Prospects, Innovation, and Business

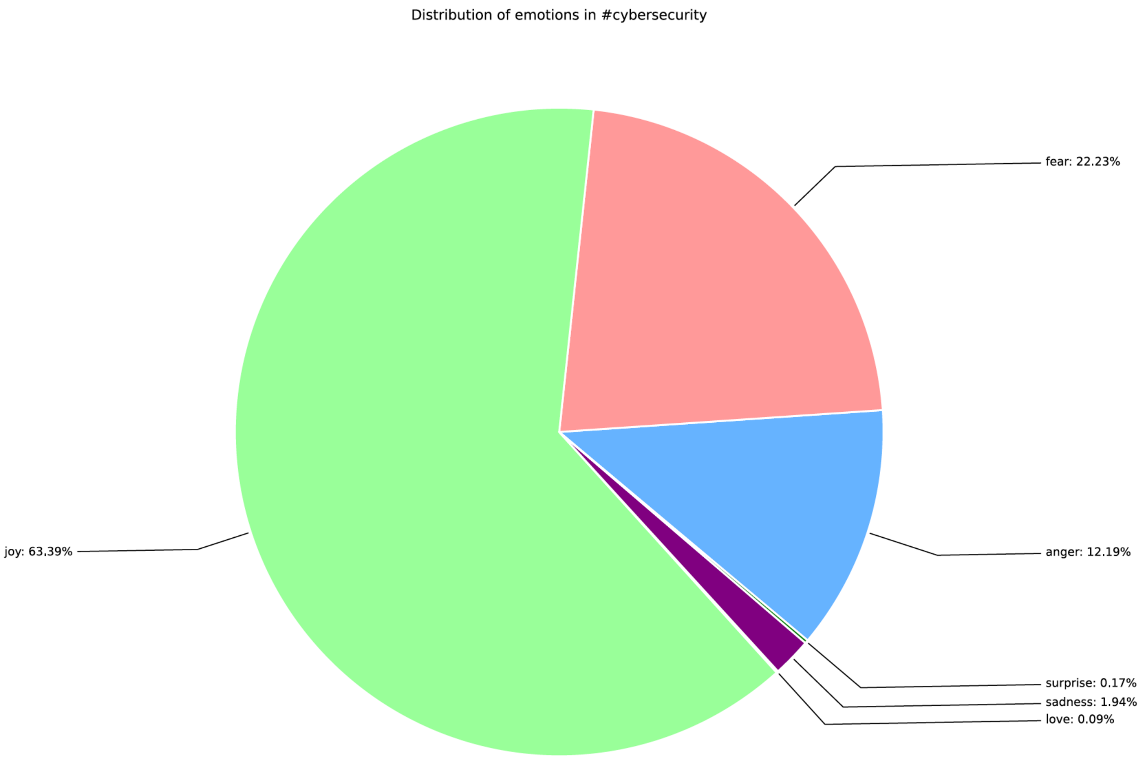
3.4.2 Emotion analysis

In general, all hashtags contained significantly different distributions than the baseline. Also, the proportion of most emotions within each hashtag were either significantly larger or significantly smaller than the respective emotion in the baseline set. All tables and pie-chart distribution of emotions can be seen in the appendix. These analyses reveal multiple observations that need attention.

Firstly, hashtags such as ‘AI’, ‘IoT’, ‘quantumcomputing’, and ‘blockchain’ consist of significantly fewer anger, fear and sadness related emotions. In contrast, joy and surprise-dominated tweets are significantly more present. Thereby it seems like users are positive about these subjects. The reason for the more positively loaded discussion might stem from the fact that the hashtags are rather technical and therefore the terms used are fact-based and neutral. This idea is reflected in the topic models of the hashtags where most labels relate to research and technology. Also, users might express surprise because of discussions concerning new technology within the subject. This is as well emphasized by the topic models results where “breakthrough” for example is one of the keywords in a topic in ‘quantumcomputing’.

Further, we see a pattern in relation to ‘cybersecurity’ and ‘privacy’. The emotions here are primarily dominated by a significantly larger amount of fear and anger, suggesting that there exists a lot of concern in relation to security. Instead of being hashtags that users apply to express a feeling of safety and security it is used for the opposite. More specifically, it might reflect insecurity on the internet as the keywords “hacker”, “hacking”. “malware”, and “phishing” occur in cybersecurity and privacy topics.

Lastly, some final notes on ‘5g’, ‘hatespeech’, ‘fakenews’, and ‘gdpr’ need to be highlighted. In relation to ‘5g’, we see a surprising finding; the hashtag is primarily dominated by a larger amount of positive and neutral tweets and negative tweets are significantly less present. One could have expected that the hashtag would consist of conspiracies in relation to consequences of the 5g technology. However, either users are not discussing these aspects on Twitter or there is a more neutral counterpiece or even a positive contra movement to the discussion. On the other hand, tweets on 'hatespeech' - not surprisingly - mostly consist of angry tweets. Fakenews is the hashtag which reflects the largest amount of different emotions where all emotions except joy are significantly more present suggesting the topic is complex. Lastly, 'gdpr' is interesting as it both reflects a larger amount of anger but as well joy; the attitude towards the regulation both consists of resistance and support.

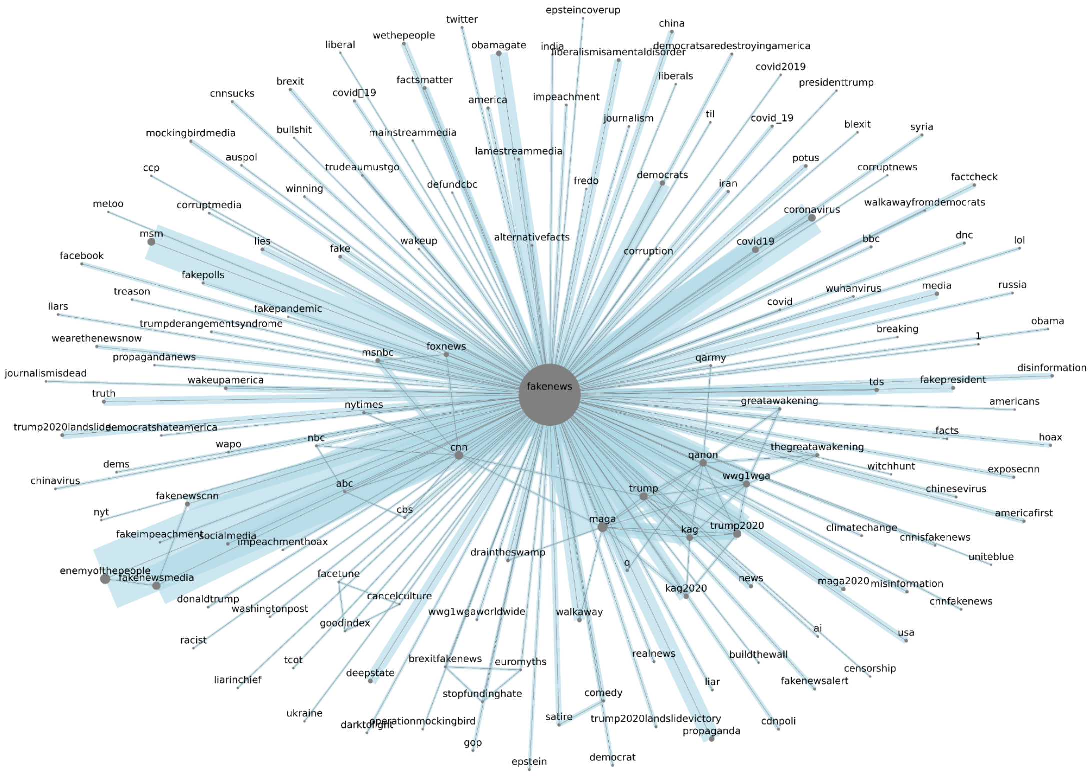


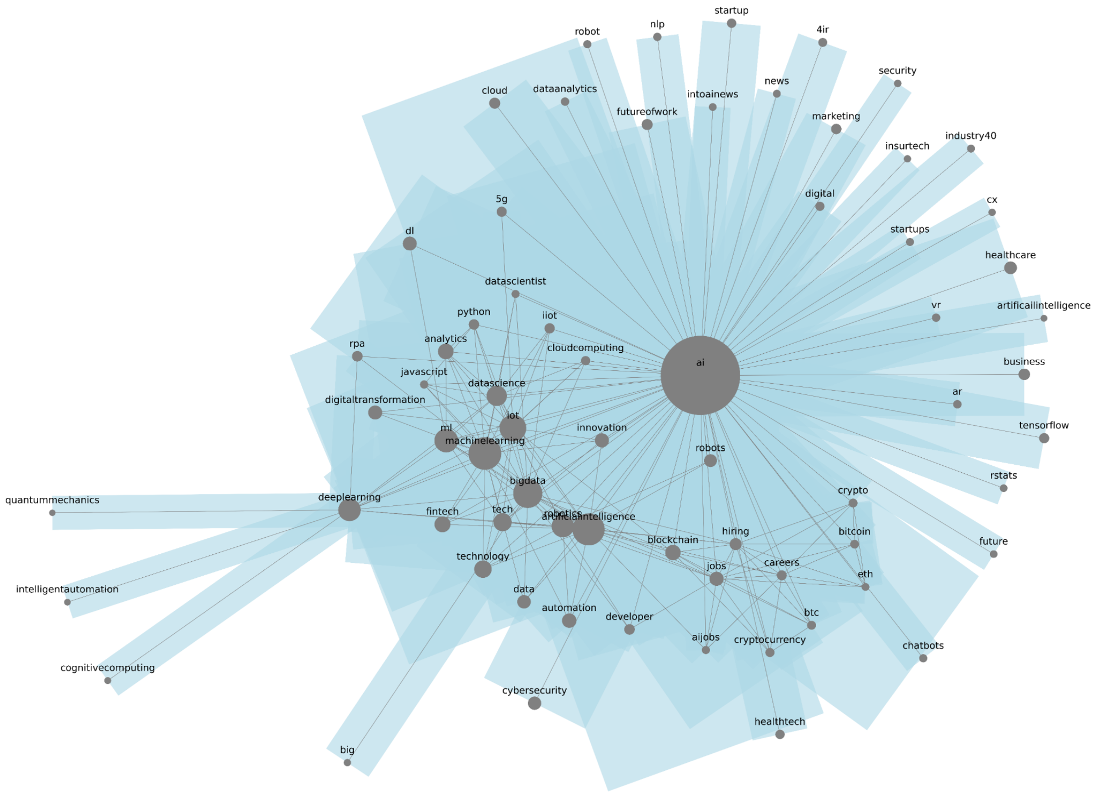
3.4.3 Network analysis

To create a network which would not be too big to inspect but still contain relevant and highly occurring nodes, we selected the 200 most co-mentioned hashtags in the network. To see network size differences between the 10 hashtags, we refer to the appendix. The hashtag in question is the center of every network, the thickness of the edges represent how much it is mentioned with another node/hashtag. The size of the nodes represent how many tweets the hashtag is mentioned in itself.

Generally, we see that the resulting networks primarily fall in two categories. Either the hashtag is very central itself and has a lot of co-mentioned hashtags that are not connected to other hashtags than the hashtag in focus (example is the first figure below). Or the hashtag is connected to other hashtags that are quite large as well and form subnetworks of connections (second figure below). Examples of the first are '5g', 'quantumcomputing' and 'fakenews' and of the latter are 'cybersecurity', 'blockchain', 'IoT' and 'AI'. The networks in 'IoT' and 'AI' reflect both similar structure and content. In all networks, we see subnetworks with the co-mentioned hashtags ‘artificialintelligence’, ‘bigdata’ and ‘machinelearning’. From the networks, ‘AI’ contains the densest relation to its co-hashtags with many more co-mentions.

The reason for the organization of the networks might be because some hashtags are very specific while others have many words that can be associated with it. One example here is the hashtag ‘hatespeech’ in contrast to ‘AI’. But it can also reflect that when using some hashtags, the discussion concerns a larger field whereas discussions related to some of the other hashtags are more specific.





3.5 Discussion

We find that the majority of the hashtags contain discussions about business as well as security. On one hand, users treat the hashtags and subjects as having potential and being a possible business. On the other hand, the subjects raise concerns about security and privacy among the users. Especially the discussions of ‘privacy’ are negatively loaded, containing fearful and angry tweets.

Furthermore, we find stronger connections and similarities between the technology-related hashtag (‘AI’, ‘quantumcomputing’, ‘blockchain’, ‘IoT’, ‘5g’) than communication (‘fakenews’, ‘hatespeech’) and data privacy-related hashtags (‘cybersecurity’, ‘privacy’, ‘gdpr’).

In general, joy is the predominant emotion in the tweets with few exceptions such as ‘fakenews’ and ‘hatespeech’. The algorithm classifying these emotions has been trained on the Emotion Twitter Dataset, where there are no neutral labels, possibly resulting in that some neutral sentences are labelled as joyful. A few examples are “i do however want you to know that if something someone is causing you to feel less then your splendid self step away from them” or "on a boat trip to denmark"[4]. Therefore, it is unclear to which extent the distribution of joyful tweets are in fact joyful and not simply neutral.

[4] Can be found at https://huggingface.co/datasets/viewer/?dataset=emotion

Though there seem to be clear indications of which topics are discussed and how the discussion is emotionally loaded on Twitter, one can reflect on which voices they represent. Users interact differently with the platform, and some specific users are more active than others. In fact, Pew Research Center found that many tweets stem from a small minority of tweeters; though the amount of tweets in our dataset is large, the results can have been affected by the interests of the few users who tweet exceptionally much (Wojcik & Hughes, 2019).

3.6 Conclusion

This article performed an extensive analysis including topic modelling, emotion analysis and network of 10 trending hashtags on Twitter. The goal was to map and nuance the topics dealt with in the hashtags as well as detect the similarities across hashtags. We found that users express a need to discuss security and technology within the majority of the hashtags. These co-occurring topics are relevant for NGI to go in-depth with in the future as they seem strongly manifested in the Twitter community.

4 Deep dive into gender inequality

Several of the ten key rights and principles of human rights online mention equality and diversity online, namely the principles of 'universality and equality', 'diversity' and 'network equality'. Similarly, one of the six umbrella topics identified by our NGI research consortium in early 2021 was 'access, inclusion and justice', with equality and discrimination as subtopics. We make a deep dive into this topic by examining gender inequality associated with artificial intelligence. We apply the gender concept in a binary form, being aware that many different gender identities exist (though to a smaller extent). Here we refer to an already published article (Bechmann & Bowker, 2019), where we have extended and trained an up-to-date neural network anew. This section is an extract of a paper to be submitted to an academic journal. We center the analysis on a case study of Facebook picture albums for a sample mirroring the Danish national population, a country which historically has had a liberal approach to gender equality.

4.1 Background, research questions and hypotheses

Critical algorithmic studies and science and technology studies have criticized algorithms for discriminating against social classes when implementing automated-decision making (Caliskan et al., 2017; Citron & Pasquale, 2014; Eubanks, n.d.; Howard, 2006; Levin, 2016; Sweeney, 2013). Such discrimination has especially been documented towards etnic minorities and females who have been retained in non-empowering positions, amplifying an already discriminating pattern (Ananny & Crawford, 2018; Boyd & Crawford, 2012; Cheney-Lippold, 2017; Citron & Pasquale, 2014; Elish & Boyd, 2018; O'neil, 2016; Sandvig et al., 2016). Here it is important to emphasize that it is seldom the mathematical models themselves that discriminate but it results from the contribution of the human components - those that make decisions about what model to choose for what purpose, what data to train the algorithm on, how to interpret thresholds, etc. (Bechmann & Bowker, 2019). Also, the human component as a potential discriminator is visible in the structures of the datasets that are used for training. If the dataset shows discriminatory patterns, the algorithm will reproduce and amplify such patterns in the recommendations of the automated decision making. For example, afro-americans get ads that align with already existing social classes and thereby reinforce such positions (Sweeney, 2013), or females are kept in a stereotypic interpretative pattern (Bechmann, 2017; Bolukbasi et al., 2016).

Here, we investigate the potentially discriminatory patterns that an algorithm learns from a dataset of uploaded Facebook images. Our main research question was: can an algorithm learn patterns in uploaded pictures and predict the gender of the uploader? We hypothesize that there is a detectable difference in pictures uploaded by men and by women, but that there also is a large overlap in the type of pictures uploaded. We further set out to explore the question of which patterns the algorithm learns to correctly predict the gender category. Do these patterns correspond to perceived stereotypes of each gender? We expect to find some patterns that relate to stereotypical behavior. Lastly, we ask which patterns the algorithm contributes falsely to the other gender category, as those patterns reflect behavior that does not fit the expected gender behavior. It is here that most societal issues occur, and it is problematic when an algorithm amplifies existing biases already present in a dataset. A gender bias that correctly distinguishes between gender is not causing problems, but a gender bias that wrongly attributes characteristics to a gender category is.

4.2 Facebook dataset

The dataset was collected in 2014 through the Facebook API with a multi-step written informed consent procedure of the research participants (Bechmann & Vahlstrup, 2015) and permitted by the Danish Data Agency. The international internet panel Userneeds provided participants for the study and we included 1,000 Danes sampled to mirror the Danish Facebook population stratified on age, gender and area of residence (Bechmann & Bowker, 2019; Bechmann, 2019). We collected all uploaded Facebook images (from the Facebook photo albums) available at the point of collection and obtained a total of 340,000 images. The self-identified (binary) gender by the participants was used as labels for the study. We split the datasets into pictures uploaded by males and females respectively and preprocessed the data to contain no profile pictures (of themselves) and no metadata that explained the picture. Furthermore, we cropped the images to equal size in order not to create white space as a signal in the model. Only participants with at least one image apart from the profile picture were included, and the number of uploaded images ranged from one to 13,125 (Bechmann & Bowker, 2019). After the data cleaning, we had 238,173 images uploaded by 486 females and 106,797 images uploaded by 397 males, indicating that females had uploaded more images than males. Only one individual had uploaded more than 10,000 images, 85 individuals more than 1,000 images, 441 individuals more than 100 images, and 722 individuals more than 10 images.

4.3 Analysis

4.3.1 Convolutional neural networks

Convolutional neural networks (CNN) are deep learning networks, which are often applied to analyze visual images. They learn the spatial patterns of an image by extracting features at small, mid and large spatial scales. We trained a CNN using the PyTorch framework with the pretrained ResNet-18 model (He et al., 2015). This network architecture was chosen because it is small and therefore fast to train and less prone to overfitting.

4.3.2 Data preprocessing

The dataset was randomly split into a training set of 90% and a test set of 10% of the images. The split was done using a random seed to ensure reproducible results (the split was random but deterministic). We restricted the number of images per individual to a maximum of 100 images (chosen at random), to deal with the large differences in number of uploaded images. Without this restriction, a few individuals with a large number of images would dominate the result of the classification and the results would not be generalizable. Even though this results in less images for training the model, it will make the prediction better, as the number of individuals stays the same and the number of pictures per individual is more balanced. We also addressed the imbalance in the number of pictures in each gender category by using a sampler. The sampler ensures that, on average, each training batch contains an equal number of female and male images. Further, we used data augmentation to reduce the risk of model overfitting. The augmentation included random cropping and resizing, random rotation (+/- 15 degrees), horizontal flipping, and color jitter.

4.3.3 Data analysis

The classification results were summarized in a confusion matrix, which depicts the predicted gender category versus the actual gender category. The accuracy of the model was calculated as Accuracy = (true positives + true negatives) / (true positives + true negatives + false positives + false negatives). The model returned a score of the predicted category for each image. A score close to zero or one means that the model is very confident about its prediction. Specifically, a score close to one indicates a true positive where the model is very confident about its prediction. Likewise, a score close to zero indicates a false negative where the model also is very confident. In this study we are primarily interested in the images where the model is very confident about its prediction, as the true positive and false negative images display patterns that were correctly or falsely attributed to a gender class with high confidence, respectively. The high confidence reflects a strong detectable pattern within a gender category and points to stereotypical behavior. Lastly, heatmaps displaying the areas that the model focused on in the image to make its prediction were generated using GradCam (Selvaraju et al., 2020).

4.4 Results

4.4.1 Prediction of gender based on uploaded images

Figure 1 displays the actual and predicted gender categories in a confusion matrix. The accuracy of the model was 64%. For the images uploaded by men, 60% could be classified correctly, whereas 40% were misclassified. Likewise, for the images uploaded by women, 68% were classified correctly and 32% were misclassified. The results show that the model has learned some gender-specific patterns, as otherwise the predictions would have been only 50% correct. This means that there are detectable differences in the images that are uploaded by women or men. The model achieved a higher prediction rate for women, meaning that it detected more patterns in pictures uploaded by women than by men. However, not all images were classified correctly, which shows that there is considerable overlap in the images uploaded by both males and females.

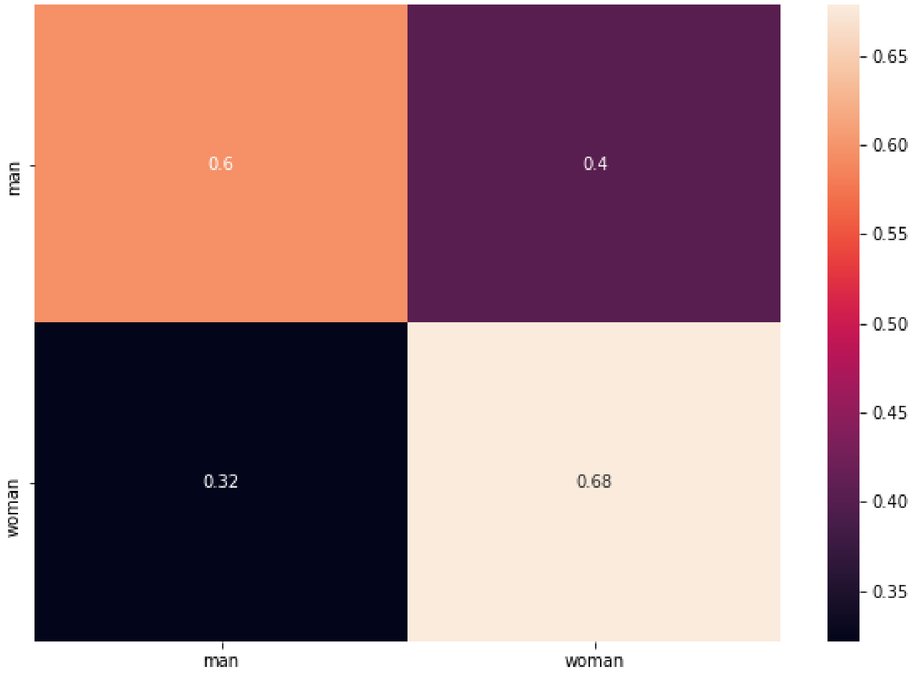


Figure 1: Confusion matrix on the testset. The actual gender categories are displayed as rows and the predicted categories as columns.

4.4.2 Detected gender patterns in the uploaded images

First, we looked at the patterns in the images that were correctly predicted with high certainty by the model. Here, we describe the images but cannot display them due to privacy reasons. When looking at the 200 pictures with the largest score that were correctly predicted to be uploaded by women, several patterns are noticeable. Most images are of people, and there are many close-up images of one or two people as well as pictures of children. Further, many pictures display dogs, horses, weddings, flower decorations, and cartoon characters.

The top 200 pictures that were correctly classified as being uploaded by men contained noticeably less images of people and more images of objects. Those images that showed people were either showing the face of a man, people doing sports, or group pictures (mainly of sport teams). The pictured objects were buildings, vehicles (cars, ships, trains, planes), signs, food, or bottles containing alcohol. Many pictures showed nature, wild animals, snow, views of both natural and city landscapes, text, and shots at night. Overall, the top 200 correctly predicted pictures were more diverse for men than for women. Another striking difference is that the pictures uploaded by women consisted of people looking straight into the camera, whereas those by men were more pictures showing the photographer's view naturally.

The heatmaps show which parts of the images the model focused on for its prediction. For the correctly predicted female category, the model used mainly the middle part of the image. They displayed people, pets, flowers, cartoon characters, and bridal couples were often in the center of the image. In those images, where this was not the case, the model focused on people or faces that were off-center in the image. In a few cases, objects, buildings, pets and cakes were shown where the model focused on other parts than the center.

Correspondingly, for the correctly classified male category, the model often focused on the middle of the image, but less often and with a more widespread area than for the correctly classified female category. Those images contained people, animals, vehicles, and objects. When the focus was off-center, the images showed people off-center, objects, text, groups of people, nature, or food. For both female and male categories, the model often focused on the area containing people, faces, vehicles, or objects when present in the picture.

4.4.3 Falsely assumed gender patterns

Subsequently, we looked at the images that were classified wrongly with a high certainty to the other gender category. Those results point to situations where algorithms might show gender biases. We looked at the top 200 pictures that were falsely predicted to be uploaded by women, but that were in fact uploaded by men. They show people, especially children or women. They further depict flowers, weddings, pets (dogs, cats, horses), water, and buildings.

On the other hand, the top 200 pictures that were falsely classified to be uploaded by men but that were uploaded by women, contained alcoholic beverages, snow, wild animals, text, racing cars, buildings, nature, sunsets, food, art, cats, wild animals, group pictures, or people in action (events and sports).

The heatmaps with the focus of the model also show interesting patterns. For the wrongly predicted images assigned to women but actually uploaded by men, the focus was often in the center. Exceptions were images containing several people, objects, nature, or faces off-center. Interestingly, the model focused on faces when adults were displayed but on the area just below the face when babies and small children were displayed.

