1. What is “fake news” and “hate speech” and how do they work in practice?

Authors: Lea Bader and Jochen Bender
Academic Supervisor: Silvia Ručinská and Catalin Vrabie

DOI: 10.24989/ocg.v.342.1

1.1. Introduction and definitions

Fake news and hate speech are not phenomena of the Internet age. Fake news and hate speech have been around since the beginning of human history – people have always lied and insulted. However, the emergence of social media has changed how, where and with what effects fake news and hate speech occur. Where lies and insults used to take place outside the Internet, fake news and hate speech are now increasingly shifting to social media.

To bring light to the concepts, the following pages attempt to create definitions for classifying fake news and hate speech. It will be shown what harmful impacts and what effects fake news and hate speech can have, what are the reasons for them to exist, trying, at the same time, to understand the people behind them. Furthermore, the article will provide facts and figures regarding people who are exposed to these impacts. The authors will identify legal and technical problems that are to be faced in fighting against fake news and hate speech.

For a better understanding, in chapter 1, the terms fake news and hate speech are analyzed separately, while in the rest of the article it will be shown the connection between the concepts with detailed explanations.

1.1.1. Fake news

“Fake”, or better “false”, are relatively newly introduced terms according to philosophy. The Latin term “falsum” originates back from the Romans; classical (pre-socratic) Greek philosophy had no corresponding word for that. The Greek term ψευδος or pseudos did not mean false, but rather “hidden”, “camouflaged” (i.e., an ancient Greek citizen who produced a ψευδος did this with intent, he did not err, but told something intending to mislead others [1-3]). The Roman “falsum”, on the other hand, requests a process of understanding and hence the establishment of absolute truth – not the highly subjective pseudos. Thomas Aquinas wrote, “Veritas est adaequatio rei et intellectus”, which implies precisely this existence of an absolute truth and uses the terms “correspondentia” and “convenientia” in this context [1-3] Further down in the article, when presenting the “Follow the science”-movement, explanations of the word “convenientia” will be given; one can see that a single opinion of the scientific community is hard to find.

To analyze fake news, we have to state that “fake” is an absolute which implies that its status of being not true but false can be determined without any doubt. And that such a “fake” can happen with or without intent (i.e., also by error). Individuals or computer apps spreading fake news need not necessarily know whether they are fake or not. In the case of apps, such as social bots or scripts, they do not “know” about their status by default, because computers lack judgement [1-13].

According to the Collins dictionary, the term “news” can be used in different contexts. On the one hand, it means information about the change of a situation or person in a general way and, on the other hand, it is used for information published in a newspaper or said on the radio or television. For the purpose of this article, we consider the second meaning much closer to our approach because we will analyze published news that seems to be real, but they aren't.[1]

However, the European Commission defined "disinformation" as being: "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit. The risk of harm includes threats to democratic political processes and values" [1-8].

Unfortunately, we do not have an agreed definition of "fake news", as the "Digital Resistance handbook for teachers" remarks [1-6].

The definition of "disinformation" contains, similar to the Greek word "pseudos" mentioned above, the intent of the spreader. The concept of fake news is more complex since it can be spread without harmful intent. Therefore, for the purpose of this book, we simply added "or without intention" to the above-mentioned definition bearing in mind that sometimes no public harm is intended, but rather the opposite. The famous "Pizzagate" shooter serves as a good example because he acted without harmful intent, rather the opposite[2] – details of this incident will be provided in Chapter 2.

According to the above, fake news can be spread with or without intent, therefore, there is a need to differentiate between different the two.

When false information is shared without any intention of causing harm, the proper term to use is "misinformation"; if there is an intention, "disinformation" should be used and, nonetheless, if the information is true but shared with the intention of causing harm, it is called "malinformation".

Phenomena like satire could be seen as fake news, however, this is not yet very clear because it can only cause harm if people don't understand the background context.[3]

This leads us to the problems this definition has. The term itself was used for many different phenomena over the past years. The inflationary use of it generated chaotic approaches with multiple definitions and phenomena (i.e., "hoax" is another term, very much used concerning fake news). Some researchers consider a published study, which turns out wrong afterwards, as being also fake news [1-14].