Looking at the images that were falsely classified as men but did actually belong to women, the focus was both on and off-center. It was often on the displayed objects, people, mouths, and animals. When the model did not focus on the displayed items, the images often contained food, text, landscapes, buildings and other uniform patterns.

4.5 Discussion and conclusion

Our goal was to shed light on gender inequality in algorithmic decision making. We focused on the patterns learned by training a neural network on Facebook pictures uploaded by users. Our model could predict the correct gender category with an accuracy of 64%, with a better performance for images uploaded by females than by men. For the correctly classified images, females had uploaded many images containing people and children, and further dogs, horses, weddings, flowers, and cartoon characters. On the other hand, images uploaded by males displayed more often objects and less often people. They contained more varied subjects such as male faces, people in action and doing sports, sports teams, buildings, vehicles, signs, food, alcohol, nature, wild animals, snow, views, text and night shots. Further, the pictures of women were more set into the scene, whereas the pictures of men took an on-lookers view. This was reflected in the parts of the images that the model focused on for its classification, with the images of females having the focus in the middle of the image and on the displayed people. For the images of males, the focus was also in the center but less often and on a wider area. Lastly, for the falsely classified images, the images predicted to be uploaded by women but actually uploaded by men often displayed women and children, as well as flowers, weddings, pets, water and buildings. Here, the model focused often on the center of the image and on faces, except for displayed babies where the focus was just below the face. On the other hand, the images predicted to be uploaded by men but that were uploaded by women depicted alcohol, snow, wild animals, text, racing cars, buildings, nature, sunsets, food, art, cats, wild animals, group pictures, and people in action. The focus of the model was both on and off-center.

Gender stereotypes still influence online behavior (Bailey et al., 2013; Eisenchlas, 2013; McAndrew & Jeong, 2012; Oberst et al., 2016; G. Park et al., 2016; Rose et al., 2012). Studies have confirmed both perceived differences (Bailey et al., 2013; Eisenchlas, 2013) as well as actual differences (Makashvili et al., 2013; McAndrew & Jeong, 2012; Muscanell & Guadagno, 2012) between women and men. Here, we found several differences in the online posting behavior of women versus that of men. For example, women uploaded more pictures than men in our dataset, which might be a result of women spending more time on Facebook than men (Frison & Eggermont, 2016; McAndrew & Jeong, 2012) and using it for the purpose of uploading pictures (Makashvili et al., 2013). Women also posted more pictures that were set in scenes compared to men, which fits previous reports of women using Facebook for impression management (McAndrew & Jeong, 2012). However, the largest difference was found in that women posted more pictures of people on Facebook. This gender difference can be embedded in the framework of the social role theory (Eagly, 1997). This theory allocates different roles to men and women, which influence their upbringing, beliefs and behavior. Men are commonly attributed to be more task and information oriented and women more communal and interpersonally orientated (Guadagno et al., 2011). Similarly, individualistic traits are viewed cross-culturally as more masculine and collectivistic traits as more feminine (Cuddy et al., 2015). Also, women are associated with being dependent and men with being independent (Cuddy et al., 2010; Guadagno et al., 2011; Rose et al., 2012). This theory and the literature offers an explanation for why women upload more pictures containing people, as women focus their behavior on engaging with others. It also explains why the other pictures of women often contained pets (signifying a family-like connection), weddings (celebrating the love of a couple), and flower decorations (creating a nicer environment for people). Those subjects are closely related to communal events and family bonds. Another Facebook study similarly found that female discussions on Facebook contained the topics 'friends', 'family' and 'social life' (G. Park et al., 2016). For men, the theory explains why the pictures uploaded by men contained more objects and action shots, as those subjects relate to tasks and information. Another study also found that males were discussing objects more compared to females when using Facebook (G. Park et al., 2016). Similarly, Rose et al. evaluated self-selected Facebook profile pictures and found that pictures of males were associated more with the trait 'active' compared to females, as their images often showed athletic gear, sports or outdoor settings (Rose et al., 2012). Our findings corroborate those results, as we found many images displaying action or sports, but also nature, wild animals, snow, and night pictures.

The algorithm also falsely classified some pictures to the opposite gender category. Those pictures often contained the same subjects that the opposite gender category had shown in the pictures that were correctly classified to that gender category with a high score of confidence. This finding makes sense, because stereotypical behavior is predominantly shown by one class but is often not only shown by that one class. This is not a problem in itself; it only becomes one when it has negative consequences for the class not conforming to expected stereotypes. For example, Otterbacher and colleagues investigated the images that were shown for online queries (Otterbacher et al., 2017). They found that the query 'person' yielded more photos of men than of women. This can proliferate into a subconscious cultural understanding that men are more 'persons' than women and lead to subsequent beliefs that they have more rights. Such negative consequences from behavior that is not conforming to stereotypes are called backlashes. Otterbacher et al. reported a larger backlash effect for the search term 'competent women' than for 'warm men', but both terms were associated with negative depictions.

Possible solutions for fairer algorithms are more caution by and education of developers, as well as regulation through policies as this area is barely regulated right now (Bechmann & Bowker, 2019). Algorithms can also be adjusted to include subclasses or alternative classes (Bechmann & Bowker, 2019). Explicit removal of the bias is also a possibility as in one study, where the authors removed gender associations of words to generate an algorithm that decreased the gender bias (Bolukbasi et al., 2016). Another possibility is using fairness in machine learning, or adding explicit rules to the implicitly learned patterns (Caliskan et al., 2017).

To conclude, gender differences were shown to be present in uploaded pictures online. They can in general be related to females showing social behavior and males behavior oriented towards action and objects. However, the classification algorithm also falsely attributed images to the other gender category, which displayed the same subjects that were often displayed by the opposite gender category and accordingly often correctly classified. With this study we show how choices made by algorithm developers might result in gender discrimination because the dataset contained gender differences. It is important to be aware of unwanted biases in datasets and to rectify the disadvantage some groups are experiencing at the hand of discriminating algorithms.

5 Deep dive into disinformation – sentiments of COVID-19 misinformation on Twitter

This chapter will make a deeper dive into the topic of disinformation and misinformation as a next generation internet challenge balancing between diversity, expression and association (as addressed in the introduction - the ten key rights and principles of human rights online) and the safety of society in the light of COVID-19. The chapter is an extract of the scientific paper accepted to be published in the academic journal Big Data & Society in 2021. We have made an extract of the academic publication text in this section for convenience, but we refer readers to the full online open access version and in case you want to cite the results please use: Charquero-Ballester, M, Walter, J.G., Nissen, I.A. & Bechmann, A. (2021). Different types of COVID-19 misinformation have different emotional valence on Twitter, Big Data & Society.

5.1 Background, research questions and hypotheses

“We’re not just fighting an epidemic; we’re fighting an infodemic”, the director-general of the WHO declared in February 2020. Since then, the spread of false and misleading information about COVID-19 on social media has only intensified (Brennen et al., 2020). Our aim was to contribute to a better understanding of COVID-19 misinformation by looking at the extent and associated emotional valence as one potential factor influencing spreading on social media. Our focus on COVID-19 related misinformation and its emotional valence was sparked by inconsistencies on the amount and character of misinformation and emotions in the context of COVID-19 (e.g. Brennen et al., 2020; Kouzy et al., 2020; Singh et al., 2020).

We supplement previous studies with a broader understanding of emotional valence of misinformation content related to COVID-19 on Twitter and investigated, if we empirically saw any differences in this valence depending on the type of misinformation. We expected that the emotional valence of COVID-19 related misinformation tweets depend on the type of misinformation as also (Li et al., 2020) found a relation between topic and emotion. Our study: 1) was based on all English fact-checked stories from Google Fact Check Explorer in March 2020, 2) coded these stories manually into types of misinformation to increase the sample size per type for a reliable sentiment analysis, 3) included the creation of a classifier for each story and its application to tweets from March for related COVID-19 hashtags, and 4) was continued by the manual coding of a random subset of the tweets for misinformation to increase sensitivity (finding true positives) and specificity (avoiding false positives) before 5) it reported the measure of the emotional valence scores and differences between types.

The results of the study are a first step to understand how the misinformation content itself might ignite different emotional valence when it spreads on social media. Even though tools for automatic detection of misinformation are still improving, detecting misinformation in real time is likely to remain a significant and enduring challenge due to the high velocity, volume and variety of dis- and misinformation. Thus, a better understanding of any of its contexts and drivers of diffusion is essential for minimizing its potential impact on society.

5.2 Misinformation dataset

5.2.1 Misinformation stories on Google Fact Check Explorer

First, we extracted misinformation that was debunked by fact-checking organizations through the Google Fact Check Explorer. We collected all English language misinformation stories (irrespective of the rating result) related to COVID-19 from 1st until 31st of March 2020, as this was when the pandemic was globally wide-spread and the amount of misinformation was high due to still being in the early stages of the crisis. This yielded 247 debunked stories, and after removing duplicates 226 stories remained.

5.2.2 Twitter sample

The initial sample of tweets was obtained through a publicly available coronavirus Twitter dataset containing over 123 million tweets, with over 60% of them in English collected through 76 hashtags related to COVID-19 (E. Chen et al., 2020). We narrowed down the sample to March 2020, selected tweets in English language and removed retweets, which gave us a sample of 17,463,220 tweets.

5.2.3 Selecting tweets related to misinformation stories

Tweets related to the misinformation stories were found by matching the tweets to keywords selected from the misinformation stories. In a first step to decrease the size of the data, a sample of tweets for each debunked story was identified by selecting all tweets that contained a primary keyword. This primary keyword was the most central single word or bigram of the misinformation story title, and was selected manually by two annotators independently. The annotators selected identical keywords in 89% of the stories, and for the rest the final keyword was selected in discussion. In a second step, false positives (tweets that were captured in the first step but were not related to the debunked stories) were filtered out based on secondary keywords. Those secondary keywords were also manually selected by the two annotators and contained all relevant words of the story title (excluding common or unspecific words and repetitions). At least one word of the secondary keywords was required for selection of a tweet, such that the selected tweets contained the first keyword and one or more of the secondary keywords. This two-step approach was necessary to reduce processing time, as filtering tweets based on all keywords at once would be computationally too costly. In the end, we had 690,004 tweets that discussed misinformation related content.

When checking the selected tweets we noticed that some of the keywords were better than others in selecting tweets relevant to the debunked story, due to the generic character of some stories and subsequent keywords (e.g. vaccine, virus). To increase the validity of the study by reducing false positives, we needed to manually select a subset of tweets that contained true positives rather than running the emotional valence analysis on tweets containing many claims unrelated to the debunked stories. We therefore randomly selected 100 tweets for each of the 226 debunked stories or, if fewer tweets were found for a story, all related tweets. From this subset of tweets we selected those that were related to the story and deleted recurring tweets, which provided us with a sample of 2,097 tweets for the sentiment analysis.

5.3 Analysis

5.3.1 Types of misinformation

We categorized the debunked stories into six types of misinformation, which were made bottom up by two researchers from the stories at hand: “cure, prevention & treatment” (shortly “cures”), “conspiracy”, “political measures” (shortly “politics”), “vaccine & test kits” (shortly “vaccine”), “virus characteristics & numbers” (shortly “virus”) and “other”. The “other” category contained stories related to individual fates, economy or criminal behavior and stories that could not be assigned to the remaining five categories. The clustering was inspired by a typology of narratives provided by the EU DisinfoLab (2020): health fears, conspiracy theories, lockdown fears, false cures, identity, societal and political polarization. Two independent raters assigned the stories to the six categories with an acceptable inter-rater reliability (Krippendorff, 2018) of 0.7 (α= 0.74), and agreed upon a category if the rating differed. Figure 1 provides an overview of the datasets, filtering process and distribution of the types of misinformation.

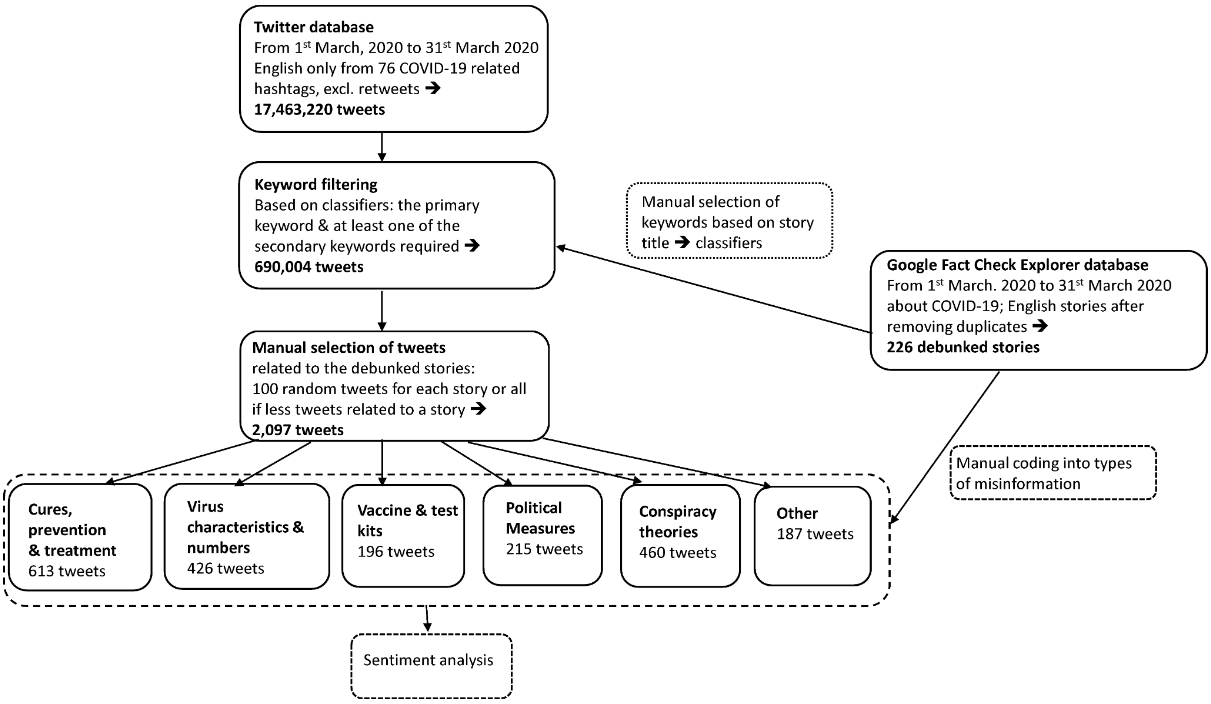

Twitter database
From 1st March, 2020 to 31st March 2020
English only from 76 COVID-19 related hashtags, excl. retweets ➔
17,463,220 tweets
Keyword filtering
Based on classifiers: the primary keyword & at least one of the secondary keywords required ➔
690,004 tweets
Manual selection of keywords based on story title ➔ classifiers
Google Fact Check Explorer database
From 1st March. 2020 to 31st March 2020 about COVID-19; English stories after removing duplicates ➔
226 debunked stories
Manual selection of tweets related to the debunked stories:
100 random tweets for each story or all if less tweets related to a story ➔
2,097 tweets
Manual coding into types of misinformation
Cures, prevention & treatment
613 tweets
Virus characteristics & numbers
426 tweets
Vaccine & test kits
196 tweets
Political Measures
215 tweets
Conspiracy theories
460 tweets
Other
187 tweets
Sentiment analysis

Figure 1. Illustration of the filtering process and distribution of tweets across the types of misinformation

5.3.2 Sentiment analysis

We estimated the emotional valence of the tweets using sentiment analysis by applying VADER (https://github.com/cjhutto/vaderSentiment) (Hutto & Gilbert, 2014). This model is based on a list of lexical features specifically attuned to sentiment analysis. We chose this model because it is specifically made for short texts and shows a high reliability (Ribeiro et al., 2016). The model produces four different scores: positive, negative, neutral, and a compound score. Positive, negative and neutral scores are ratios for the proportion of text that falls into each category. The compound score provides a single score of sentiment for any given sentence or tweet and is normalized between -1 (most negative) and +1 (most positive).

To test whether the sentiment of the different types of misinformation (e.g., disease prevention vs. political measures) was significantly different from each other, we ran a two-sample t-test permutation testing (10,000 permutations) on the compound scores. We corrected for multiple comparisons using the False Discovery Rate by (Benjamini & Hochberg, 1995).

5.4 Results

5.4.1 Overview of tweets related to misinformation

Out of the 226 debunked stories, 195 stories could be associated with tweets. On average, for the 195 debunked stories we had 3538 ± 21,527 (average ± standard deviation) tweets per story. The percentage of tweets that were manually confirmed to be related to the debunked story varied widely across the different stories. Keywords that were specific for a well-defined concept often captured tweets discussing the misinformed claim (such as 'RFID' 7/9 (78%), 'secret terrius' 62/100 (62%), and 'Vitamin C' 89/100 (89%)). In the cases where the primary keyword appeared in a different much-discussed context that was unrelated to the claim, the number of selected tweets was very low (such as 'Netherlands' 4/100 (4%), 'Amazon' 4/100 (4%), and 'sun' 2/100 (2%)). The average number of manually selected tweets for each story was 9.3 ± 16.6 related tweets (range: 0 - 89). The number of selected tweets per type of misinformation is shown in the appendix (Table A9.3.1).

5.4.2 Is the discussion about misinformation related to emotion?

All tweets discussing the misinformation claims taken together had an average compound score of -0.0151. Most tweets were neutral (86.15%) and the proportion of positive (6.85%) or negative (6.99%) valenced tweets was similarly small. The neutral tone of the tweets could possibly be explained with the reasoning that the tweets often stated facts without adding emotional words. Additionally, the positively and negatively loaded tweets balance each other out in the entire dataset. Following this line of reasoning, certain types of misinformation could differ from a neutral sentiment by evoking one predominant feeling. To test this hypothesis, we analyzed the sentiments of the six different types of misinformation.

5.4.3 What are the sentiments in the different topics?

As shown in figure 1, the number of analyzed COVID-19-related misinformation tweets for each of the six types was the following: cures: n = 613, virus: n = 426, vaccine: n = 196, politics: n = 215, conspiracy: n = 460, and other: n = 187. While the sample sizes were not equal across the types, the statistical tests used (i.e. permutation testing) do not suffer from sample size bias, and the results therefore preserve statistical validity.

The compound scores showed a difference in valence for the different types of misinformation (see table 9.3.2 in the appendix). Namely, the types 'virus' and 'conspiracy' had a negative compound score (-.124 and -.098, respectively), meaning that they had a negative valence. Both types significantly differed from the other types of misinformation, which had a positive compound score and hence positive valence, with 'cures' being most positive (.073), 'other' and 'vaccine' slightly less positive (.050 and 0.54 respectively) and 'politics' almost neutral (.007).

Conspiracy-related misinformation might be more negative than misinformation related to other types of misinformation since, especially shortly after dramatic events, conspiracies elicit a negative emotional response, a higher emotionality in dramatic situations potentially drives people towards conspiracies and, in general, emotions contribute to the spreading of conspiracies (Samory & Mitra, 2018; Sunstein & Vermeule, 2009). Misinformation related to virus characteristics and numbers are probably associated with a high uncertainty and especially in the beginning of the pandemic also fuel anxieties about a potentially deadly disease and are therefore especially negative.

Misinformed claims about false cures for COVID-19 are potentially damaging to society by fostering reckless behavior and thereby advancing the spread of the disease. Nonetheless, the associated sentiment with potential cures and vaccines for COVID-19 had an overall positive valence. The tweets with positive valence contained words related to hope (for example: help, treat, progress). Even though this type of misinformation expressed positive emotions, the effect on society is not necessarily positive but probably rather negative as it can enhance transmission of the disease.

The tweets related to politics were about government regulations mainly for curbing the spreading of COVID-19. Here the limiting consequences in people's daily lives in combination with a feeling of hope for things to get better as a result, or the fact that most regulations stated were just reported as facts, might explain the neutral valence.

Continuing the line of the previous reasoning, positive sentiments relating to hope for cures would naturally produce stronger positive sentiments than the rather neutral fact-relating sentiments for political measures.

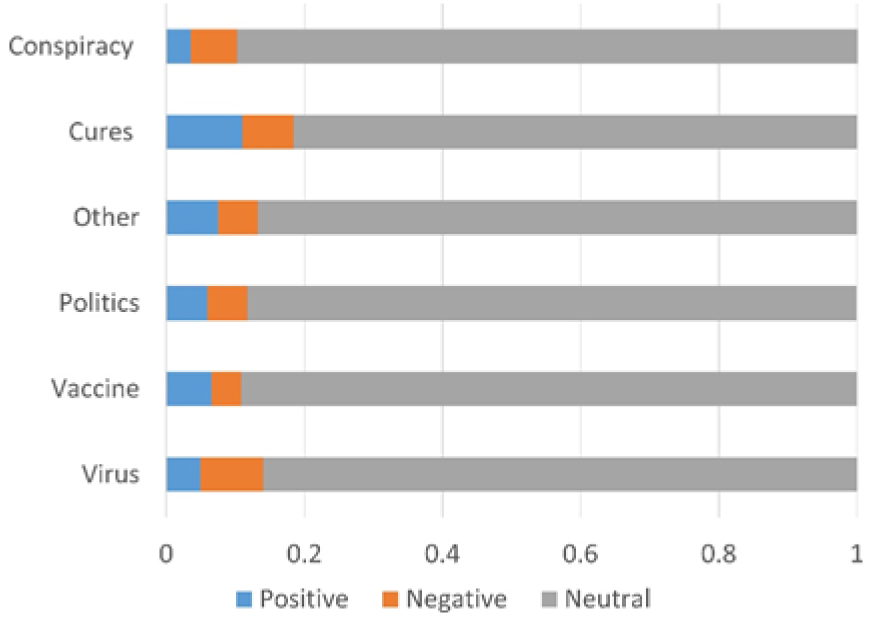


Figure 2. Average for the sentiment valence within each type of misinformation

To have a clearer overview of the factors driving the results of the compound scores, we computed the mean for positive, negative and neutral scores for each of the types of misinformation (Figure 2). For every type of misinformation, the neutral score was above 0.82, highlighting the predominance of neutral language independently of misinformation type. However, when using the most common threshold value to classify sentences as neutral (compound score < .05 or > -.05) as stated by (Hutto & Gilbert, 2014), only politics would qualify as a neutral type of misinformation based on its average compound score. This indicates that, despite the high neutral scores, all the other types of misinformation contained some positive and/or negative words that resulted in the differences between types of misinformation. Regarding misinformation around conspiracy theories and virus characteristics and numbers, we observed in both cases that the language was still mainly neutral, but that the ratios for positive and negative words showed a more pronounced imbalance than for the other types of misinformation. Furthermore, we observed that misinformation around cures contained a relatively higher percentage of both positive and negative words than other types, but that they tended more towards positive words.

5.5 Discussion and conclusion

In this study we showed how the emotional valence of COVID-19 related misinformation within Twitter differed by type of misinformation, thus looking at health-related misinformation, emotions and adding the aspect of types of misinformation. We identified 2097 tweets related to misinformation that was debunked by fact-checkers and accessible via the Google Fact Check Explorer. Within the tweets we analyzed, we found that approximately 29% were related to misinformation, however, extrapolating this finding to our starting sample lead to the estimation that only a very small number of the tweets related to misinformation. Overall, the tweets included in the final dataset, did not show a clear positive or negative emotional valence, but only a slight tendency towards a negative emotional valence. However, looking at different types of misinformation, we found significant differences. Communication around misinformation related to “conspiracy” and “virus characteristics and numbers” was characterized by stronger negative emotional valence than misinformation related to “cures, prevention and treatment”, “vaccine and test kits”, “political measures” or “other”.

Our findings support the argument for a more differentiated analysis of COVID-19 related misinformation. We suggest that strategies for fighting COVID-19 misinformation should focus on a fast response to misinformation regarding conspiracies and virus characteristics, as well as on reported numbers, given that previous research has shown that negative (mis)information spreads faster (cp. Vosoughi et al., 2018). However, since Vosoughi et al. (2018) focus on political misinformation, the relationship between emotional valence and the spreading of misinformation should be analyzed in more detail for health-related misinformation as well. Furthermore, our findings showed that communication about cures, prevention and treatments has a more positive emotional valence. Future studies should investigate the link between misinformation with positive emotional valence and its consequences for society - i.e., is misinformation with a positive language, which potentially elicits positive emotions, particularly dangerous because it leads to health risking behavior?

Overall, our study provides evidence for differences in emotional valence of the different types of COVID-19-related misinformation, highlighting the importance of investigating sentiment within meaningful clusters of misinformation. Our results, together with insights from previous research showing that negative misinformation spreads faster (Vosoughi et al., 2018), suggest focusing on misinformation around virus characteristics and conspiracies should be prioritized when combating misinformation. This includes a focus of policies, information campaigns or projects on these topics. Furthermore, in-depth analyses on the emotional characteristics of misinformation, and how these affect society, could further help to optimize strategies aiming to counteract the spread of misinformation.

6 Deep dive into privacy

This chapter is a deep dive into privacy on social media and relates to the human right key principle 'privacy and data protection' (see introduction - then ten key rights and principles of human rights online). This section contains an extract of the NORDMEDIA 2021 conference paper *Privacy settings cannot be predicted from Facebook Group topics, but gender ratio can: a large-scale study of open, closed and secret Facebook Groups* that is being submitted for publication in an academic journal. In the paper we outline how topics based on content in Facebook Groups cannot predict privacy settings and subsequently discuss the implications of this finding.

6.1 Background, research questions and hypotheses

Facebook Groups provide a place where users have the possibility of being exposed to information from topic-related or network-related, stronger or weaker ties. Facebook Groups are increasing in popularity, and not just among users. Facebook news feed algorithms have even changed to give priority to groups over Facebook pages. In a political climate where social cohesion yet again is an issue in debates about increasing polarization in relation to exactly Facebook as the dominating international media platform, this article zooms in on studying Facebook Groups at a larger scale.

Despite the overwhelming number of users of Facebook Groups, we know very little scientifically, if anything, on whether there is a strong correlation between privacy settings and the content of the groups. Thus, the research question is: can privacy settings be predicted from group topics (extracted from the content in the group) across open, closed and secret groups? The GDPR (General Data Privacy Regulation in EU) states that certain content is more sensitive than other content (such as political, religious, sexual and health content), which should therefore be protected regarding privacy. We therefore expect that topics can predict privacy settings in groups. If this is the case, sensitive content is protected sufficiently, if this is not the case, privacy settings in Facebook groups should be regulated more strictly. More generally, privacy settings could be controlled by algorithms to protect sensitive content on discussion platforms.

Why Denmark?

We have chosen Denmark to study Facebook groups because it is one of the countries with the highest internet and Facebook penetration. Our sample of 1,000 Danish Facebook users was drawn in April 2014 and data collection resulted in a collection of information about a total of 14,608 groups including the metadata of these groups and all communication within them starting from the date of creation to the date of data collection. The retrieved total number of active communicating users in the groups are 2,884,911 in all 14,608 groups.

76,7% of the Danish population above 12 years old has a social media profile and 69% (16-89 years olds) visit social media at least once daily. There is a small gender difference on Facebook use with 73% of females (18-89 years old) and 64% of males in this age group connect to Facebook daily. We focus on Facebook since - after entering the market in 2006 - it is consistently by far the most used social media site in Denmark with 97% of all social media users using Facebook in 2015 (Ministry of Culture, 2015). In an international comparison, Danes are - according to a survey reported in 2011 - the most active sharers in the EU on social network sites. 37% have shared photos, videos, and music compared to the average of 22% in the EU (Statistics Denmark, 2011 p. 24; Bechmann, 2014).

Danes being involved to this extent in social media platforms and Facebook in particular, enables us to analyze a diverse and active sample.

6.2 Theory

Facebook Groups have been the focus of several existing studies. However, most of these studies are limited in scope. Qualitative and computational studies which aim at understanding for example communication within open groups often focus on a specific topic, e.g. health (Al Mamun et al., 2015; Bender et al., 2011), education (De Villiers, 2010) or politics (Fernandes et al., 2010; Marichal, 2013; Woolley et al., 2010) and often rely on information of public groups. Other studies rely on survey data (Park et al., 2009) or are conceptualized as ethnographic studies (e.g., Miller, 2011) in order to add to our understanding of Facebook Groups. For example, findings are that also sensitive information about health is shared (Asiri et al., 2016). Al Mamun et al. (2015) shows that for the specific topic of hypertension most observed groups are global, do not serve commercial purposes and around a third does not show a high amount of activity. These studies contribute to a better understanding of a certain type of groups and the relationship between users within these groups. But they do not provide us with a more holistic understanding of the general topics discussed within Facebook Groups. Our aim is to add to our understanding of Facebook Groups and to map topics of Facebook Groups and their relation to other characteristics of these groups. This will lead to a more nuanced understanding of Facebook usage and ultimately contribute to the discussion on digital sociology and Facebook as a central platform for a large portion of communication in our everyday lives. Besides mapping topics of Facebook Groups, we also want to link them with the communicative structures and composition of Facebook Groups.

Interface-related communicative structure – privacy settings

Since its foundation, Facebook has been a growing social networking site (SNS) not only regarding the number of users but also concerning the number of provided features and the merge with other platforms such as Instagram or WhatsApp. This expansion also leads to new privacy concerns as it enables data merging across different sites. Users therefore face new decisions regarding privacy and are probably more aware about privacy issues. In the EU, the exposure to the discussion and changes associated with the introduction of the new GDPR in 2018 probably increased awareness about privacy as well. We observe, for example, that some scholars find that less information is shared on Facebook over time (Fiesler et al., 2017). This trend occurs even though individuals are not entirely in control of or aware of their privacy settings and its consequences, that is about with whom and which information they share (Acquisti et al., 2015). Facebook established settings in which you can choose between different degrees of sharing information - from public to "friends only". Facebook changed default settings over time with "friends only" being the default since 2014 (Mondal et al., 2019). In general, Facebook encourages its users to disclose information and to share feelings, activities and thoughts (Aharony, 2016). One feature of Facebook are Facebook Groups that aim at facilitating communication about shared interests among their members and friends (Chu, 2011). Here the privacy settings differ between "secret" (members only can access the group and its posts), "closed" (everyone can see the group but only members see posts) or "open" (all content is public). Furthermore, access to the groups can be restricted by the creator of the group (free access, access upon approval, access via invitation) (Chu, 2011). In this paper, we want to examine how Facebook Groups are characterized by privacy settings. Are privacy settings for example related to group topics?

Some studies already provide some insights about patterns of privacy setting use. Raynes-Goldie's (Raynes-Goldie, 2010) study of Facebook points to a greater concern with social privacy than institutional privacy. That is, participants were not concerned about how Facebook would use their data, but they were concerned about controlling their data towards their (potential) circle of friends. Additionally, Marwick and Boyd (Marwick & boyd, 2014) demonstrate how Facebook users perform social privacy as social norms of sharing and hiding information, for instance through encoded messages. Most of these studies focus on the individuals' own privacy choices and not on privacy settings of Facebook Groups. However, they allow for some conclusions regarding Facebook Groups as well. There exist several explanations about why people disclose information about themselves on social media platforms. They are divided into interpersonal goals such as social approval, social control or intimacy related issues and intrinsic goals such as the need for identity clarification and distress relief (Andalibi et al., 2017). Other studies point out the importance of privacy concerns and trust for choices regarding privacy settings (Gupta & Dhami, 2015). These goals express themselves differently in the use of different Facebook features with e.g. social approval being more prominent in public posts than in messages or wall posts (Andalibi et al., 2017; see also Hollenbaugh & Ferris, 2014). Differences in privacy behavior and concerns are also found based on demographic characteristics such as gender (Chakraborty et al., 2013; Lewis et al., 2008; Reynolds et al., 2011) and age (Aharony, 2016; Aljohani et al., 2016; Fiesler et al., 2017). Furthermore, attitudes about and own privacy preferences are not always linked with actual behavior (Acquisti et al., 2015; Reynolds et al., 2011), making it more difficult to analyze privacy settings via survey data. Besides individual characteristics, also context characteristics affect behavior. For example, cultural differences related to privacy issues can be observed (L. Chen & Tsoi, 2011; Nemati et al., 2014). Studies also show that behavior in their networks affects individuals' privacy behavior and attitudes (Acquisti et al., 2015; Lewis et al., 2008). In addition, the platforms themselves affect privacy behavior e.g. via their default settings (Acquisti et al., 2015). Studies so far focus on posting behavior of individuals and not Facebook Groups. Regarding groups, Bechmann (2014) shows how young Danes choose groups as their main privacy filter to avoid data being shared with all their friends on Facebook and friends in other Facebook connected services. This finding indicates that group privacy settings matter and are used deliberately. We assume that also regarding groups privacy settings networks play a role as can be observed for individual privacy settings. The group creator might take settings of similar groups as a benchmark. Furthermore, groups with similar topics have similar purposes as well (see for hypertension Al Mamun et al., 2015).

We therefore assume to find a link between group privacy and group topic.

6.2 Methods

6.2.1 Dataset

We have chosen Denmark to study Facebook Groups because it is one of the countries with the highest internet and Facebook penetration. Our sample of 1,000 Danish Facebook users was recruited through Userneeds to mirror the Danish internet population stratified on age, gender, education and area of residence. Data was drawn with first degree informed consent in April 2014 and the data collection resulted in a collection of information about a total of 13,672 groups including the metadata of these groups and all communication within them starting from the date of creation to the date of data collection. The total number of active communicating users in the groups are 3,981,950 in all 14,608 groups.

76,7% of the Danish population above 12 years old has a social media profile and 69% (16-89 years old's) visit social media at least once daily. There is a small gender difference on Facebook use with 73% of females (18-89 years old) and 64% of males in this age group connect to Facebook daily. We focus on Facebook since - after entering the market in 2006 - it is consistently by far the most used social media site in Denmark with 97% of all social media users using Facebook in 2015 (Ministry of Culture, 2015). In an international comparison, Danes are - according to a survey reported in 2011 - the most active sharers in the EU on social network sites. 37% have shared photos, videos, and music compared to the average of 22% in the EU (Statistics Denmark, 2011 p. 24; Bechmann, 2014). Danes being involved to this extent in social media platforms and Facebook in particular, enables us to analyze a diverse and active sample.

6.2.2 Analysis

We computed a latent semantic model (LDA) across post and comments in all open, closed and secret group to identify latent variables which can be interpreted as 'group topics' - which leads to the identification of 50 latent topics. We then looked at the relation between group topics and privacy settings. 1000 samples were created and before running each model, data in every sample were split into train (70% of the data) and validation set (30% of the data). As the dependent variable, group privacy, were three-level factors, the classifier chosen was a multinomial logistic regression.

We computed which model predicting privacy settings is the most optimal based on the Akaike Information Criteria (AIC). This measure returns the explanatory power of the model with a penalty for overfitting. We controlled for group duration and group size (number of posts and comments). Regarding privacy settings, all models perform worse, with the one, which included only topic as the independent variable, performing best. We also examined the mean accuracy of all models.

We wished to investigate how each individual topic was related to each of the privacy settings. To examine this, we applied the trained model to a dataset which only contained one of the 50 topics, and instead of extracting the final predicted class we extracted the probabilities for all of the classes. . By this we could see the connection between each topic and each privacy setting in isolation. This process was repeated for every topic within every sample out of all 1000 samples. We extracted the mean probability of all topics and privacy setting combinations.

6.3 Results

The results of the tests on the ability of the topics to predict privacy setting level of the group are presented in Table 1. Our analyses show that the models for the prediction of privacy settings are not very accurate (around 40%), indicating that there is only a weak link between privacy settings and group topics.

Table 1. Accuracies and standard deviations

Model	Accuracy (empirical / random classifier)	Standard deviation
Privacy settings ~ topic	40.23 % / 33.32 %	2.47 / 1.16

Note: All models were tested against a baseline model being a random classifier where privacy settings was randomly predicted; mean accuracy and its standard deviation

Table 2 reports the topics for public, closed and secret Facebook Groups with the highest probabilities. From the 50 latent topics 20 are most likely discussed in open groups, 26 topics are most likely assigned to closed groups and the remaining 4 topics cannot be determined (balanced equally). None of the latent topics is most likely discussed in secret groups. Table 2 reports the three topics with the highest probability for each group privacy setting. Regarding open groups we find that “politics”, “trading” and “gardening” are most likely discussed in this kind of group. Regarding closed groups we find that “security”, “horse riding” and “international” are most likely discussed in this kind of group. The topics which have the highest probability to be discussed in secret groups are “business”, “news” and “fans”. However, the probability to be discussed in open or closed groups for these topics is still higher. These results must be interpreted carefully, as topics are not good predictors for privacy settings in itself.

Table 2: predicted probabilities for being open, closed or secret group by topics

Topic	Open	Closed	Secret	Most common nouns English translation
trading	0.78 (0.19)	0.18 (0.15)	0.05 (0.05)	Seeking, oa[i], boy, girl, picture, condition, year, sale, price, alot
politics	0.73 (0.15)	0.19 (0.12)	0.08 (0.08)	Politics, year, government, party, society, debate, share, people, member, election
gardening	0.71 (0.16)	0.37 (0.15)	0.02 (0.02)	year, flower, picture, plants, couple, thank you, idea, time, soil, advice
security	0.20 (0.16)	0.72 (0.21)	0.08 (0.06)	guard, our, interest, all, hat, euro, mask, photo, politics, case
horseback riding	0.20 (0.17)	0.72 (0.22)	0.08 (0.06)	Horse, year, dressage, leap, (saddle) pad, price, pony, stable, picture, horse show
international/travel[ii]	0.25 (0.20)	0.70 (0.21)	0.05 (0.04)	Share, events, page, ref, link, eid, euro, text message, mail
business	0.12 (0.09)	0.52 (0.14)	0.35 (0.14)	Party, form, link, customs, interview, petition, country, condition, feature, list
personal news	0.34 (0.22)	0.35 (0.20)	0.31 (0.21)	Somebody, year, tip, news, verdict, share, day, group, trip
fans	0.48 (0.24)	0.22 (0.16)	0.30 (0.22)	Fans, start, page, teams, form, summe, customs, rest, post, sport

Note: Standard deviation in parentheses; assigned topic labels

6.3 Discussion and conclusion

Our analyses of Facebook Groups show that 50 latent group topics can be distinguished and the use of group topics to predict privacy settings is undeniably weak as it is well below chance. The implications of such a weak connection between group topics (based on the content) and the privacy setting are that we cannot assume that because something is set to be private that it then contains a private topic. In a profiling context opening groups for third party data brokers thus may result in unintentional data leakage and on the other hand we cannot rely on privacy settings that are not aligned with the content in regulatory contexts such as GDPR if the aim is to have a general understanding of what type of content is to be regarded as sensitive (e.g., political standpoint, sexual orientation and religion) instead of what can be expected using privacy settings as a signal indicating this from the platform's side.

The result that we cannot predict privacy setting from the topic/content type of the group does not immediately correspond with people using more closed groups as a way to control the visibility and circulation of content (Bechmann, 2014), but on the other hand such existing studies do not indicate in details what type of content goes where. At the same time the lack of prediction power in topics when it comes to privacy settings may support existing studies suggesting default settings on Facebook to be powerful and controlling for the end result because the users do not change this setting (Stutzman et al., 2013; Van Dijck, 2013). However, when we measure the mean of the number of open (619), closed (663) and secret (217) groups these numbers do not indicate an overwhelming use of closed groups (default setting) compared to open groups. A third explanation for the weak prediction may be due to people either disagreeing on what is private or that they simply follow the administrators choice in a privacy paradoxical fashion (Barnes, 2006; Nissenbaum, 2009, 2011) where convenience or the communicating with peers on a certain topic is more important than the privacy setting is and that this setting may be indicated by an administrator that has other incentives such as heighten visibility (Bucher, 2012; Nahon & Hemsley, 2013). This in turn can explain the high number of open groups.

A limitation of the study is that the dataset was constructed in 2014 and trace data access has been closed down since then to private groups. This means that we are not able to continuously account for the current and future effects algorithmic changes have on the predictive power of privacy settings. In other words, if Facebook chooses to provide an algorithm that is more sensitive to the content and changes therein, then the predictive power of topics on privacy settings may increase as a way to nudge users to align these two aspects. This can be done through more intelligent machine learning models that detect topics and suggest setting accordingly on a running basis so that setting also can change over time if the content or for instance the size of the group changes. At the time of writing, we have no accounts of this having taken place. However, future studies are needed to continuously understand the predictive power due to the potential effect of algorithmic adjustments but also changes in the platform and general media landscape that can affect the way Facebook Groups are used as well. For now, the study has shown that we cannot treat privacy settings as a proxy for the type of content behind.

7 Conclusion

We have presented several analyses of discussions on the social media platforms Reddit, Twitter and Facebook. Our analysis indicates that privacy is both a trending and much discussed topic, which gives rise to concern and negative emotions. Similarly, cyber security is a related topic and also showed up several times in our analysis. These topics should be a focus point in further steps taken towards an improved next generation internet. Our deep dive into privacy points to the fact that discussions in Facebook groups might not be sufficiently protected against unintentional data leaks. Other discussed topics were related to technology (cryptocurrency, hacking, artificial intelligence, machine learning, big data), while other issues were focused on the societal impact (censorship, having alternative choices, communication, and business opportunities). We showed how algorithms can display gender bias based on the selected dataset, which underlines that choices made by algorithm developers should account for potential biases in the dataset beforehand. In recent years, disinformation on social media has become an issue, and the types of disinformation that elicit negative emotions should be a target to curb the spread of false information. Overall, our analysis showed that technological developments are accompanied by various social issues, and that several issues are discussed with negative emotions and concern. The next generation internet can implement targeted legal regulations, provide alternative choices, and enhance awareness about certain issues in order to provide a more citizen-centered internet in the future.

8 References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Aharony, N. (2016). Relationships among attachment theory, social capital perspective, personality characteristics, and Facebook self-disclosure. *Aslib Journal of Information Management*.

Al Mamun, M., Ibrahim, H. M., & Turin, T. C. (2015). Social Media in Communicating Health Information: An Analysis of Facebook Groups Related to Hypertension. *Preventing Chronic Disease*, *12*, 140265. https://doi.org/10.5888/pcd12.140265

Aljohani, M., Nisbet, A., & Blincoe, K. (2016). *A survey of social media users privacy settings &amp; information disclosure* [PDF]. https://doi.org/10.4225/75/58A693DEEE893

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.

Andalibi, N., Ozturk, P., & Forte, A. (2017). Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1485–1500. https://doi.org/10.1145/2998181.2998243

Asiri, E., Khalifa, M., Shabir, S.-A., Hossain, M. N., Iqbal, U., & Househ, M. (2016). Sharing sensitive health information through social media in the Arab world. *International Journal for Quality in Health Care*, intqhc;mzw137v1. https://doi.org/10.1093/intqhc/mzw137

Bailey, J., Steeves, V., Burkell, J., & Regan, P. (2013). Negotiating with gender stereotypes on social networking sites: From “bicycle face” to Facebook. *Journal of Communication Inquiry*, *37*(2), 91–112.

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*. https://doi.org/10.5210/fm.v11i9.1394

Bechmann, A. (2014). Managing the interoperable self. In *The ubiquitous Internet* (pp. 66–85). Routledge.

Bechmann, A. (2017). Keeping it real: From faces and features to social values in deep learning algorithms on social media images. *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, *6*(1), 2053951718819569.

Bechmann, A., & Vahlstrup, P. B. (2015). Studying Facebook and Instagram data: The Digital

Footprints software. *First Monday*, *20*(12), 1–13. https://doi.org/10.5210/fm.v20i12.5968

Bender, J. L., Jimenez-Marroquin, M.-C., & Jadad, A. R. (2011). Seeking Support on Facebook: A Content Analysis of Breast Cancer Groups. *Journal of Medical Internet Research*, *13*(1), e16. https://doi.org/10.2196/jmir.1560

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, *29*, 4349–4357.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). *Types, Sources, and Claims of COVID-19 Misinformation* (p. 13). Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Chakraborty, R., Vishik, C., & Rao, H. R. (2013). Privacy preserving actions of older adults on social media: Exploring the behavior of opting out of information sharing. *Decision Support Systems*, *55*(4), 948–956. https://doi.org/10.1016/j.dss.2013.01.004

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance*, *6*(2), e19273. https://doi.org/10.2196/19273

Chen, L., & Tsoi, H. K. (2011). Privacy Concern and Trust in Using Social Network Sites: A Comparison between French and Chinese Users. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2011* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-642-23765-2_16

Cheney-Lippold, J. (2017). We Are Data. In *We Are Data*. New York University Press.

https://www.degruyter.com/document/doi/10.18574/9781479888702/html

Chu, S.-C. (2011). Viral Advertising in Social Media: Participation in Facebook Groups and Responses among College-Aged Users. *Journal of Interactive Advertising*, *12*(1), 30–43. https://doi.org/10.1080/15252019.2011.10722189

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Wash. L. Rev.*, *89*, 1.

Cuddy, A. J., Crotty, S., Chong, J., & Norton, M. I. (2010). Men as cultural ideals: How culture shapes gender stereotypes. *Boston, MA: Harvard Business School*.

Cuddy, A. J., Wolf, E. B., Glick, P., Crotty, S., Chong, J., & Norton, M. I. (2015). Men as cultural ideals: Cultural values moderate gender stereotype content. *Journal of Personality and Social Psychology*, *109*(4), 622.

De Villiers, M. R. (2010). *Academic use of a group on Facebook: Initial findings and perceptions*.

Desilver, D. (2016). *5 facts about Twitter at age 10 | Pew Research Center*. https://www.pewresearch.org/fact-tank/2016/03/18/5-facts-about-twitter-at-age-10/

Eagly, A. H. (1997). *Sex differences in social behavior: Comparing social role theory and evolutionary psychology.*

Eisenchlas, S. A. (2013). Gender Roles and Expectations: Any Changes Online? *SAGE Open*, *3*(4), 2158244013506446. https://doi.org/10.1177/2158244013506446

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, *85*(1), 57–80.

Eubanks, V. (n.d.). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. HeinOnline.

Fernandes, J., Giurcanu, M., Bowers, K. W., & Neely, J. C. (2010). The Writing on the Wall: A Content Analysis of College Students' Facebook Groups for the 2008 Presidential Election. *Mass Communication and Society*, *13*(5), 653–675. https://doi.org/10.1080/15205436.2010.516865

Fiesler, C., Dye, M., Feuston, J. L., Hiruncharoenvate, C., Hutto, C. J., Morrison, S., Khanipour Roshan, P., Pavalanathan, U., Bruckman, A. S., De Choudhury, M., & Gilbert, E. (2017). What (or Who) Is Public?: Privacy Settings and Social Media Content Sharing. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 567–580. https://doi.org/10.1145/2998181.2998223

Frison, E., & Eggermont, S. (2016). Gender and Facebook motives as predictors of specific types of Facebook use: A latent growth curve analysis in adolescence. *Journal of Adolescence*, *52*, 182–190.

Guadagno, R. E., Muscanell, N. L., Okdie, B. M., Burk, N. M., & Ward, T. B. (2011). Even in virtual environments women shop and men build: A social role perspective on

Second Life. *Computers in Human Behavior*, *27*(1), 304–308. https://doi.org/10.1016/j.chb.2010.08.008

Gupta, A., & Dhami, A. (2015). Measuring the impact of security, trust and privacy in information sharing: A study on social networking sites. *Journal of Direct, Data and Digital Marketing Practice*, *17*(1), 43–53. https://doi.org/10.1057/dddmp.2015.32

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *ArXiv:1512.03385 [Cs]*. http://arxiv.org/abs/1512.03385

Hollenbaugh, E. E., & Ferris, A. L. (2014). Facebook self-disclosure: Examining the role of traits, social cohesion, and motives. *Computers in Human Behavior*, *30*, 50–58. https://doi.org/10.1016/j.chb.2013.07.055

Honnibal, M., & Montani, I. (2017). Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *Unpublished Software Application. Https://Spacy. Io*.

Horne, B. D., Adali, S., & Sikdar, S. (2017). Identifying the social signals that drive online discussions: A case study of Reddit communities. *ArXiv:1705.02673 [Cs]*. http://arxiv.org/abs/1705.02673

Howard, P. N. (2006). *New media campaigns and the managed citizen*. Cambridge University Press.

Hutto, C. J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. 10.

Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E., & Baddour, K. (2020). Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*. https://doi.org/10.7759/cureus.7255

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 591. https://doi.org/10.1145/1772690.1772751

Levin, S. (2016). A beauty contest was judged by AI and the robots didn't like dark skin. *The Guardian*, *9*.

Lewis, K., Kaufman, J., & Christakis, N. (2008). The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. *Journal of Computer-Mediated Communication*, *14*(1), 79–100. https://doi.org/10.1111/j.1083-6101.2008.01432.x

Li, X., Zhou, M., Wu, J., Yuan, A., Wu, F., & Li, J. (2020). Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions. *ArXiv:2005.14464 [Cs]*. http://arxiv.org/abs/2005.14464

Madani, A., Boussaid, O., & Zegour, D. E. (2014). What's Happening: A Survey of Tweets Event Detection. *ICC 2014*.

Makashvili, M., Ujmajuridze, B., & Amirejibi, T. (2013). Gender Difference in the Motives for the Use of Facebook. *Asian Journal for Humanities and Social Studies (AJHSS)*, *1*(03), 130–135.

Marichal, J. (2013). Political Facebook groups: Micro-activism and the digital front stage. *First Monday*, *18*(12). https://doi.org/10.5210/fm.v18i12.4653

Marwick, A. E., & boyd, danah. (2014). *Networked privacy: How teenagers negotiate context in social media—Alice E Marwick, danah boyd, 2014*. https://journals.sagepub.com/doi/10.1177/1461444814543995

McAndrew, F. T., & Jeong, H. S. (2012). Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use. *Computers in Human Behavior*, *28*(6), 2359–2365. https://doi.org/10.1016/j.chb.2012.07.007

Miller, D. (2011). *Få Tales from Facebook af Daniel Miller som Paperback bog på engelsk—9780745652108*. SAXO.com. https://www.saxo.com/dk/tales-from-facebook_daniel-miller_paperback_9780745652108

Mondal, M., Yilmaz, G. S., Hirsch, N., Khan, M. T., Tang, M., Tran, C., Kanich, C., Ur, B., & Zheleva, E. (2019). Moving Beyond Set-It-And-Forget-It Privacy Settings on Social Media. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 991–1008. https://doi.org/10.1145/3319535.3354202

Muscanell, N. L., & Guadagno, R. E. (2012). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*, *28*(1), 107–112.

Nahon, K., & Hemsley, J. (2013). *Going viral*. Polity.

Nemati, H., Wall, J. D., & Chow, A. (2014). Privacy Coping and Information-Sharing Behaviors in Social Media: A Comparison of Chinese and U.S. Users. *Journal of Global Information Technology Management*, *17*(4), 228–249. https://doi.org/10.1080/1097198X.2014.978622

Nielbo, K. L., Vahlstrup, P. B., Gao, J., & Bechmann, A. (2019). *Sociocultural trend signatures in minimal persistence and past novelty*. Manuscript submitted for publication.

Nissenbaum, H. (2009). *Privacy in Context: Helen Nissenbaum: 9780804752374*. https://www.bookdepository.com/Privacy-in-Context-Helen-Nissenbaum/9780804752374?redirected=true&utm_medium=Google&utm_campaign=Base2&utm_source=DK&utm_content=Privacy-in-Context&selectCurrency=DKK&w=AF4ZAU9SB230V3A8038M&pdg=kwd-293946777986:cmp-1597361031:adg-58873824845:crv-303010908953:pid-9780804752374:dev-c&gclid=CjwKCAiA_P3jBRAqEiwAZyWWaGes5fE-4mkm7tgbEDU4MBnKIC-DZVSGqqAyo6IOZuf-iu098bnCUBoCBisQAvD_BwE

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, *140*(4), 32–48.

Oberst, U., Renau, V., Chamarro, A., & Carbonell, X. (2016). Gender stereotypes in Facebook profiles: Are women more female online? *Computers in Human Behavior*, *60*, 559–564.

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Otterbacher, J., Bates, J., & Clough, P. (2017). Competent men and warm women: Gender stereotypes and backlash in image search results. *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems*, 6620–6631.

Park, G., Yaden, D. B., Schwartz, H. A., Kern, M. L., Eichstaedt, J. C., Kosinski, M., Stillwell, D., Ungar, L. H., & Seligman, M. E. P. (2016). Women are Warmer but No Less Assertive than Men: Gender and Language on Facebook. *PLOS ONE*, *11*(5), e0155885. https://doi.org/10.1371/journal.pone.0155885

Park, N., Kee, K. F., & Valenzuela, S. (2009). Being Immersed in Social Networking Environment: Facebook Groups, Uses and Gratifications, and Social Outcomes. *CyberPsychology & Behavior*, *12*(6), 729–733. https://doi.org/10.1089/cpb.2009.0003

Raynes-Goldie, K. (2010). Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday*.

Reynolds, B., Venkatanathan, J., Gonçalves, J., & Kostakos, V. (2011). Sharing Ephemeral Information in Online Social Networks: Privacy Perceptions and Behaviours. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2011* (pp. 204–215). Springer. https://doi.org/10.1007/978-3-642-23765-2_14

Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench—A benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1), 23. https://doi.org/10.1140/epjds/s13688-016-0085-1

Rose, J., Mackey-Kallis, S., Shyles, L., Barry, K., Biagini, D., Hart, C., & Jack, L. (2012). Face it: The Impact of Gender on Social Media Images. *Communication Quarterly*, *60*(5), 588–607. https://doi.org/10.1080/01463373.2012.725005

Samory, M., & Mitra, T. (2018). *Conspiracies Online: User Discussions in a Conspiracy Community Following Dramatic Events*. 10.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). Automation, algorithms, and politics| when the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. *International Journal of Communication*, *10*, 19.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, *128*(2), 336–359.

https://doi.org/10.1007/s11263-019-01228-7

Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., & Wang, Y. (2020). A first look at COVID-19 information and misinformation sharing on Twitter. *ArXiv Preprint ArXiv:2003.13907*.

Stutzman, F. D., Gross, R., & Acquisti, A. (2013). Silent listeners: The evolution of privacy and disclosure on Facebook. *Journal of Privacy and Confidentiality*, *4*(2), 2.

Sunstein, C. R., & Vermeule, A. (2009). Conspiracy Theories: Causes and Cures*. *Journal of Political Philosophy*, *17*(2), 202–227. https://doi.org/10.1111/j.1467-9760.2008.00325.x

Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, *56*(5), 44–54.

Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.

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

Wojcik, S., & Hughes, A. (2019). Sizing up Twitter users. *Pew Research Center*, *24*.

Woolley, J. K., Limperos, A. M., & Oliver, M. B. (2010). The 2008 Presidential Election, 2.0: A Content Analysis of User-Generated Political Facebook Groups. *Mass Communication and Society*, *13*(5), 631–652. https://doi.org/10.1080/15205436.2010.516864

9 Appendices

9.1 Appendix for Trend detection

A9.1.1. Overview of the ten key rights and principles with an explanation, the shortened statements and the extracted keywords.

1. **Universality and equality**: all humans are born free and equal in dignity and rights, which must be respected, protected and fulfilled in the online environment

   Freedom and equality online

   1. Freedom online
   2. Equality online
2. **Rights and social justice**: The Internet is a space for the promotion, protection and fulfilment of human rights and the advancement of social justice. Everyone has the duty to respect the human rights of all others in the online environment

   Human rights and social justice online

   1. Human rights online
   2. Social justice online
3. **Accessibility**: Everyone has an equal right to access and use a secure and open Internet

   Access to secure and open internet

   1. Access to secure internet
   2. Access to open internet
4. **Expression and association**: Everyone has the right to seek, receive, and impart information freely on the Internet without censorship or other interference. Everyone also has the right to associate freely through and on the Internet, for social, political, cultural or other purposes

   No censorship, free association online

   1. Censorship online
   2. Free association online
5. **Privacy and data protection**: Everyone has the right to privacy online. This includes freedom from surveillance, the right to use encryption, and the right to online anonymity. Everyone also has the right to data protection, including control over personal data collection, retention, processing, disposal and disclosure

   Freedom from surveillance, encryption, online anonymity, data protection online

    1. Surveillance online
    2. Online encryption
    3. Online anonymity
    4. Data protection online

6. **Life, liberty and security**: The rights to life, liberty, and security must be respected, protected and fulfilled online. These rights must not be infringed upon, or used to infringe other rights, in the online environment

    Rights to life, liberty and security must be protected and respected online

    1. Right to life online
    2. Online liberty
    3. Online security

7. **Diversity**: Cultural and linguistic diversity on the Internet must be promoted, and technical and policy innovation should be encouraged to facilitate plurality of expression

    Cultural and linguistic diversity, and plurality of expression, must be promoted online

    1. Cultural diversity online
    2. Linguistic diversity online

8. **Network equality**: Everyone shall have universal and open access to the Internet's content, free from discriminatory prioritisation, filtering or traffic control on commercial, political or other grounds

    Universal and open access to the internet's content, free of discriminatory prioritisation, filtering or traffic control online

    1. Online content discrimination
    2. Online content filtering
    3. Online traffic control

9. **Standards and regulation**: The Internet's architecture, communication systems, and document and data formats shall be based on open standards that ensure complete interoperability, inclusion and equal opportunity for all

    Open standards for internet's architecture, communication systems and document and data formats online

    1. Internet architecture (open standards)
    2. Online communication systems (open standards)
    3. Online data format (open standards)

10. **Governance**: Human rights and social justice must form the legal and normative foundations upon which the Internet operates and is governed. This shall happen in a transparent and multilateral manner, based on principles of openness, inclusive participation and accountability

Openness, inclusive participation and accountability online for human rights and social justice

1. Inclusive participation online
2. Accountability online

A9.1.2. List of the ten key rights and principles and the selected subreddits.

1. **Universality and equality**
    1. Freedom online (4): r/InternetFreedom, r/FreeAsInFreedom, r/CryptoKiwis, r/degoogleyourlife
    2. Equality online: No relevant communities

2. **Rights and social Justice**
    1. Human rights online (2): r/AnonyNet, r/ACTA
    2. Social Justice online: No relevant communities

3. **Accessibility**
    1. Access to open internet (4): r/InternetFreedom**, r/InformationPolicy, r/internetdeclaration, r/opentheinternet
    2. Access to secure internet (2): r/internetdeclaration, r/CyberSec101

4. **Expression and association**
    1. Censorship online (4): r/yro, r/antisocialmedia, r/IdeasAreBeautiful, r/killingcensorship
    2. Free association online: No relevant communities

5. **Privacy and data protection**
    1. Surveillance online (6): r/yro **, r/privacytoolsIO, r/privacytools, r/thinkprivacy, r/snowden, r/IdeasAreBeautiful**
    2. Online encryption (3): r/CryptoKiwis**,r/CrpytoKiwis, r/privacytoolsIO**
    3. Online anonymity (7): r/privacy, r/conspiracy, r/technology **, r/cyberlaws, r/cyb3rs3c, r/TinfoilHatter, r/CryptoKiwis**
    4. Data protection online (1): r/CyberSec

6. **Life, liberty and security**

    1. Right to life online: No relevant communities
    2. Online liberty: No relevant communities
    3. Online security (10): r/privacy **, r/technology**, r/netsec, r/Bitcoin**, r/OnlineSecurity, r/onlinesecuritytips, r/Internet_Security, r/SmashingSecurity, r/ComputerSecurity, r/degoogleyourlife**

7. **Diversity**
    1. Cultural diversity online: No relevant communities
    2. Linguistic diversity online: No relevant communities

8. **Network equality**
    1. Online content discrimination: No relevant communities
    2. Online content filtering: No relevant communities
    3. Online traffic control: No relevant communities

9. **Standards and regulation**
    1. Internet architecture (3): r/technology**, r/turing_machines, r/InformationPolicy
    2. Online communication systems (2): r/Rad_Decentralization, r/Stellar
    3. Online data format: No relevant communities

10. **Governance**
    1. Inclusive participation online: No relevant communities
    2. Accountability online (1): r/iexec

** Subreddit appears more than once among the selection.

A9.1.3: Overview of the size of the subreddits

Subreddit name	Number of documents (posts and comments)	Analysis	Number of documents after preprocessing
ACTA	52	Not sufficient data	-
AnonyNet	3	Not sufficient data	-
antisocialmedia	186	Analyzed	185

Bitcoin	2,222,489	Analyzed	17,466 (17,572*)
ComputerSecurity	7,202	Analyzed	7,137
conspiracy	9,461,267	Processing error	-
CrpytoKiwis	6	Not sufficient data	-
CryptoKiwis	7	Not sufficient data	-
cyb3rs3c	8	Not sufficient data	-
cyberlaws	2,241	Analyzed	2,223
CyberSec	0	Not sufficient data	-
CyberSec101	175	Analyzed	171
degoogleyourlife	333	Analyzed	329
FreeAsInFreedom	880	Analyzed	873
IdeasAreBeautiful	0	Not sufficient data	-
iexec	1,400	Analyzed	1,376
InformationPolicy	1,066	Analyzed	1,066
Internet_Security	9	Not sufficient data	-
internetdeclaration	5	Not sufficient data	-
InternetFreedom	5	Not sufficient data	-
killingcensorship	31	Not sufficient data	-
netsec	64,796	Analyzed	15,206 (17,636*)
OnlineSecurity	19	Not sufficient data	-

onlinesecuritytips	0	Not sufficient data	-
opentheinternet	1	Not sufficient data	-
privacy	618,446	Analyzed	17,477 (17,585*)
privacytools	887	Analyzed	879
privacytoolsIO	215,389	Analyzed	17,360 (17,603*)
Rad_Decentralization	3,700	Analyzed	3,659
SmashingSecurity	1,958	Analyzed	1,938
Snowden	3,280	Analyzed	3,238
Stellar	96,264	Analyzed	15,336 (17,569*)
technology	3,628,038	Processing error	-
thinkprivacy	26	Not sufficient data	-
TinfoilHatter	0	Not sufficient data	-
turing_machines	24	Not sufficient data	-
yro	7	Not sufficient data	-

** Downsampled dataset size in parenthesis*

A9.1.4 Number of topics and top keywords for the 3 most representative topics for each subreddit

	Top keywords	
Topic	InformationPolicy (20 topics)	Antisocialmedia (50 topics)
1	cancel, new, culture, theresa, use, genuinely, see, study, library, status	ue, sud, social, medium, create, social, becus, destroi, platform, fair

2	twitter, digital, white, youtube, really, chinese, internet, david, communist, street	social, medium, platform, game, new, even, create, feel, wonder, time
3	twitter, political, internet, society, like, google, amp, industry, powerful, association	private, life, snap, thing, snapchat, value, keep, people, really, feel

	Top keywords	
Topic	CyberSec101 (80 topics)	Rad_Decentralization (20 topics)
1	anything, fund, sms, spouse, viber, yahoo, hey, purpose, recording, collect	use, like, want, one, make, need, way, bitcoin, would, user
2	device, thank, great, gmail, com, hacker, risk, china, recommend, work	decentralized, blockchain, use, want, one, people, get, right, well, thank
3	help, contact, hack, refer, whatsapp, computerguru, gmail, com, hacker, good	would, blockchain, gt, see, take, could, get, need, project, decentralized

	Top keywords	
Topic	SmashingSecurity (30 topics)	Privacytools (30 topics)
1	use, security, amp, password, go, fa, google, account, like, one	use, look, account, fa, would, google, really, question, share, login
2	podcast, episode, ve, thank, hacker, good, great, scam, think, call	find, matrix, change, privacy, phone, vpn, give, much, take, see
3	podcast, make, google, look, would, like, carole, new, think, episode	make, samsung, want, file, one, phone, use, store, etc, adguard

	Top keywords	
Topic	FreeAsInFreedom (30 topics)	Degoogleyourlife (80 topics)

1	read, gt, use, amp, come, police, speech, facial, recognition, make	management, open, source, access, privileged, mailbox, app, droid, encrypted, auroa
2	people, want, work, need, matter, use, issue, life, go, gab	opt, datum, facebook, simple, deep, share, company, google, link, deepspeech
3	use, smartphone, gt, people, one, rms, see, go, comment, forum	try, think, would, know, complexe, make, signal, friendly, hard, tho

	Top keywords	
Topic	Snowden (80 topics)	Netsec (3000 topics)
1	gt, use, know, re, say, guy, time, people, someone, get	information, security, portugal, brazil, blaze, service, security, publish, good, work
2	would, one, phone, america, go, us, get, step, reason, anti	company, year, get, would, find, experience, one, go, job, another
3	people, one, country, well, look, believe, like, could, right, use	attacker, use, know, challenger, doesn, car, security, linux, operating, system

	Top keywords	
Topic	Bitcoin (100 topics)	Privacy (1000 topics)
1	transaction, send, fee, long, cash, bank, lose, sure, start, actually	make, go, think, people, re, work, want, phone, gt, need
2	happen, transaction, value, put, lot, sure, high, gold, actually, guy	people, go, say, google, gt, make, even, also, re, want
3	transaction, send, fee, bank, sure, coinbase, high, long, lot, value	make, gt, see, also, datum, think, need, even, good, user

	Top keywords	
Topic	Stellar (3000 topics)	ComputerSecurity (20 topics)

1	buffett, meltdown, paychannels, abridgment, thee, maximalists, suckers, warren, fearful, cdp	use, remove, get, one, would, like, vpn, know, password, computer
2	fb, crypto, like, would, government, get, go, privacy, service, token	use, remove, drive, get, file, like, look, datum, router, program
3	stellar, network, use, xlm, gt, blockchain, make, ibm, public, know	remove, use, thank, computer, also, password, vpn, good, server, datum

	Top keywords	
Topic	Iexec (30 topics)	Cyberlaws (80 topics)
1	iexec, de, key, ico, dataset, coin, confidential, computing, remove, pump	bot, work, amp, right, notice, reg, remove, feel, runescape, owner
2	iexec, rlc, make, think, join, pool, get, isn, token, coin	market, threat, google, nso, get, phone, first, care, gt, case
3	rlc, well, bt, iexec, token, network, work, wonder, btcxbet, giv	law, cyber, computer, crime, like, lawyer, legal, would, help, issue

	Top keywords
Topic	privacytoolsIO (100 topics)
1	use, gt, app, privacy, like, would, get, one, need, people
2	use, like, gt, get, also, vpn, know, would, go, one
3	use, would, well, like, work, vpn, get, app, privacy, know

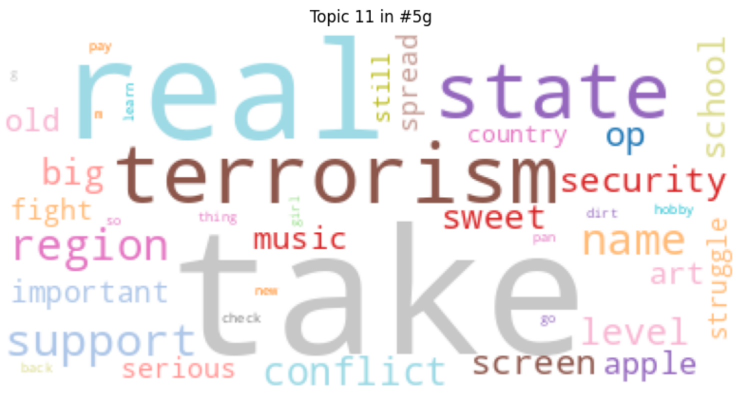
Note: the keywords ‘gt’ and ‘amp’ are probably not natural text but introduced in the process of data scraping and decoding/encoding steps during preprocessing.

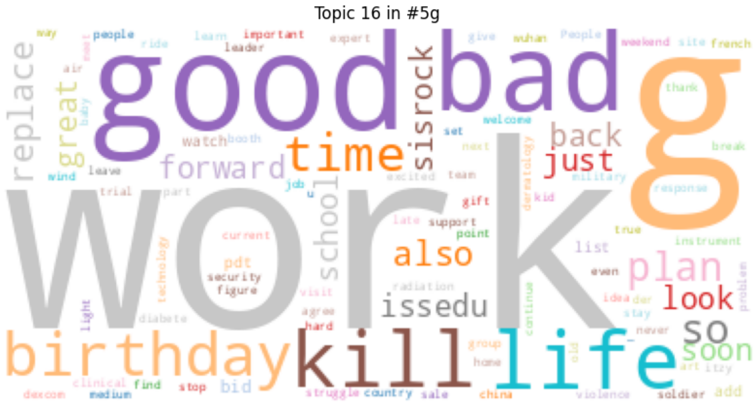
9.2 Appendix for deep dive into NGI-related hashtags

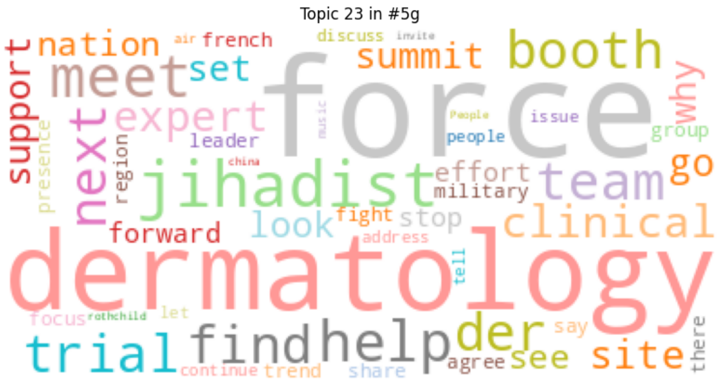
Overview of topic model keywords, label and word clouds for each of the ten hashtags

5g:

Topic #n	**Keywords**	**Label**
1	force, dermatology, jihadist, meet, help, team, find, next, booth, der	Health and military
2	take, real, terrorism, state, support, name, region, conflict, level, security	Conflict prevention
3	work, g, good, bad, life, kill, birthday, time, so, plan	Life related

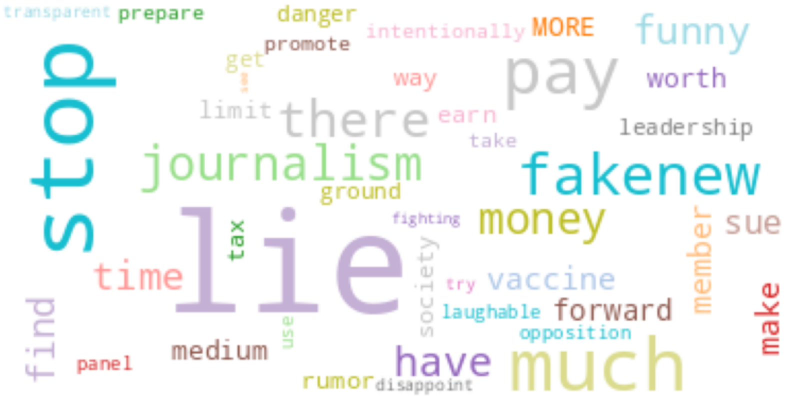




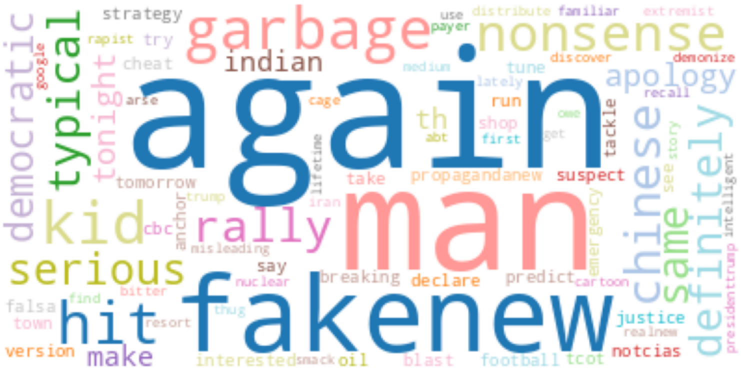


Fakenews:

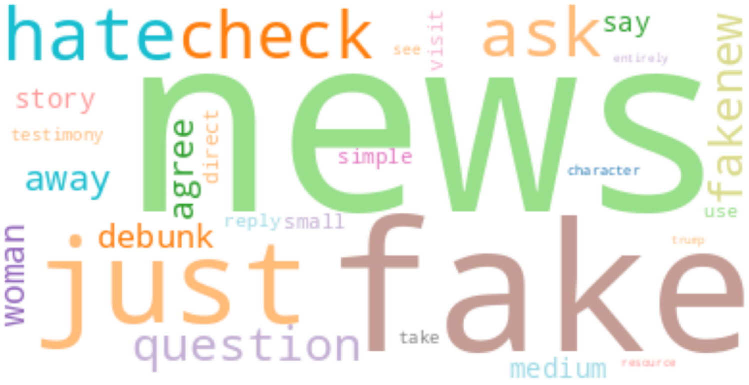
Topic #n	Keywords	Label
1	lie, stop, much, pay, fakenew, there, journalism, money, have, funny	Finances
2	again, man, fakenew, garbage, kid, hit, nonsense, chinese, rally, typical	Behaviour
3	news, fake, just, hate, ask, check, question, fakenew, away, woman	Research



Topic 22 in #fakenews



Topic 5 in #fakenews



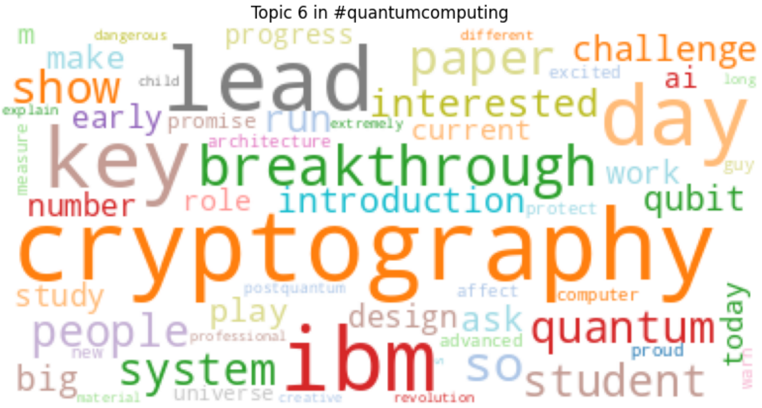
Topic 37 in #fakenews

Quantumcomputing:

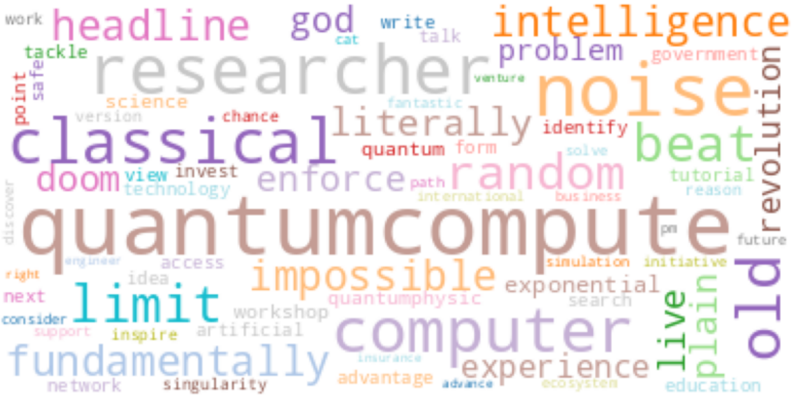
Topic #n	Keywords	Label

1	cryptography, ibm, lead, day, key, breakthrough, so, paper, show, people	Research
2	quantumcompute, noise, researcher, computer, classical, old, limit, beat, random, fundamentally	Technical
3	world, change, problem, solve, technology, how, potential, business, company, quantumcompute	Prospects, Innovation, and Business

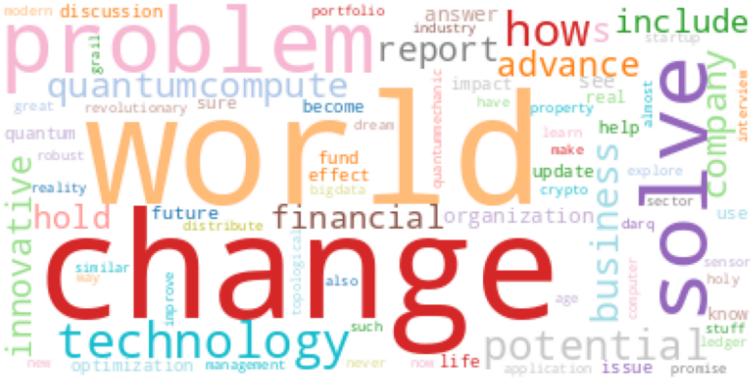
Topic 6 in #quantumcomputing



Topic 7 in #quantumcomputing



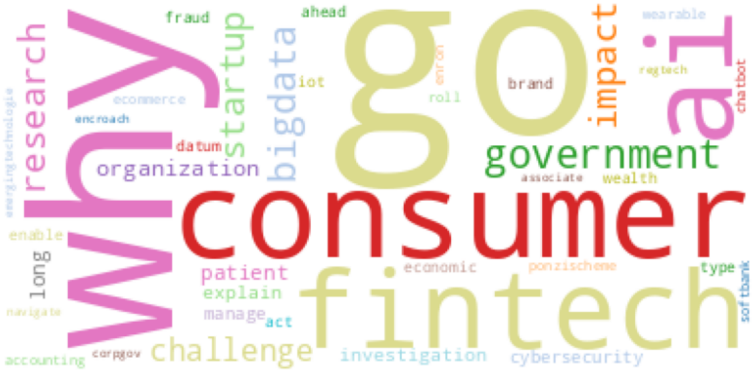
Topic 10 in #quantumcomputing

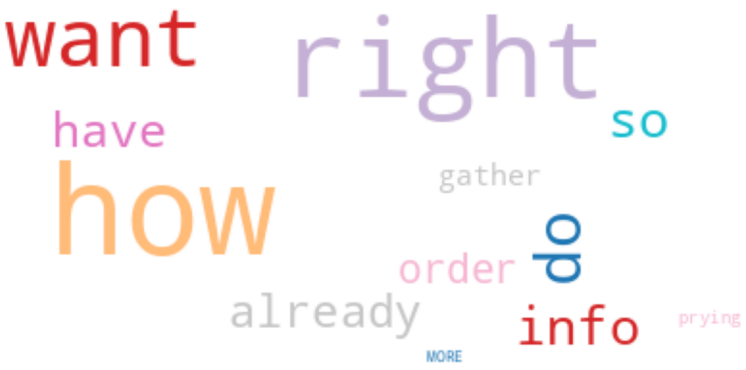


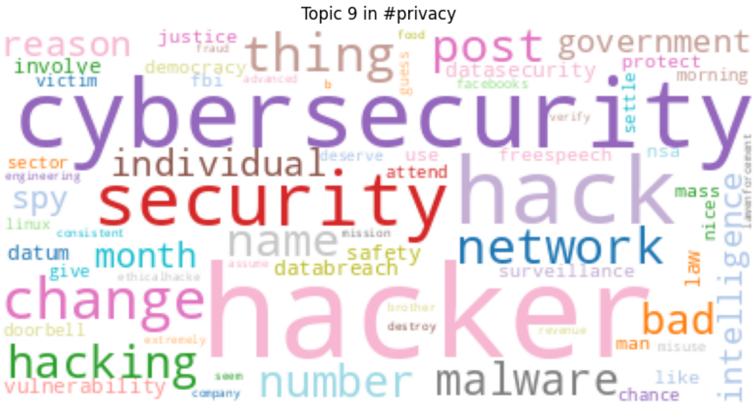
Privacy:

Topic #n	Keywords	Label

1	go, why, consumer, ai, fintech, government, bigdata, startup, research, impact	Business
2	how, right, want, do, info, so, have, already, order, gather	Action
3	hacker, cybersecurity, hack, security, change, thing, network, post, name, hacking	Cybersecurity


Topic 20 in #privacy

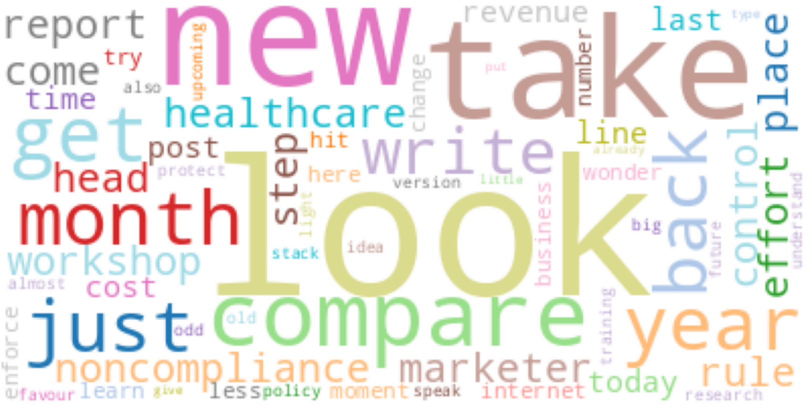

Topic 36 in #privacy



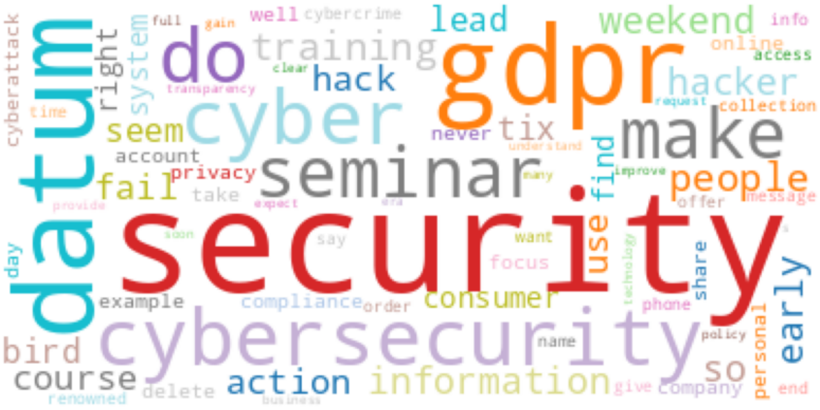
Gdpr:

Topic #n	Keywords	Label
1	look, new, take, compare, get, month, year, back, just, write	Time
2	security, datum, gdpr, cybersecurity, do, cyber, seminar, make, training, people	Security
3	gdpr, late, thank, new, cybersecurity, learn, compliance, privacy, blog, regulation	Law and manageme nt

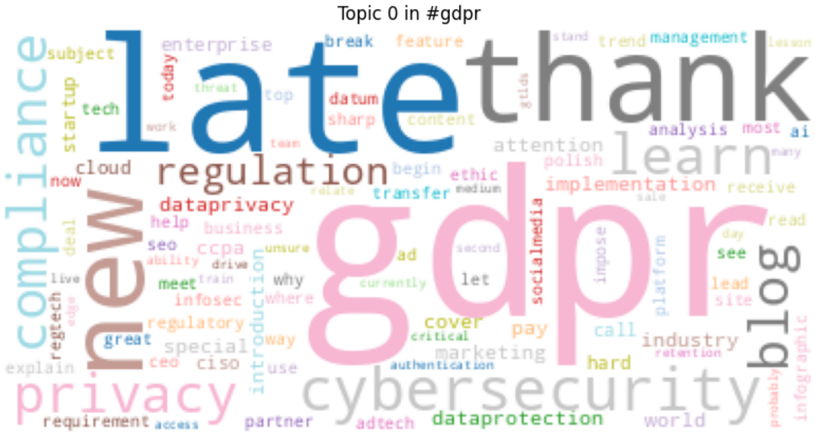
Topic 7 in #gdpr



Topic 3 in #gdpr



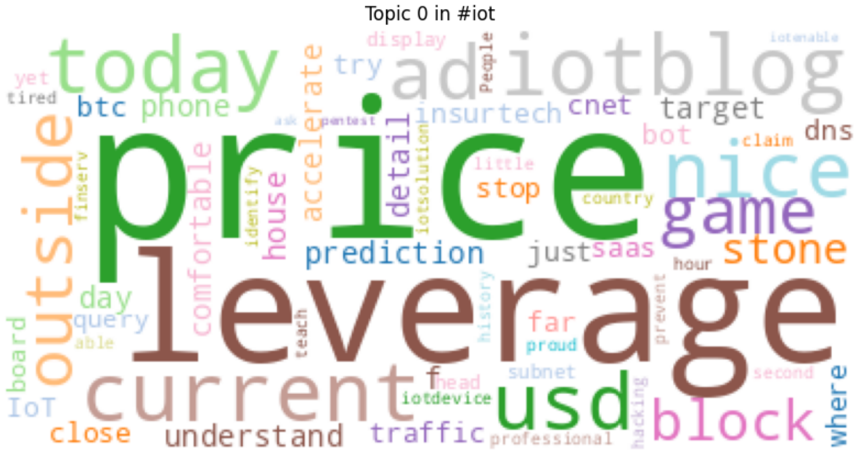
Topic 0 in #gdpr



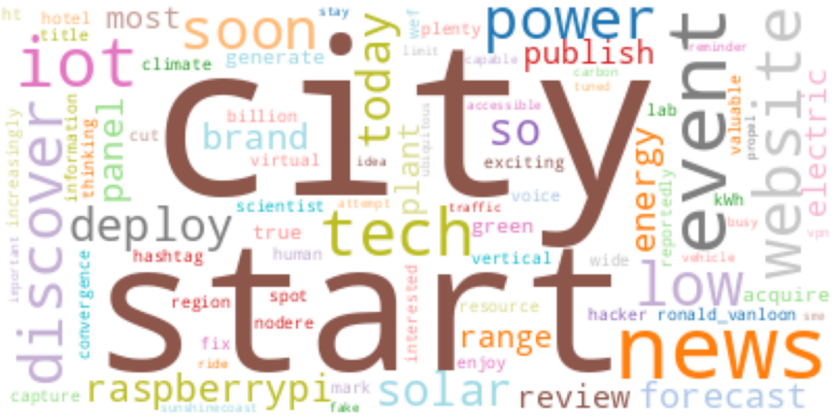
IoT:

Topic #n	Keywords	Label
1	city, start, news, event, iot, tech, website, discover, low, soon	Communication
2	tech, innovation, technology, startup, future, robot, robotic, selfdrivingcar, autonomous, discuss	Technology and Business
3	price, leverage, iotblog, current, nice, usd, today, ad, game, outside	Money

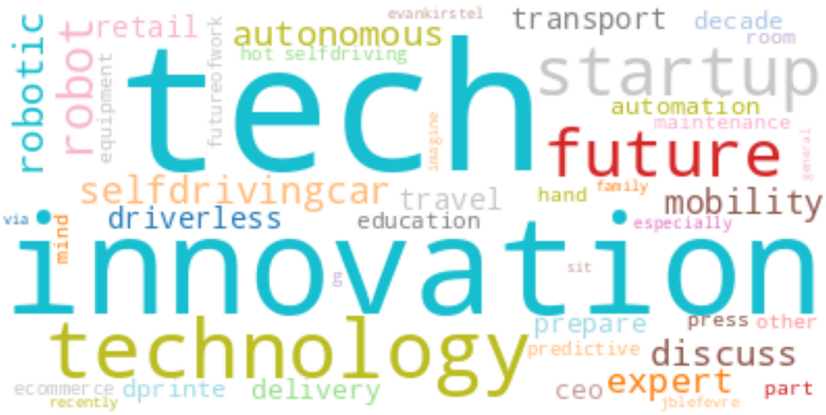
Topic 0 in #iot



Topic 7 in #iot



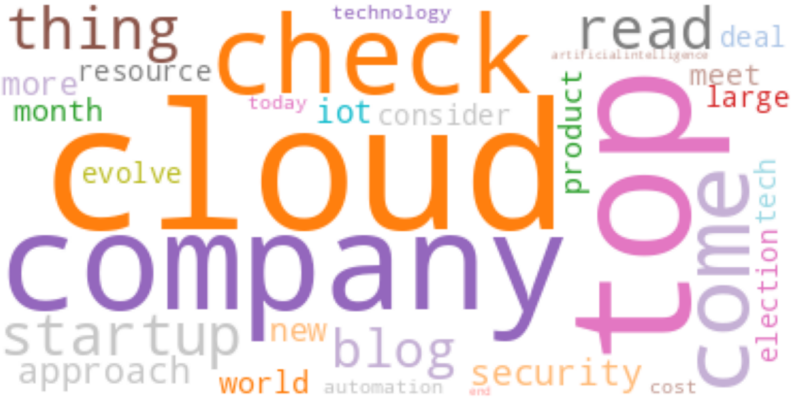
Topic 22 in #iot



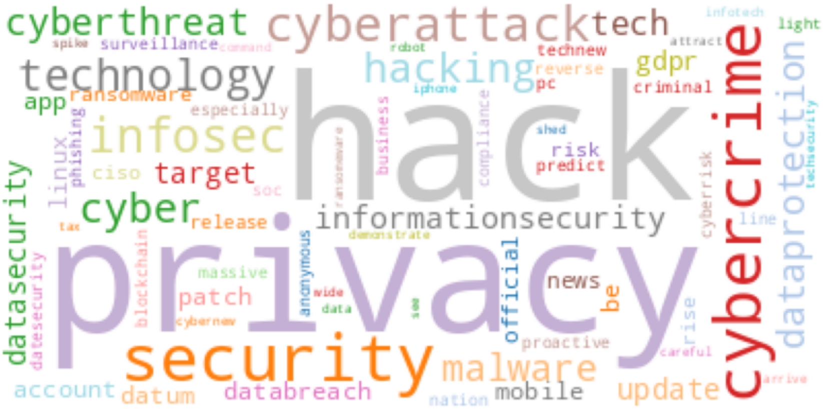
Cybersecurity:

Topic #n	Keywords	Label
1	cloud, top, company, check, come, startup, read, thing, blog, approach	Business
2	hack, privacy, security, cybercrime, cyberattack, infosec, technology, cyber, hacking, cyberthreat	Security
3	infosec, news, malware, use, new, phishe, website, today, attack, detect	Hacking

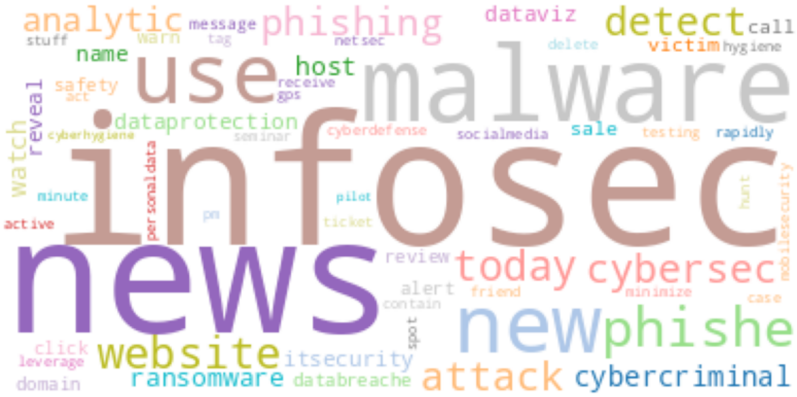
Topic 14 in #cybersecurity



Topic 6 in #cybersecurity



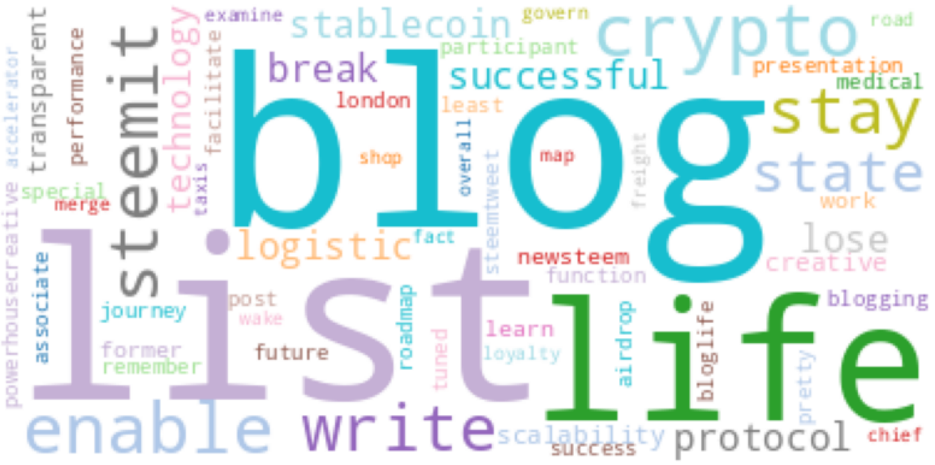
Topic 15 in #cybersecurity



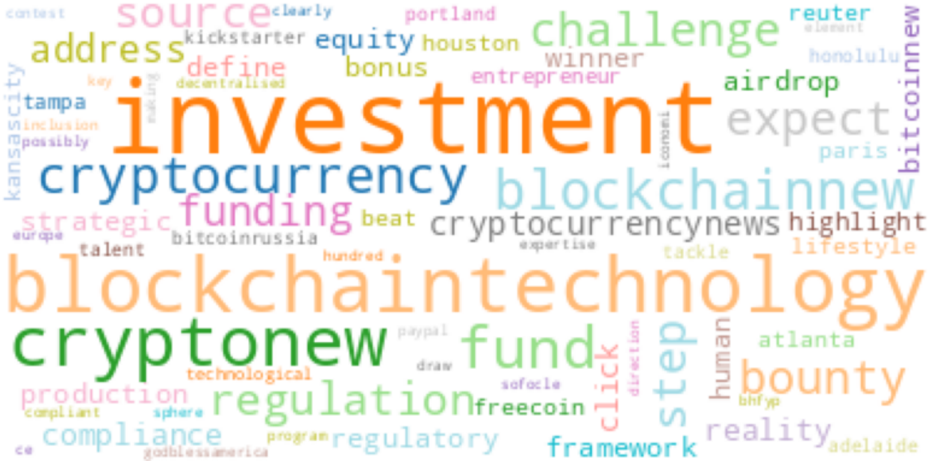
Blockchain:

Topic #n	Keywords	Label
1	investment, blockchaintechnology, cryptonew, fund, blockchainnew, cryptocurrency, bounty, challenge, step, expect	Trading and investments
2	blog, list, life, crypto, enable, steemit, write, stay, state, protocol	Documentation
3	business, digital, privacy, value, entrepreneur, fast, design, begin, art, insurance	Business

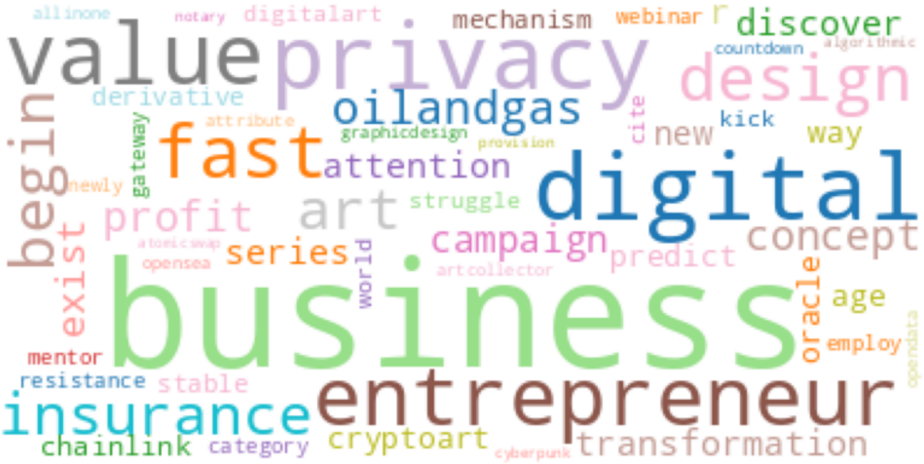
Topic 20 in #blockchain



Topic 37 in #blockchain



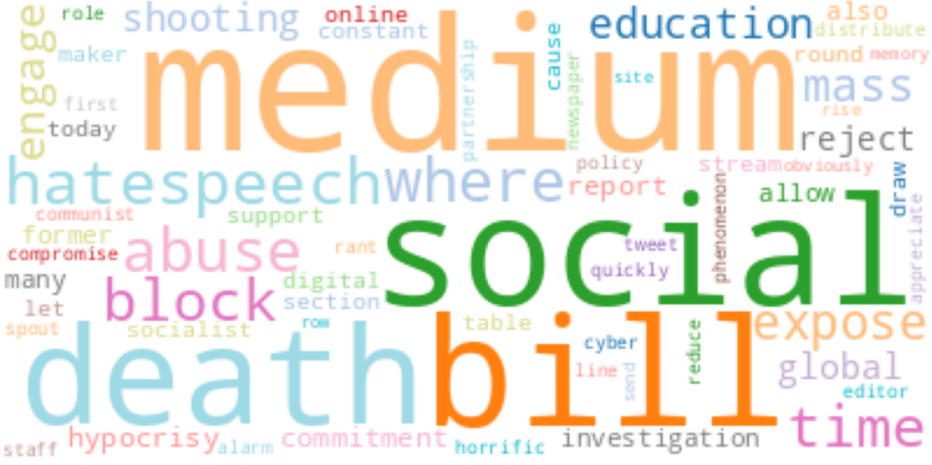
Topic 35 in #blockchain



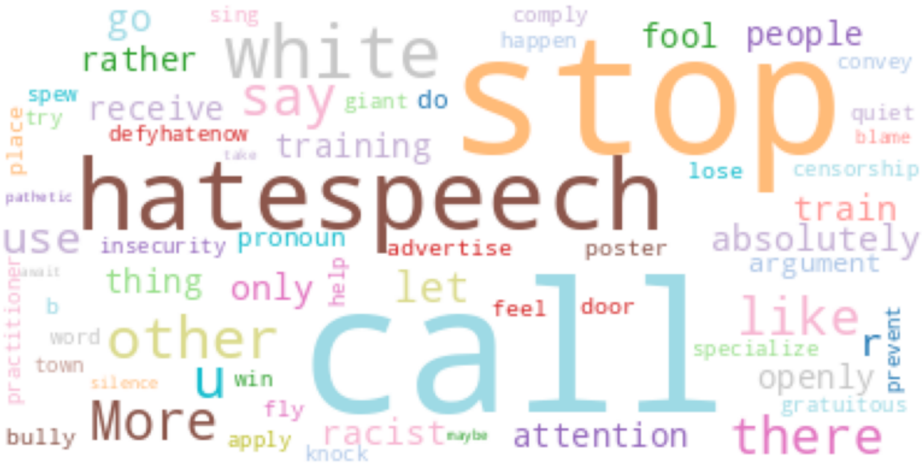
Hatespeech:

Topic #n	Keywords	Label
1	medium, social, bill, death, hatespeech, abuse, where, block, time, expose	Social media
2	call, stop, hatespeech, white, other, u, say, More, there, like	Take action
3	consider, arrest, find, hatespeech, term, flag, rise, offensive, write, true	Consequences

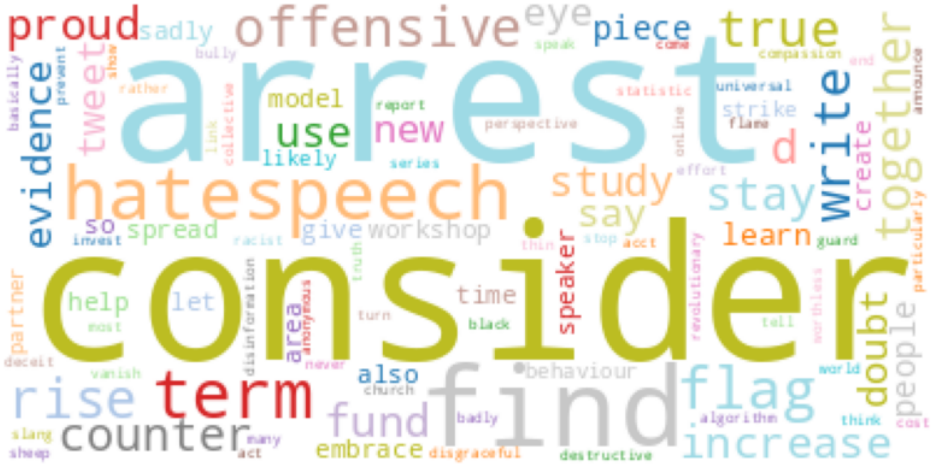
Topic 10 in #hatespeech



Topic 24 in #hatespeech



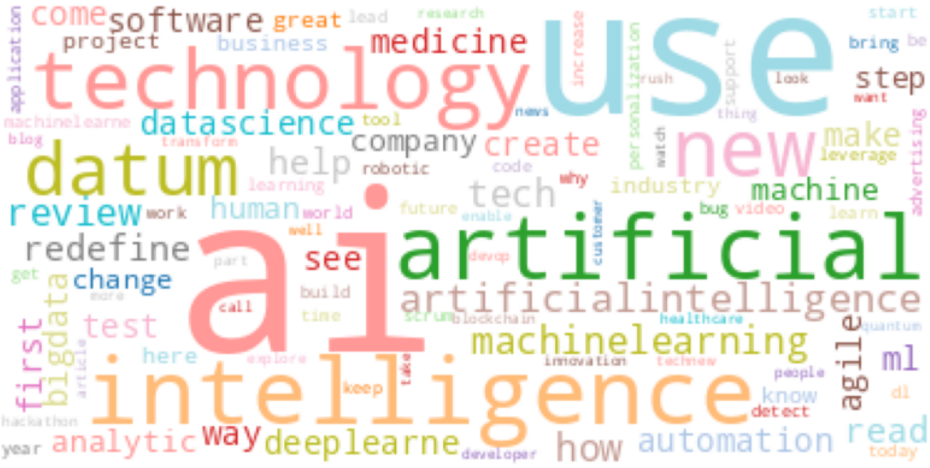
Topic 38 in #hatespeech



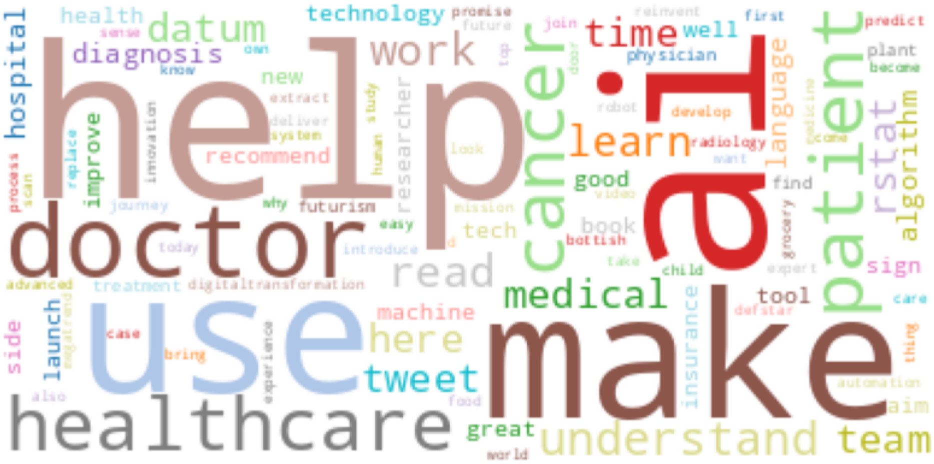
AI:

Topic #n	Keywords	Label
1	ai, use, intelligence, artificial, technology, new, datum, artificialintelligence, tech, how	Tech industry
2	ai, help, make, use, doctor, healthcare, patient, cancer, understand, work	Healthcare
3	use, ai, see, case, way, make, fight, great, human, get	Solutions

Topic 14 in #ai



Topic 39 in #ai



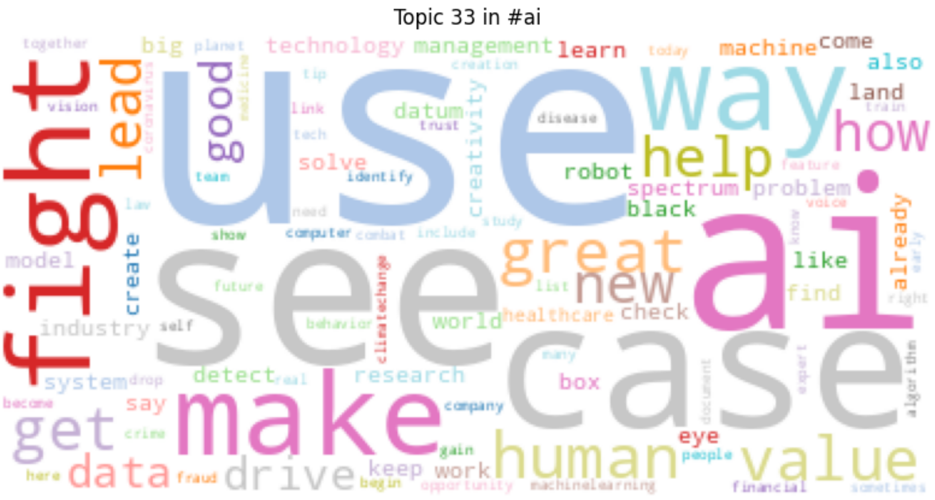
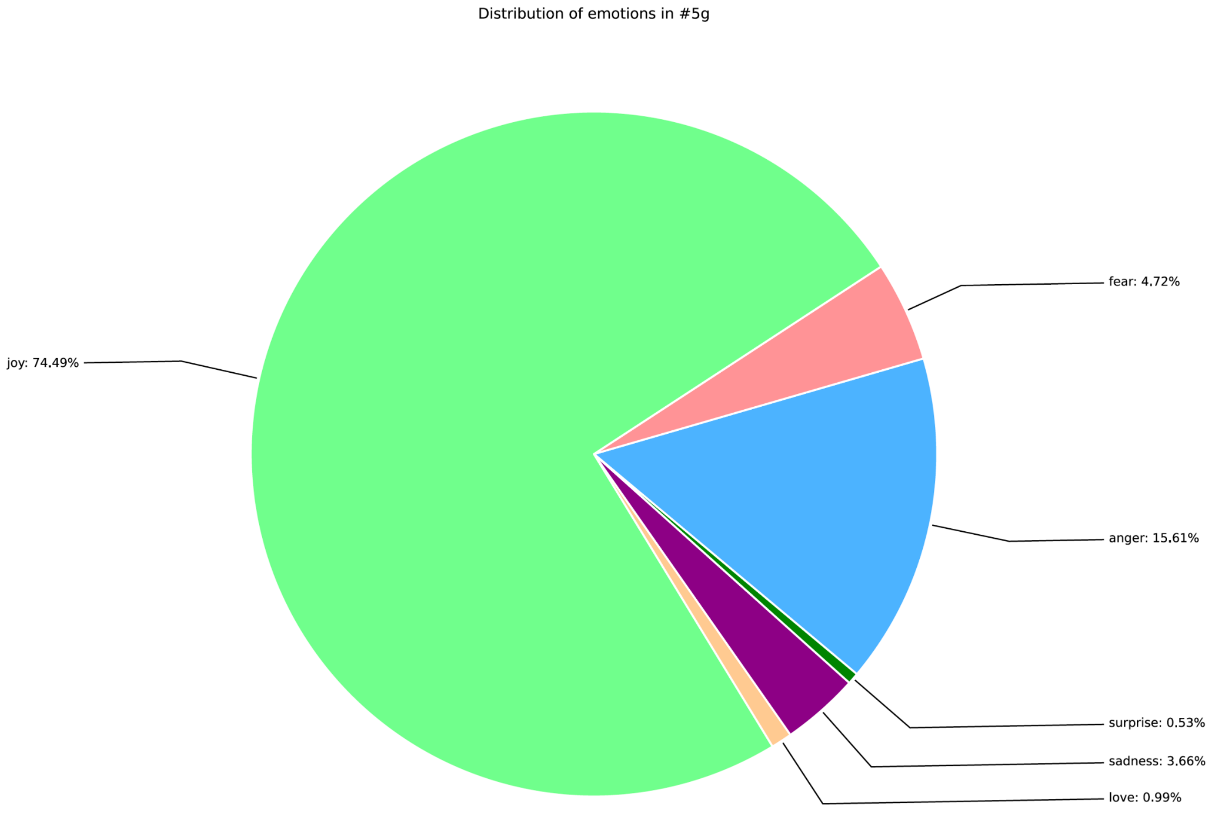
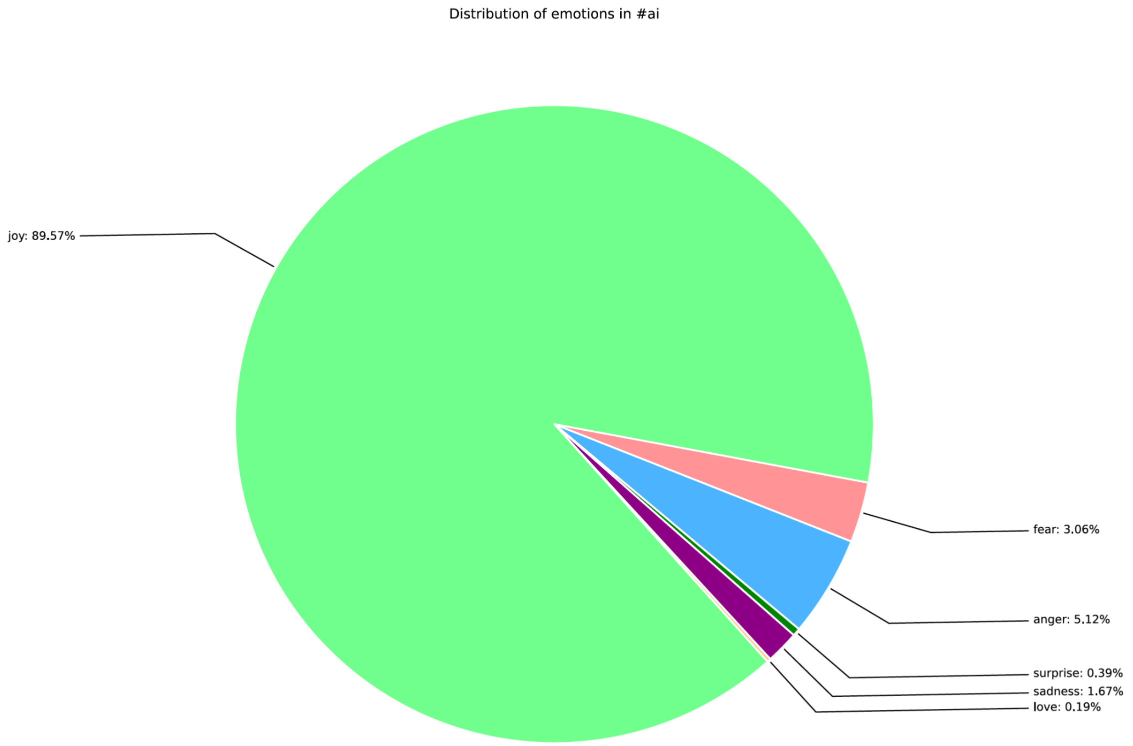


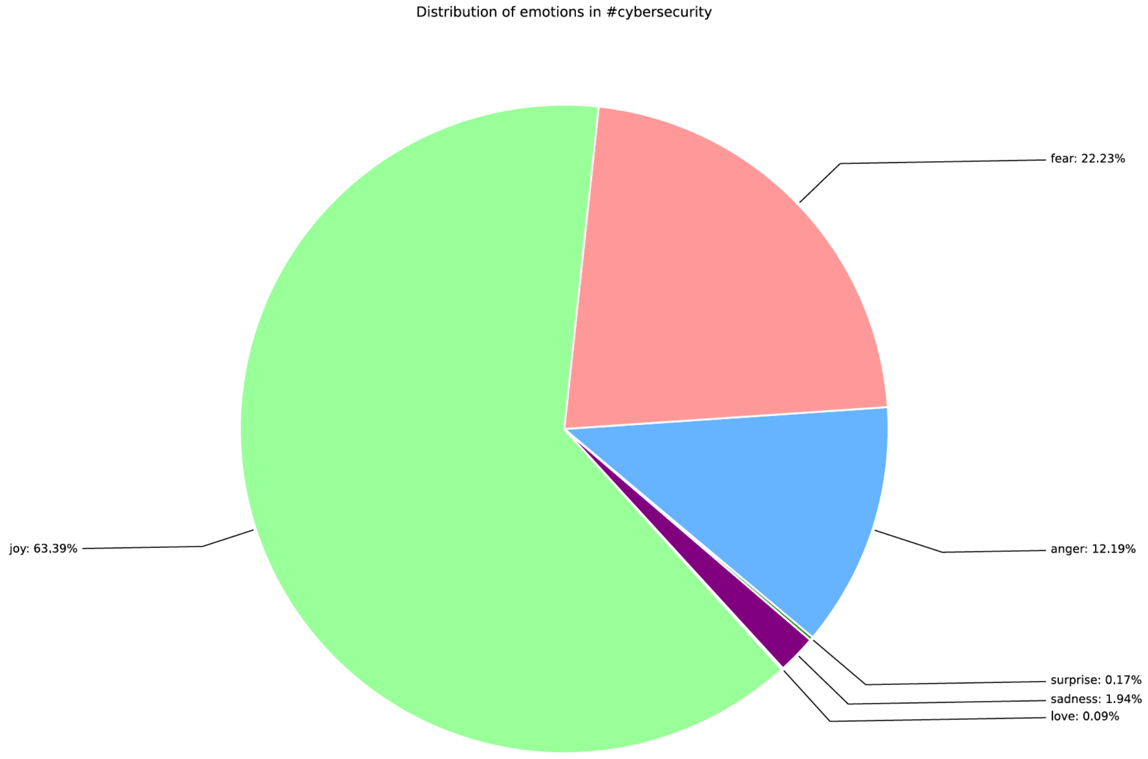
Table and pie charts for emotion analysis.

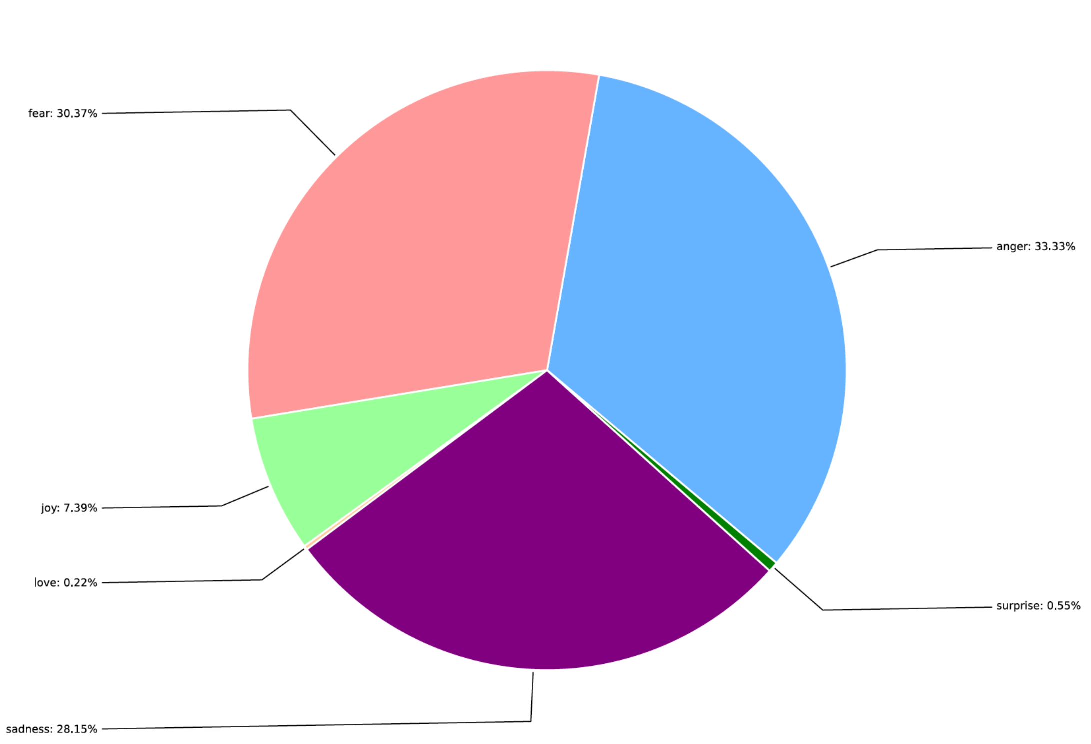
emotion_type	hashtag	adj_res	adj_sig
anger	5g	2.62	0.1
fear	5g	-8.09	**0.0**
joy	5g	5.28	**0.0**
love	5g	6.91	**0.0**
sadness	5g	-4.38	**0.0**
surprise	5g	1.17	2.91
anger	ai	-320.29	**0.0**
fear	ai	-375.94	**0.0**
joy	ai	619.35	**0.0**
love	ai	3.51	0.01
sadness	ai	-276.25	**0.0**
surprise	ai	9.64	**0.0**
anger	blockchain	-140.75	**0.0**
fear	blockchain	-242.04	**0.0**
joy	blockchain	377.4	**0.0**
love	blockchain	4.17	**0.0**
sadness	blockchain	-213.12	**0.0**
surprise	blockchain	-9.18	**0.0**
anger	cybersecurity	-27.99	**0.0**
fear	cybersecurity	289.76	**0.0**
joy	cybersecurity	-87.55	**0.0**
love	cybersecurity	-21.69	**0.0**
sadness	cybersecurity	-191.61	**0.0**

surprise	cybersecurity	-30.27	**0.0**
anger	fakenews	557.66	**0.0**
fear	fakenews	529.11	**0.0**
joy	fakenews	-1250.08	**0.0**
love	fakenews	8.83	**0.0**
sadness	fakenews	730.02	**0.0**
surprise	fakenews	34.22	**0.0**
anger	gdpr	33.68	**0.0**
fear	gdpr	-66.4	**0.0**
joy	gdpr	45.59	**0.0**
love	gdpr	8.47	**0.0**
sadness	gdpr	-46.16	**0.0**
surprise	gdpr	-2.59	0.11
anger	hatespeech	341.61	**0.0**
fear	hatespeech	-52.69	**0.0**
joy	hatespeech	-209.05	**0.0**
love	hatespeech	10.7	**0.0**
sadness	hatespeech	-7.65	**0.0**
surprise	hatespeech	-1.16	2.95
anger	iot	-175.04	**0.0**
fear	iot	-216.56	**0.0**
joy	iot	375.56	**0.0**
love	iot	-8.35	**0.0**
sadness	iot	-187.91	**0.0**
surprise	iot	-20.66	**0.0**
anger	privacy	67.02	**0.0**
fear	privacy	30.08	**0.0**
joy	privacy	-29.18	**0.0**
love	privacy	9.66	**0.0**
sadness	privacy	-78.09	**0.0**
surprise	privacy	-0.24	9.73
anger	quantumcomputing	-53.1	**0.0**
fear	quantumcomputing	-47.18	**0.0**
joy	quantumcomputing	87.21	**0.0**
love	quantumcomputing	-1.48	1.67
sadness	quantumcomputing	-33.93	**0.0**
surprise	*quantumcomputing*	*16.61*	***0.0***

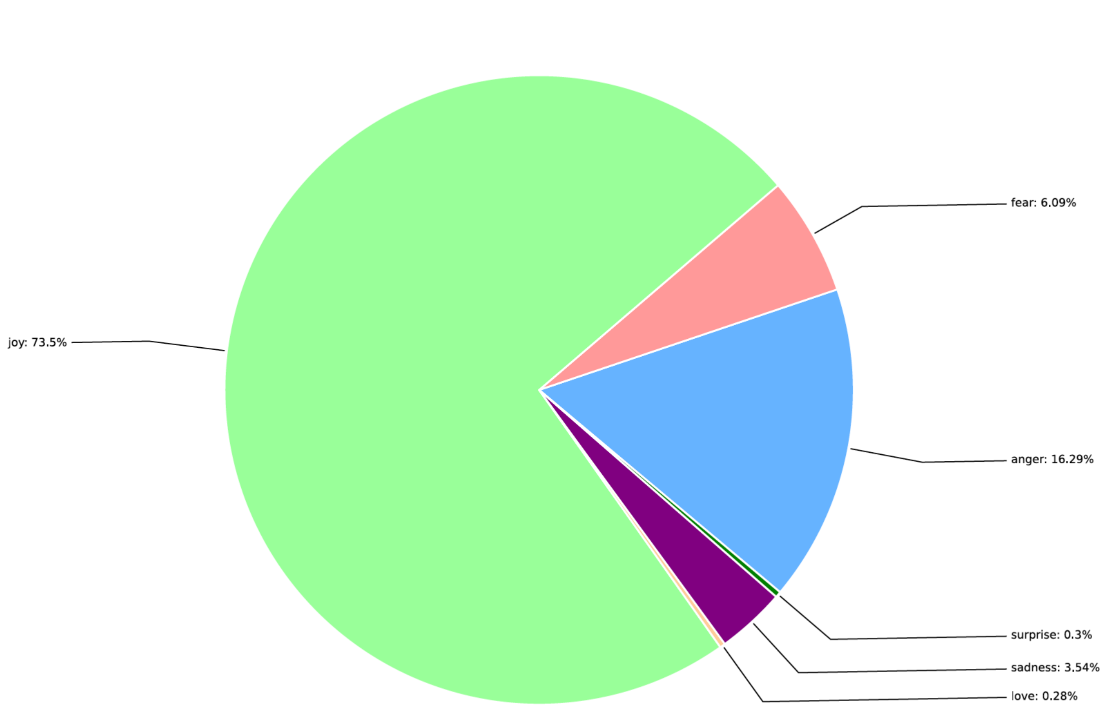

Distribution of emotions in #5g
fear: 4.72%
joy: 74.49%
anger: 15.61%
surprise: 0.53%
sadness: 3.66%
love: 0.99%


Distribution of emotions in #ai
joy: 89.57%
fear: 3.06%
anger: 5.12%
surprise: 0.39%
sadness: 1.67%
love: 0.19%


Distribution of emotions in #cybersecurity
fear: 22.23%
joy: 63.39%
anger: 12.19%
surprise: 0.17%
sadness: 1.94%
love: 0.09%

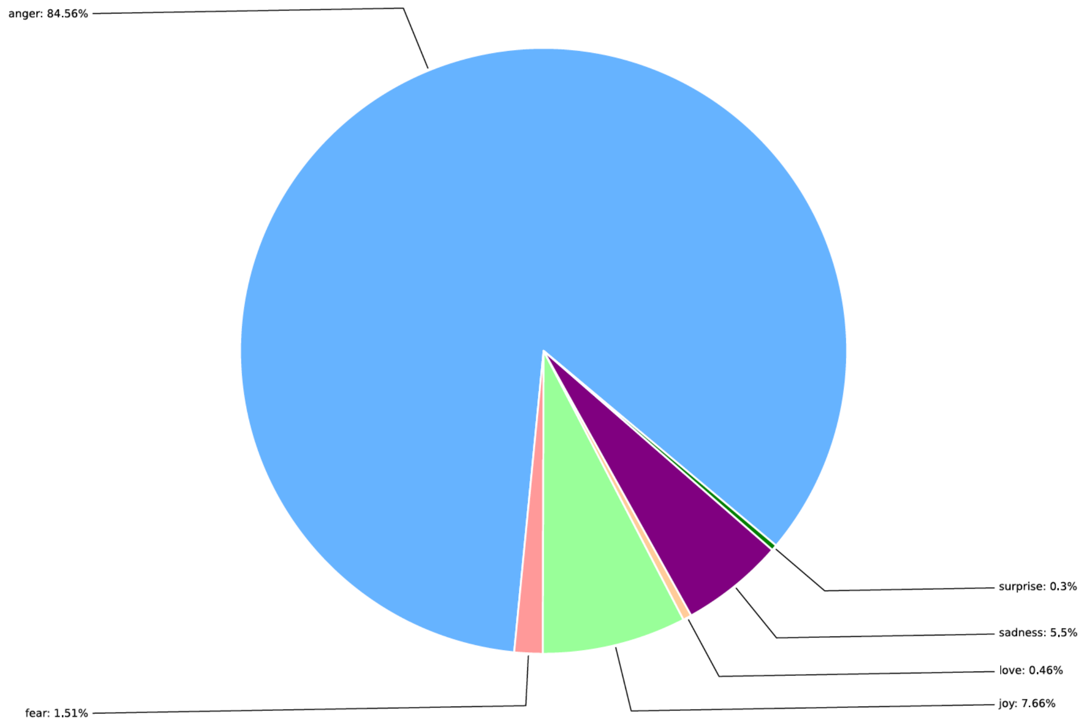


Distribution of emotions in #fakenews

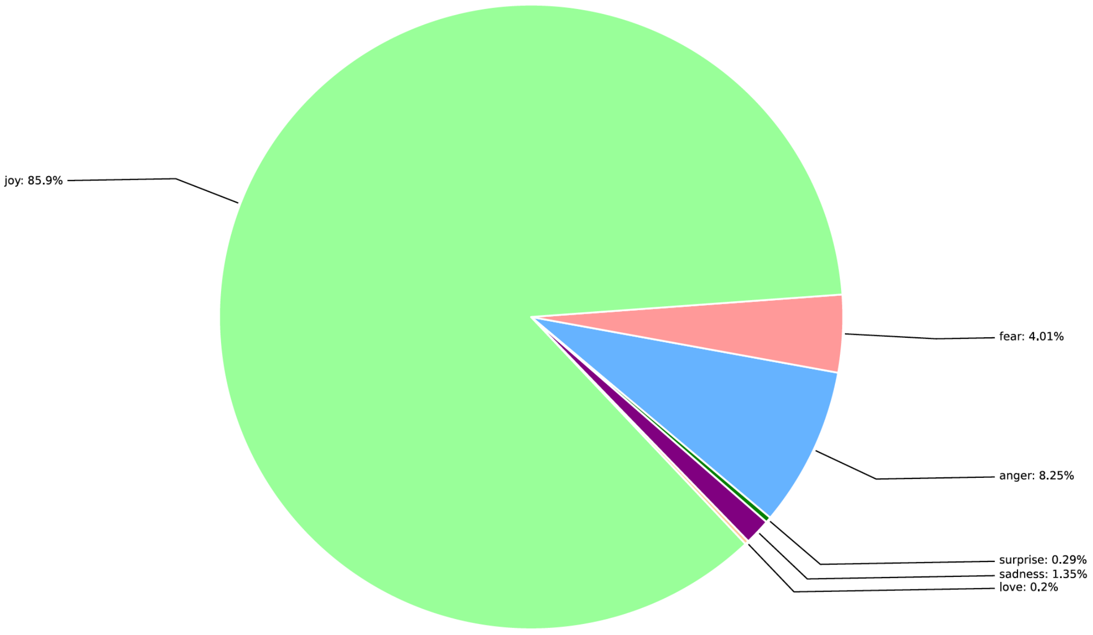


Distribution of emotions in #gdpr

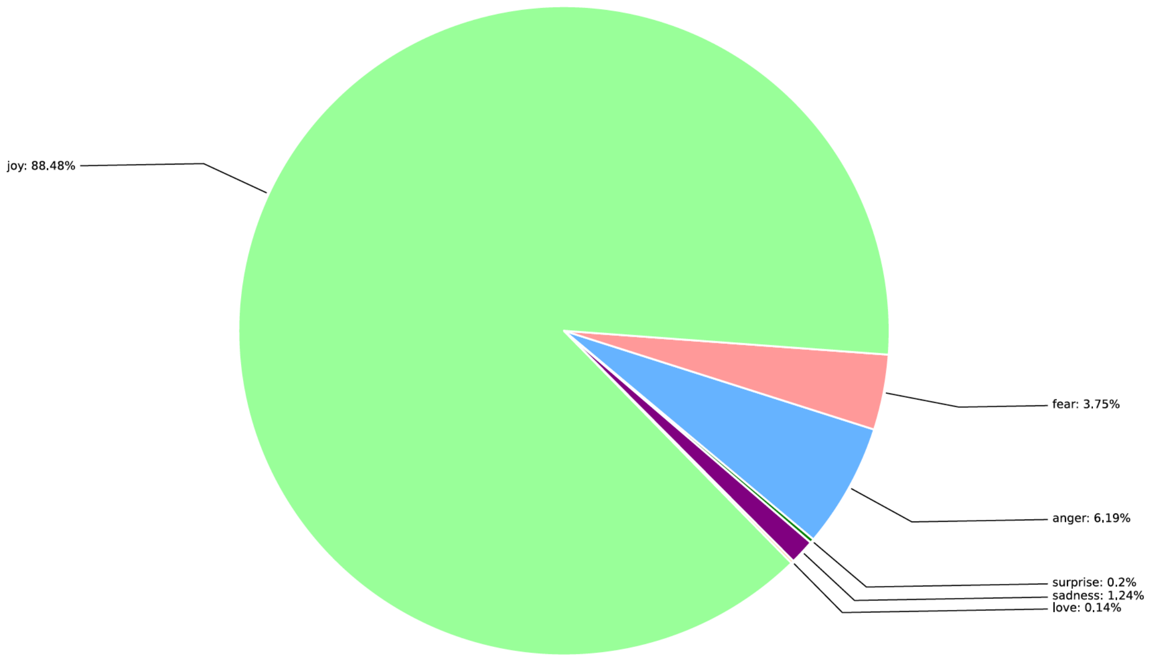
Distribution of emotions in #hatespeech



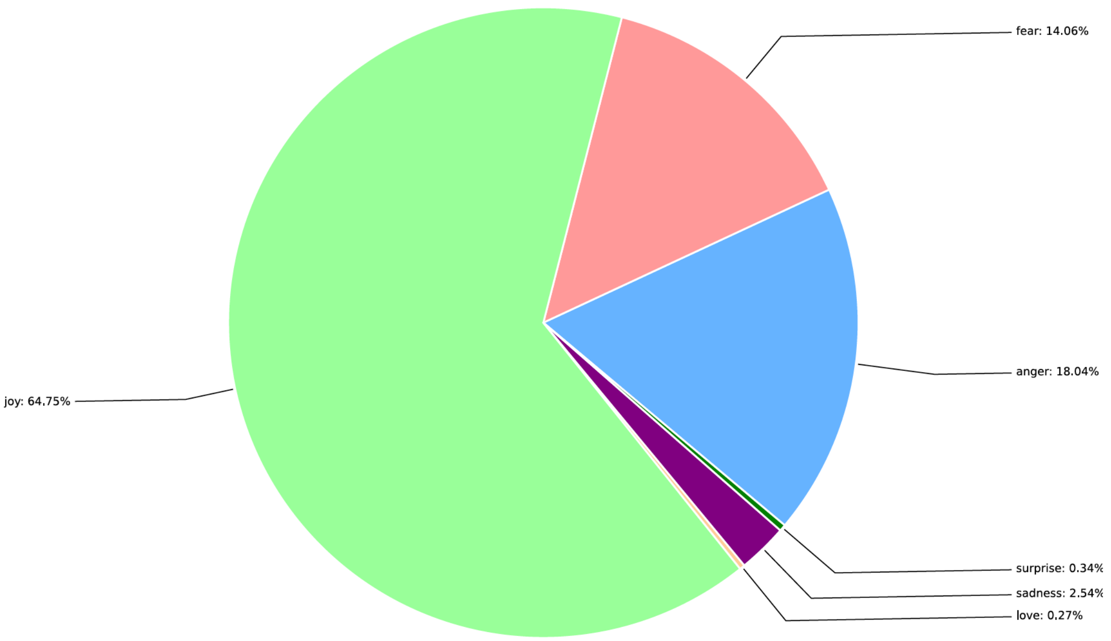
Distribution of emotions in #blockchain



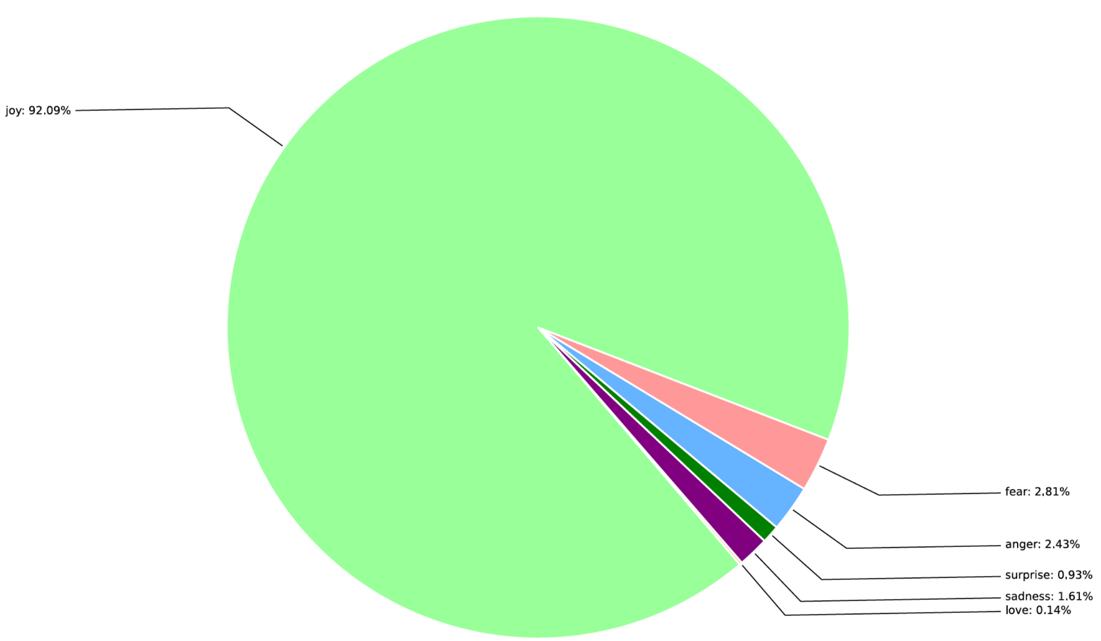
Distribution of emotions in #iot



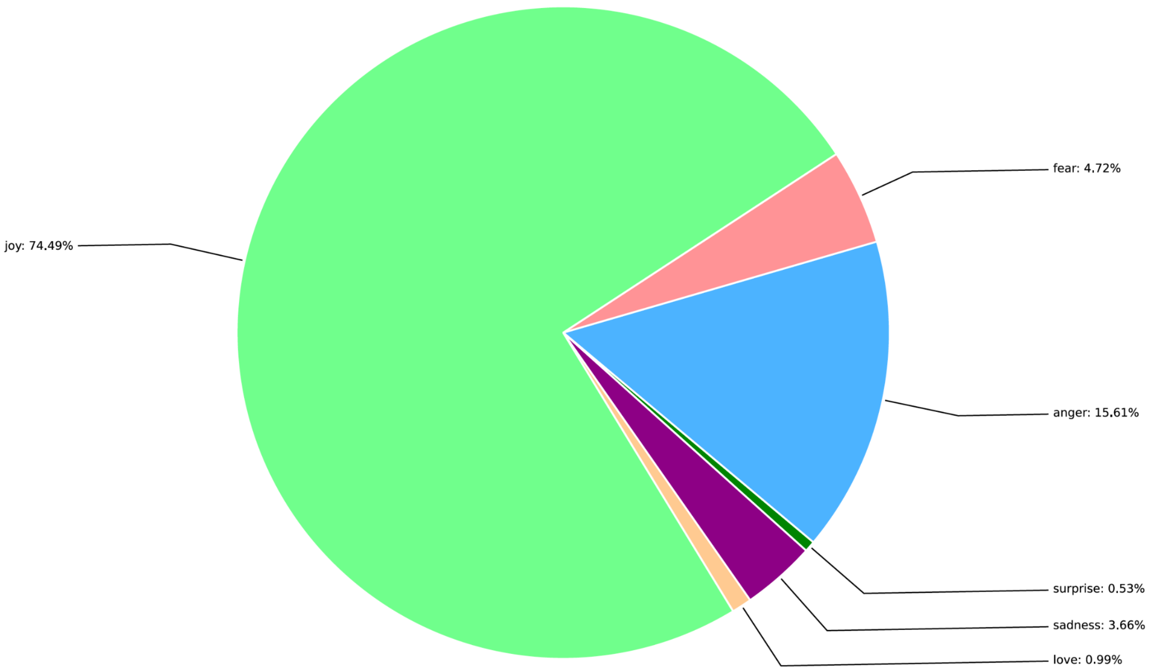
Distribution of emotions in #privacy



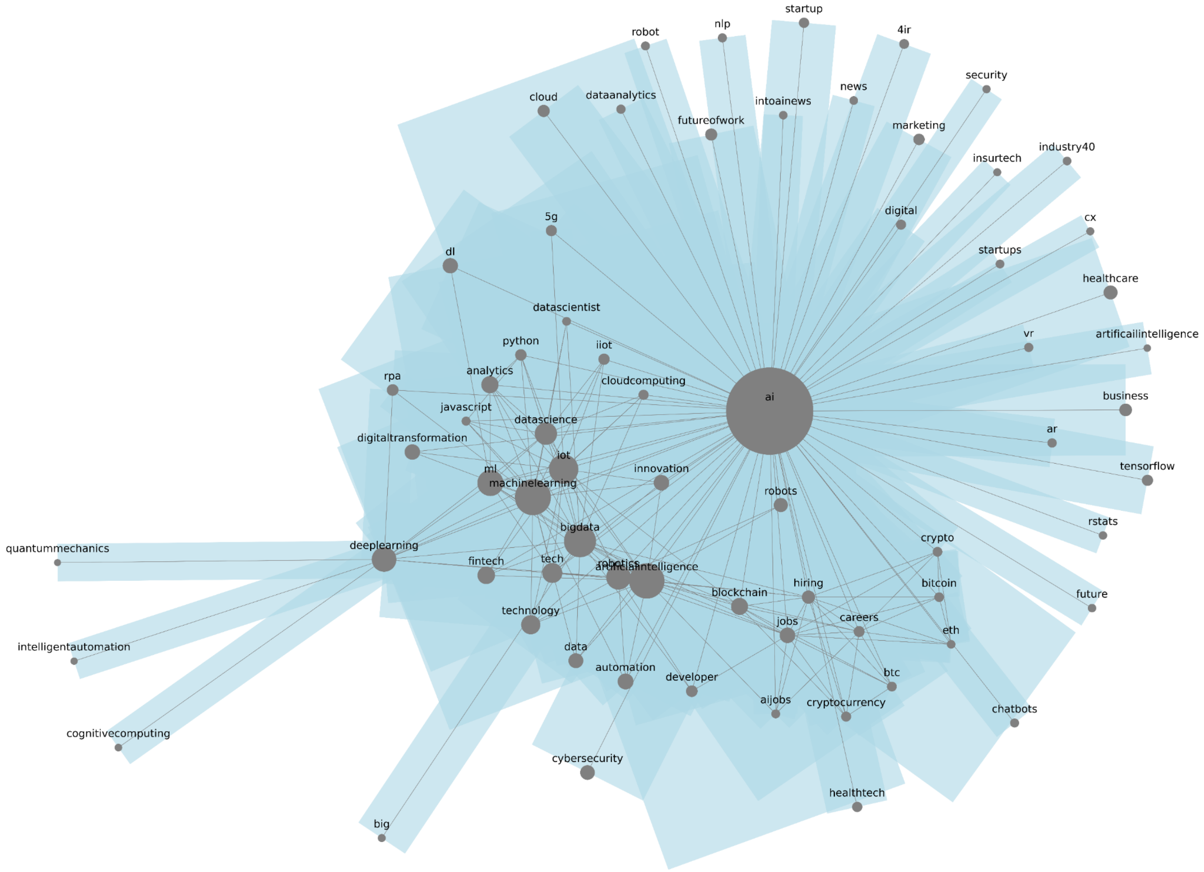
Distribution of emotions in #quantumcomputing

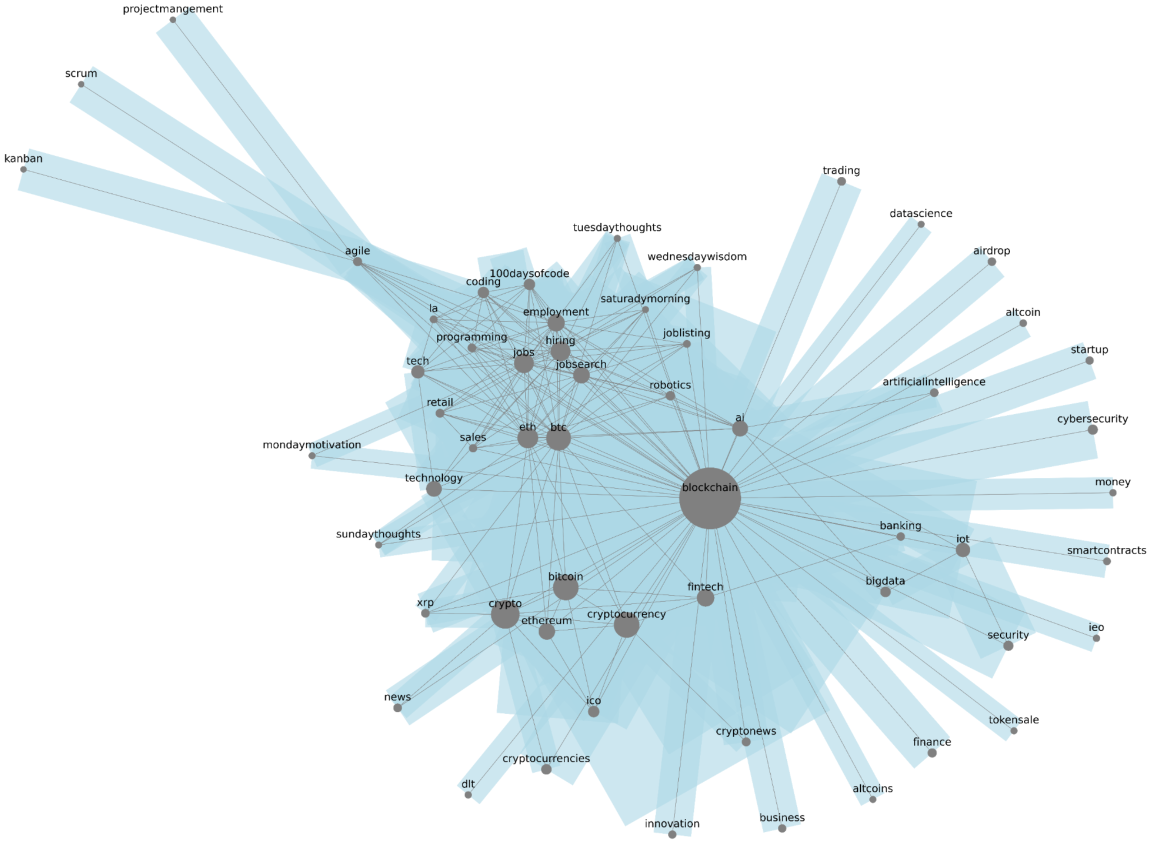


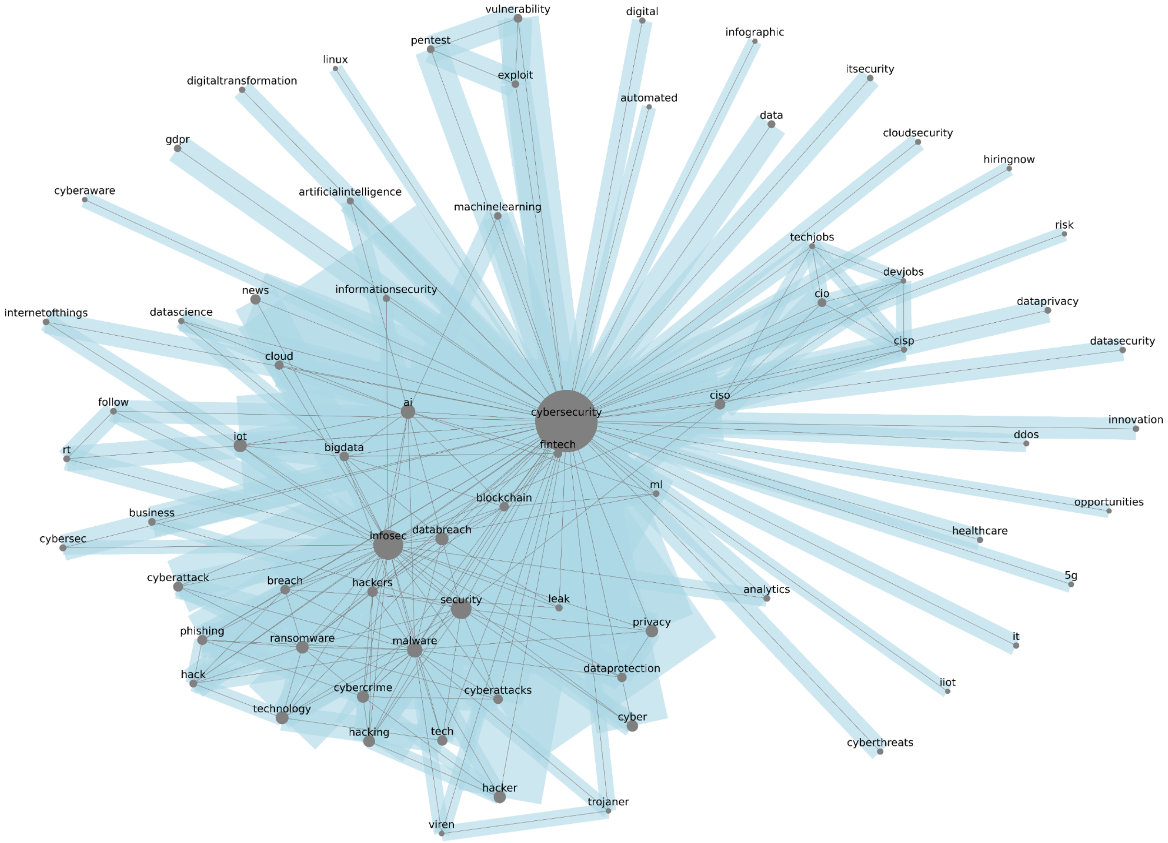
Distribution of emotions in #5g

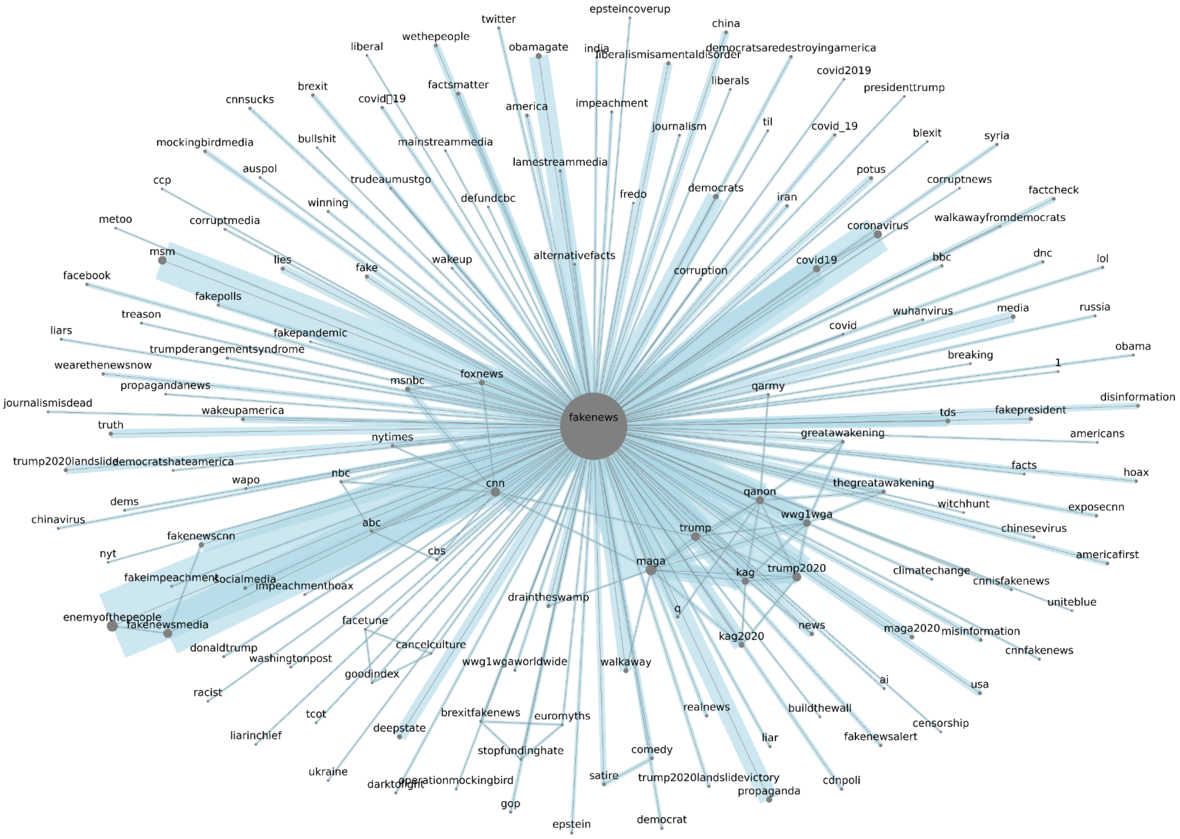


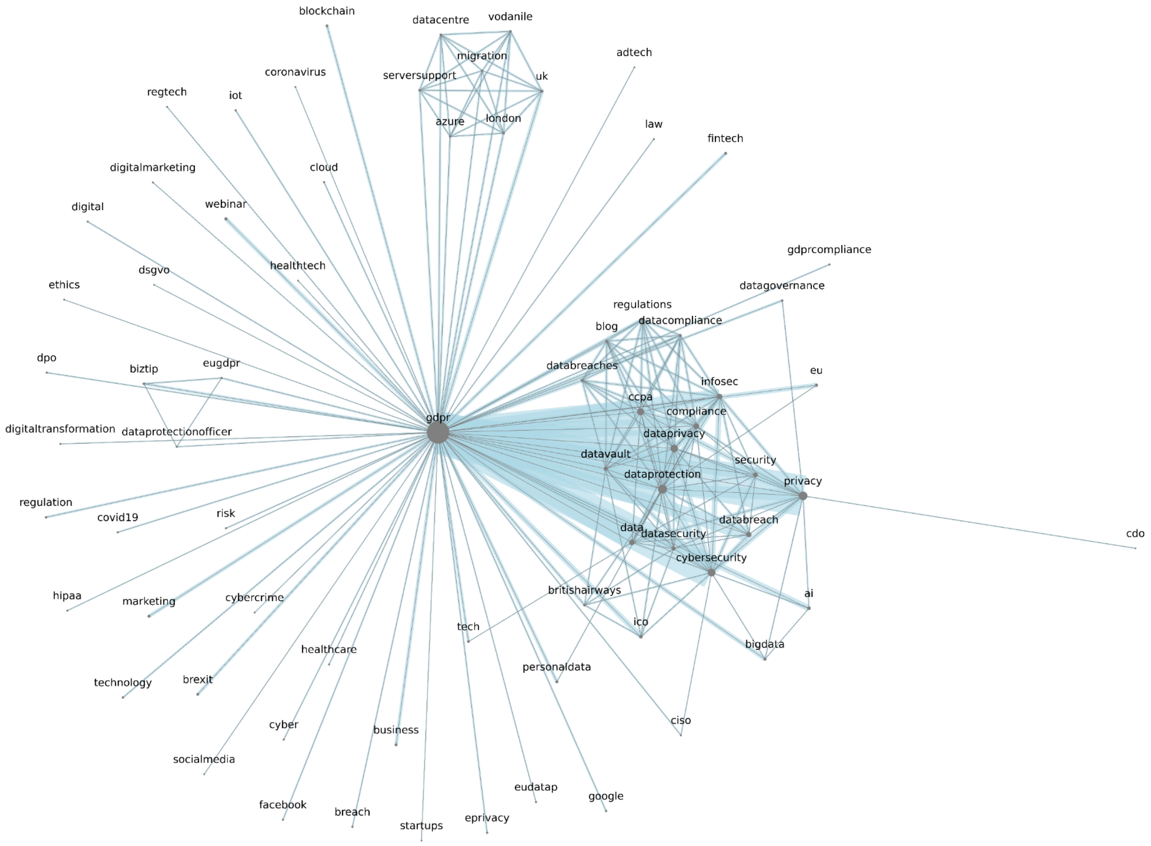
Network analysis

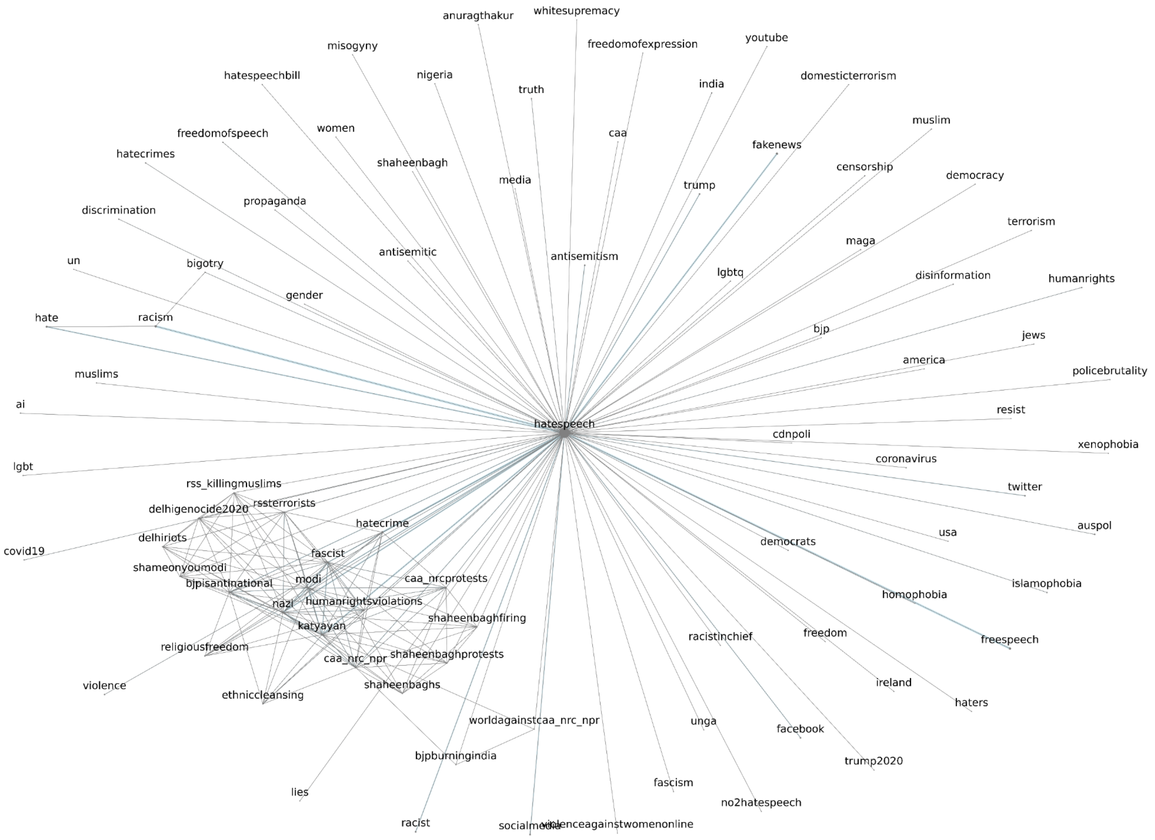

startup
nlp
robot
4ir
news
security
cloud
dataanalytics
intoainews
futureofwork
marketing
industry40
insurtech
digital
cx
5g
startups
dl
healthcare
datascientist
vr
artificailintelligence
python
iiot
analytics
rpa
cloudcomputing
ai
business
javascript
datascience
digitaltransformation
ar
iot
tensorflow
ml
innovation
machinelearning
robots
rstats
bigdata
quantummechanics
deeplearning
crypto
fintech
tech
robotics
artificialintelligence
hiring
bitcoin
blockchain
future
technology
jobs
careers
eth
intelligentautomation
data
automation
developer
btc
aijobs
cryptocurrency
chatbots
cognitivecomputing
cybersecurity
healthtech
big

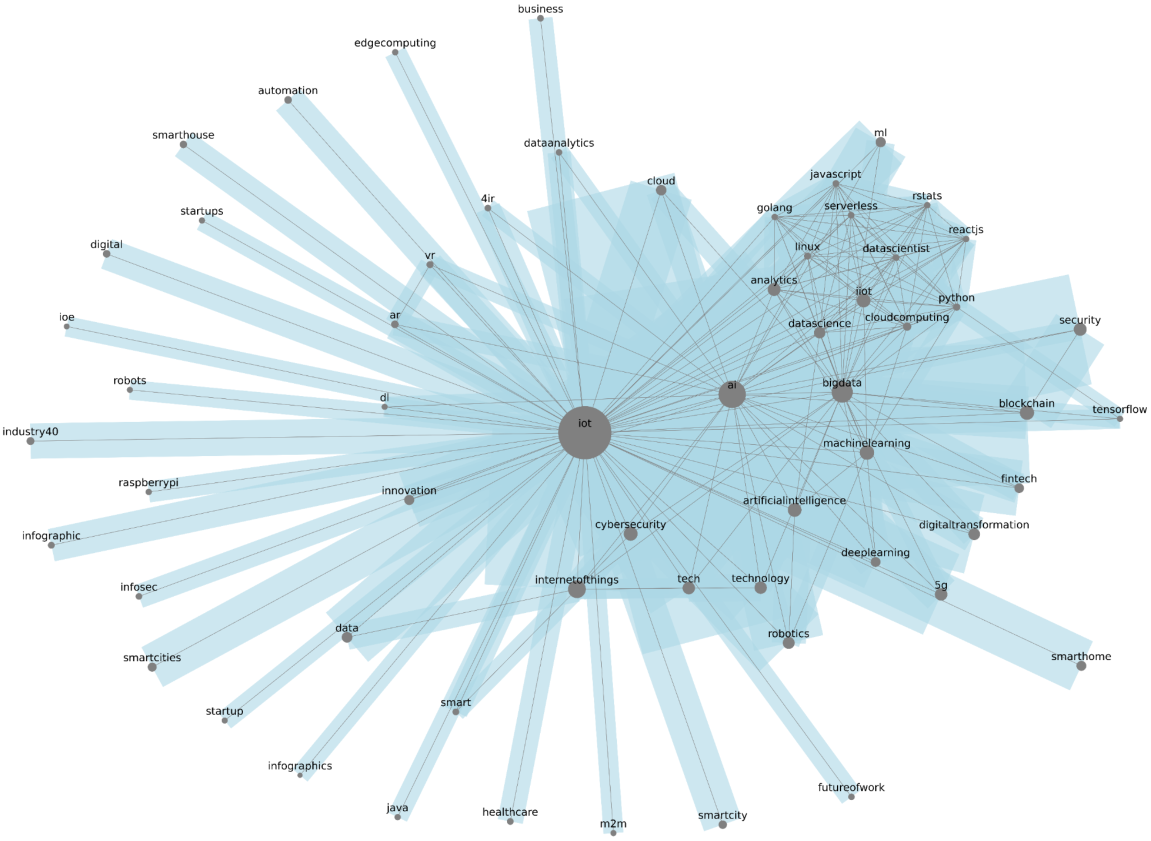

projectmangement
scrum
kanban
trading
datascience
tuesdaythoughts
airdrop
agile
wednesdaywisdom
coding
100daysofcode
altcoin
saturadymorning
la
employment
startup
joblisting
programming
hiring
artificialintelligence
tech
jobs
jobsearch
robotics
retail
ai
cybersecurity
eth
btc
sales
mondaymotivation
blockchain
technology
money
banking
iot
sundaythoughts
smartcontracts
bitcoin
bigdata
fintech
xrp
crypto
cryptocurrency
ieo
ethereum
security
news
ico
tokensale
cryptonews
finance
cryptocurrencies
dlt
altcoins
business
innovation

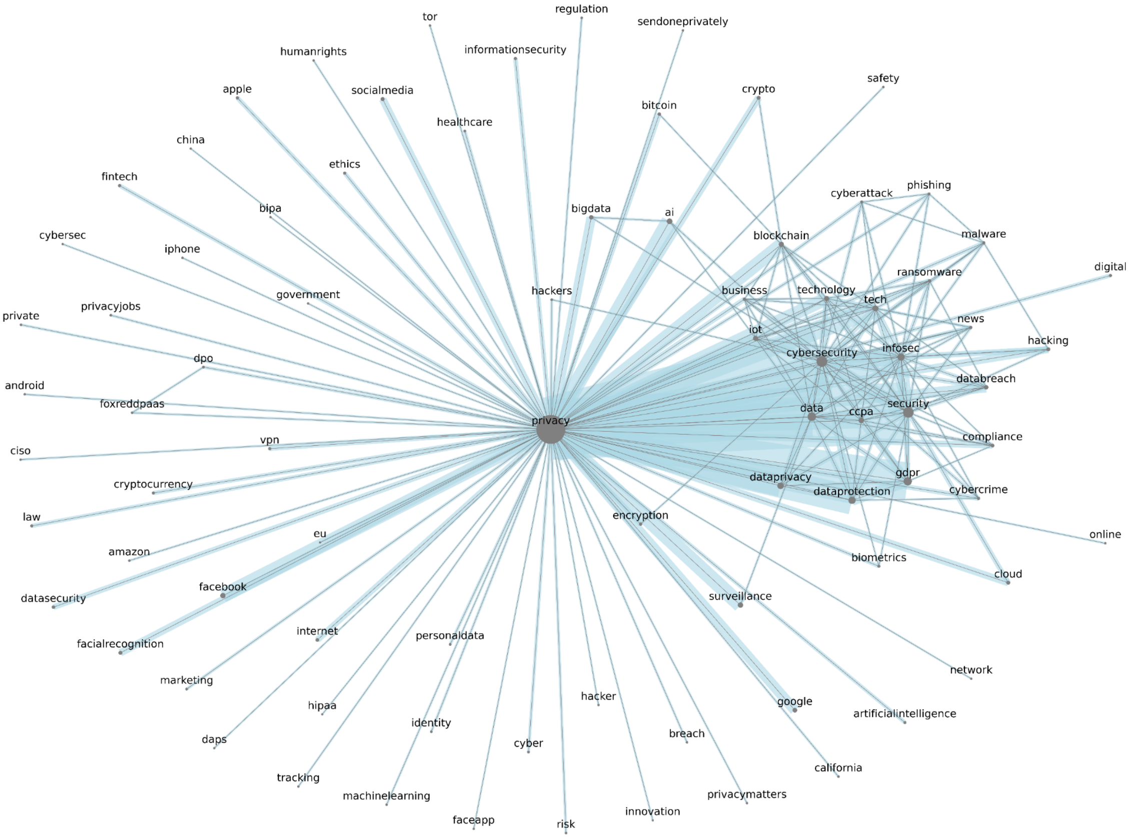

vulnerability
digital
pentest
infographic
linux
exploit
itsecurity
digitaltransformation
automated
data
cloudsecurity
gdpr
hiringnow
cyberaware
artificialintelligence
machinelearning
risk
techjobs
devjobs
cio
news
informationsecurity
dataprivacy
internetofthings
datascience
cisp
datasecurity
cloud
follow
ai
ciso
cybersecurity
innovation
iot
ddos
fintech
rt
bigdata
ml
opportunities
blockchain
business
healthcare
databreach
cybersec
infosec
5g
cyberattack
breach
hackers
analytics
security
leak
privacy
phishing
ransomware
malware
it
dataprotection
hack
iiot
cybercrime
cyberattacks
technology
cyber
cyberthreats
hacking
tech
hacker
trojaner
viren

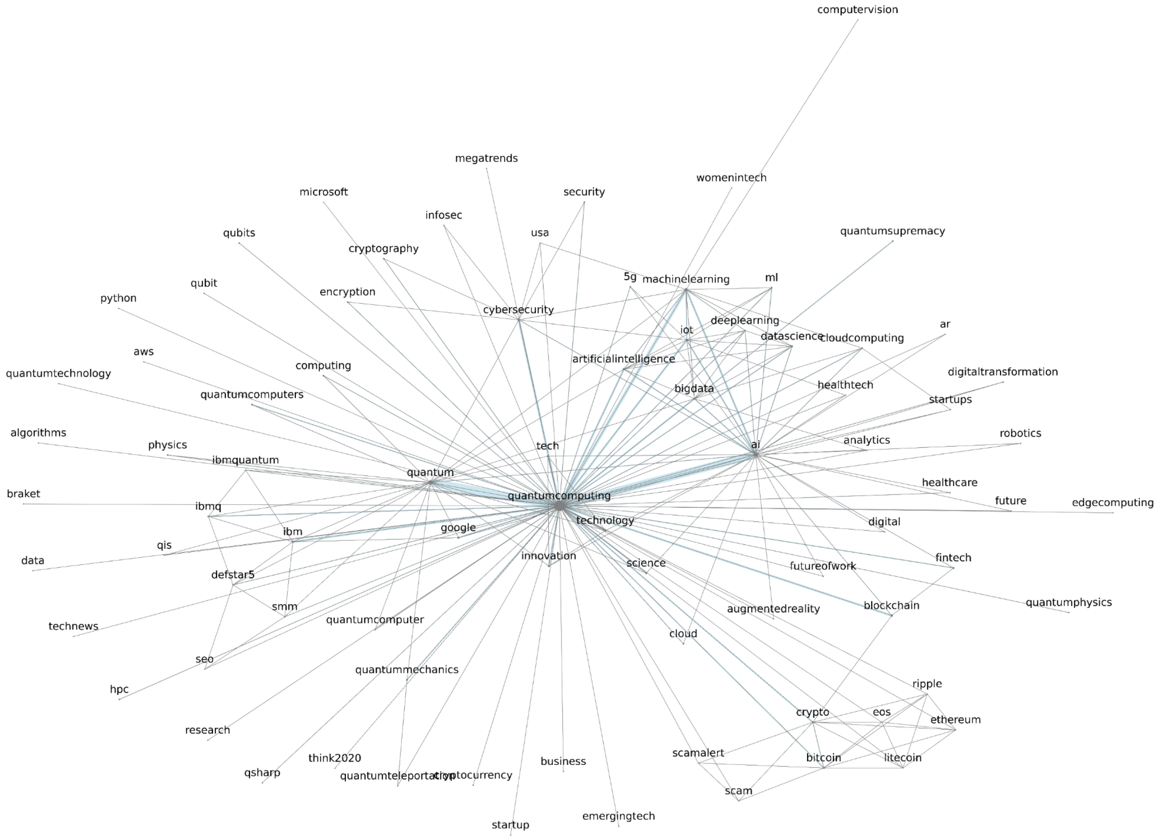

epsteincoverup
twitter
china
wethepeople
liberal
obamagate
india
democratsaredestroyingamerica
liberalismisamentaldisorder
covid2019
liberals
presidenttrump
brexit
factsmatter
cnnsucks
covid□19
america
impeachment
til
covid_19
journalism
blexit
syria
mockingbirdmedia
bullshit
mainstreammedia
lamestreammedia
auspol
potus
ccp
trudeaumustgo
corruptnews
democrats
factcheck
fredo
defundcbc
iran
winning
metoo
corruptmedia
walkawayfromdemocrats
coronavirus
alternativefacts
msm
lies
wakeup
bbc
dnc
lol
covid19
fake
corruption
facebook
wuhanvirus
fakepolls
media
russia
treason
covid
fakepandemic
liars
breaking
obama
trumpderangementsyndrome
wearethenewsnow
1
foxnews
msnbc
qarmy
propagandanews
disinformation
journalismisdead
wakeupamerica
fakepresident
tds
fakenews
truth
greatawakening
americans
nytimes
trump2020landslide
democratshateamerica
facts
hoax
nbc
wapo
cnn
thegreatawakening
qanon
dems
witchhunt
exposecnn
chinavirus
wwg1wga
abc
fakenewscnn
trump
chinesevirus
americafirst
nyt
cbs
maga
trump2020
climatechange
kag
fakeimpeachment
socialmedia
cnnisfakenews
impeachmenthoax
uniteblue
draintheswamp
q
enemyofthepeople
fakenewsmedia
facetune
news
maga2020
misinformation
kag2020
donaldtrump
cancelculture
cnnfakenews
washingtonpost
wwg1wgaworldwide
walkaway
ai
goodindex
usa
racist
realnews
buildthewall
brexitfakenews
tcot
euromyths
censorship
deepstate
liarinchief
liar
fakenewsalert
stopfundinghate
comedy
ukraine
satire
trump2020landslidevictory
cdnpoli
darktolight
operationmockingbird
propaganda
gop
democrat
epstein

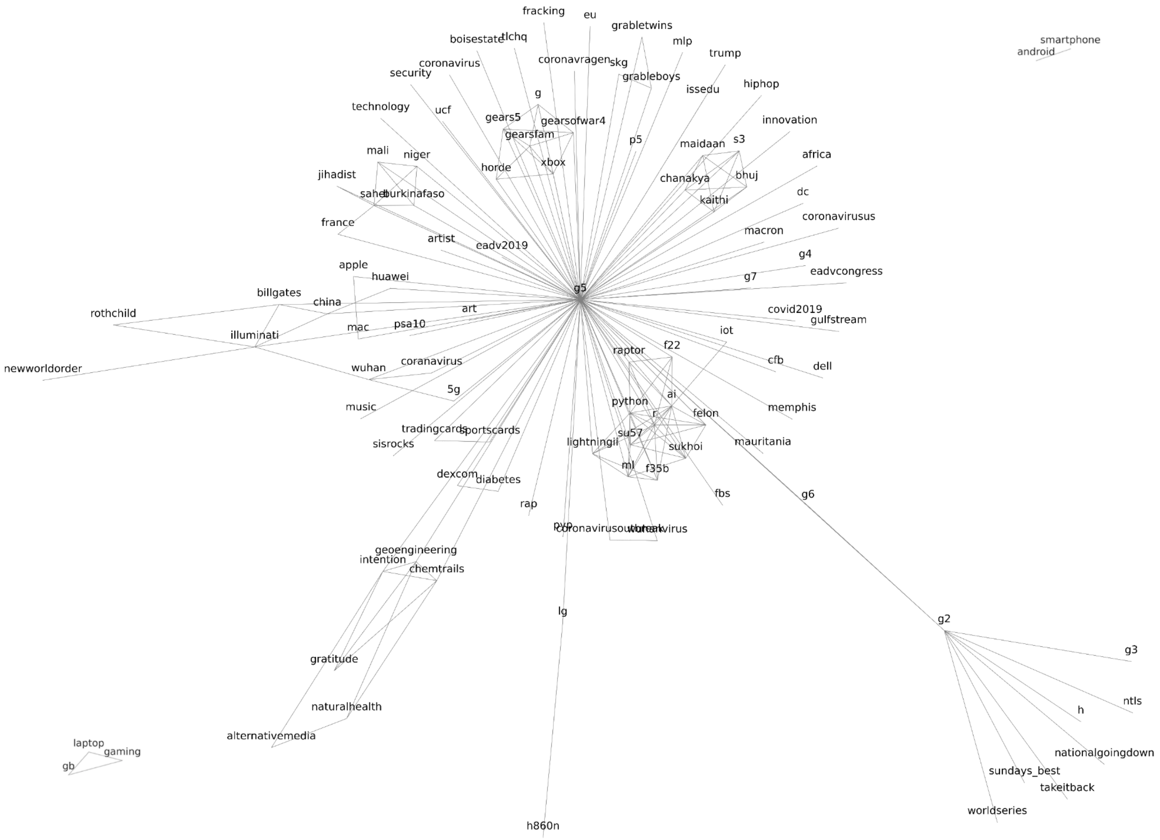

blockchain
datacentre
vodanile
migration
serversupport
uk
adtech
coronavirus
regtech
iot
azure
london
law
fintech
digitalmarketing
cloud
digital
webinar
gdprcompliance
healthtech
dsgvo
datagovernance
ethics
regulations
blog
datacompliance
dpo
biztip
eugdpr
databreaches
eu
infosec
ccpa
gdpr
compliance
digitaltransformation
dataprotectionofficer
dataprivacy
datavault
security
dataprotection
privacy
regulation
covid19
risk
databreach
data
datasecurity
cdo
cybersecurity
hipaa
marketing
cybercrime
britishairways
ai
tech
ico
healthcare
bigdata
personaldata
technology
brexit
ciso
cyber
business
socialmedia
eudatap
google
facebook
breach
eprivacy
startups


anuragthakur
whitesupremacy
misogyny
freedomofexpression
youtube
hatespeechbill
nigeria
india
domesticterrorism
truth
muslim
freedomofspeech
women
caa
fakenews
hatecrimes
shaheenbagh
censorship
democracy
media
trump
propaganda
discrimination
terrorism
maga
antisemitic
antisemitism
un
bigotry
lgbtq
disinformation
humanrights
gender
hate
racism
bjp
jews
america
muslims
policebrutality
ai
resist
hatespeech
cdnpoli
xenophobia
lgbt
coronavirus
rss_killingmuslims
twitter
rssterrorists
delhigenocide2020
hatecrime
auspol
delhiriots
usa
covid19
democrats
fascist
shameonyoumodi
caa_nrcprotests
modi
islamophobia
bjpisantinational
homophobia
nazi
humanrightsviolations
shaheenbaghfiring
racistinchief
freedom
freespeech
religiousfreedom
katyayan
shaheenbaghprotests
caa_nrc_npr
ireland
violence
shaheenbaghs
ethniccleansing
haters
worldagainstcaa_nrc_npr
unga
facebook
bjpburningindia
trump2020
fascism
lies
no2hatespeech
racist
socialmedia
violenceagainstwomenonline


business
edgecomputing
automation
smarthouse
dataanalytics
ml
javascript
cloud
rstats
4ir
golang
serverless
startups
reactjs
digital
linux
datascientist
vr
analytics
iiot
python
ar
ioe
datascience
cloudcomputing
security
robots
ai
bigdata
dl
blockchain
tensorflow
industry40
iot
machinelearning
fintech
raspberrypi
innovation
artificialintelligence
digitaltransformation
infographic
cybersecurity
deeplearning
infosec
internetofthings
tech
technology
5g
data
robotics
smartcities
smarthome
smart
startup
infographics
futureofwork
java
healthcare
smartcity
m2m


tor
regulation
sendoneprivately
humanrights
informationsecurity
safety
apple
socialmedia
crypto
bitcoin
healthcare
china
ethics
fintech
phishing
cyberattack
bipa
bigdata
ai
cybersec
blockchain
malware
iphone
digital
ransomware
government
hackers
business
technology
tech
private
privacyjobs
news
iot
hacking
infosec
cybersecurity
dpo
databreach
android
foxreddpaas
data
ccpa
security
privacy
vpn
compliance
ciso
cryptocurrency
dataprivacy
gdpr
dataprotection
cybercrime
law
encryption
online
eu
amazon
biometrics
cloud
facebook
datasecurity
surveillance
internet
personaldata
facialrecognition
network
marketing
hacker
google
hipaa
artificialintelligence
identity
breach
daps
cyber
california
tracking
machinelearning
privacymatters
innovation
faceapp
risk


computervision
megatrends
womenintech
microsoft
security
infosec
qubits
usa
quantumsupremacy
cryptography
5g
machinelearning
ml
qubit
encryption
python
cybersecurity
deeplearning
iot
ar
datascience
cloudcomputing
aws
computing
artificialintelligence
digitaltransformation
quantumtechnology
healthtech
quantumcomputers
bigdata
startups
algorithms
robotics
analytics
physics
tech
ai
ibmquantum
quantum
healthcare
braket
quantumcomputing
future
edgecomputing
ibmq
technology
digital
qis
ibm
google
data
innovation
fintech
science
defstar5
futureofwork
quantumphysics
smm
blockchain
augmentedreality
technews
quantumcomputer
cloud
seo
quantummechanics
ripple
hpc
crypto
eos
ethereum
research
scamalert
think2020
bitcoin
litecoin
qsharp
quantumteleportation
cryptocurrency
business
scam
emergingtech
startup



9.3 Appendix for deep dive into disinformation

Table 9.3.1: Overview of the numbers of selected tweets by the type of misinformation

	Number of tweets after keyword selection	number tweets for manual selection	Number of tweets after manual selection (max 100)	Number of stories	% number stories	% number tweets after keyword selection	% number tweets for manual selection	% number tweets after manual selection	mean number tweets after keyword selection per story	mean number of tweets for manual selection per story	mean number tweets after manual selection per story
cures	42285	2203	613	42	18.58	6.13	17.50	29.23	1006.79	52.45	14.60
virus	34444	2641	426	43	19.03	4.99	20.98	20.31	801.02	61.42	9.91
vaccine	239409	1803	196	22	9.73	34.70	14.32	9.35	10882.23	81.95	8.91
politics	344475	2788	215	53	23.45	49.92	22.15	10.25	6499.53	52.60	4.06

conspiracy	19190	1796	460	33	14.60	2.78	14.27	21.94	581.52	54.42	13.94
other	10201	1357	187	33	14.60	1.48	10.78	8.92	309.12	41.12	5.67
total	690004	12588	2097	226	100.00	100.00	100.00	100.00	3053.12	55.70	9.28

Table 9.3.2. Compound score results for comparisons of different types of COVID19-related misinformation. Note: p-values corrected for multiple comparisons using false-discovery rate correction according to Benjamini & Hochberg (1995).

Compound scores	Conspiracy -.098	Cures .073	Other .050	Politics .007	Vaccine .054	Virus -.124
Conspiracy -.098		1	1	1	1	.3987
Cures .073	.0004		.5882	.0617	.5882	.0004
Other .050	.0004	1		.3537	1	.0004
Politics .007	.0007	1	1		1	.0004
Vaccine .054	.0004	1	.9177	.2526		.0004
Virus -.124	1	1	1	1	1	