One very important aspect is, that the sharing of the (false) information causes harm. It is not important if this was or was not the intention of the one sharing it.

In regard to all the above, by the end of our study, we will provide a comprehensive definition of what fake news is or is not.

1.1.2. Hate speech

The term "hate speech" also lacks a clear definition [1-7]. However, its main purpose is to represent the extremely negative and threatening influence on social peace. According to the Council of Europe all statements that spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism, or other

[1] https://www.collinsdictionary.com/de/worterbuch/englisch/news (last accessed 22.12.2021)

[2] Vier Jahre Haft wegen Selbstjustiz im „Pizzagate"-Fall" https://www.faz.net/aktuell/gesellschaft/kriminalitaet/pizzagate-fall-mann-kriegt-4-jahre-haft-wegen-selbstjustiz-15073545.html (last accessed 07.01.2022)

[3] "Dealing with propaganda, misinformation and fake news" https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/dealing-with-propaganda-misinformation-and-fake-news (last accessed 22.12.2021)

forms of hatred based on intolerance are covered by the term “hate speech”.[4] A broader definition is given by the Committee of Experts on Combating Hate Speech of the Council of Europe in their background document.[5] The definition takes into account not only individuals but also groups of people and the negative stereotyping, stigmatization, or threatening of such people or groups of people based on “race”, color, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender identity, sexual orientation, and other personal characteristics or status, but also includes the term “hate speech” as a legal term, which refers to expressions that carry criminal, civil or administrative sanctions, such as incitement to hatred, insult, defamation, coercion, threat or public incitement to commit a crime. Other forms of the studied concept are to be found under the word like anti-Muslim racism, sexism, homophobia and transphobia (discrimination based on sexual orientation or gender identity), antiziganism (discrimination against Sinti and Roma), ableism (discrimination against disabled people), classism (prejudice based on social origin), lookism (discrimination based on appearance).[6] The definition of “hate speech” also reveals the different forms of it.

1.2. What can be a universal definition for fake news or hate speech?

In an effort to form a common basis for the topics of fake news and hate speech, it is necessary to define these terms. The respective definition should contribute to a better understanding and determine the meaning of the terms. With the help of the building blocks presented in Chapter 1, the definitions for fake news, hate speech, and freedom of expression will be created in this chapter.

1.2.1. Fake news

Since there is no clear definition of fake news, for the purpose of this book we define fake news as

> “false, inaccurate, or misleading information designed, presented and promoted to intentionally or unintentionally cause public harm or for profit.”

However, due to the problems of a simple definition, different forms of fake news should be distinguished. Here, the distinctions between disinformation, misinformation, and malinformation explained above lend themselves to consideration.

1.2.2. Hate speech

As mentioned already, there is no clear and universal definition of “hate speech”. A definition that most likely encompasses all the necessary aspects to show what should be included under the studied concept is the definition of the Expert Committee on Combating Hate Speech of the Council of Europe.

“Hate speech is to be understood as the advocacy, promotion or incitement in any form of denigration, hatred or disparagement of any person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat to such person or group of persons, and the justification of any of the foregoing on the grounds of 'race', color, descent, national or ethnic origin, age, disability, language, religion or belief, age, disability, language, religion or belief, sex, gender identity, sexual orientation, and other personal characteristics or status, as well as the form of public denial,

[4] “Hate Speech” - https://www.coe.int/en/web/freedom-expression/hate-speech (last accessed 20.10.2021)

[5] “Hintergrunddokument” - https://www.coe.int/en/web/committee-on-combatting-hate-speech/background-document (last accessed 19.10.2021)

[6] “Was ist hate speech” - https://www.bpb.de/252396/was-ist-hate-speech (last accessed 20.10.2021)

trivialization, justification or approval of genocide, crimes against humanity or war crimes found by courts of law, and the glorification of persons convicted of committing such crimes."

Aiming to create a distinction between "hate speech" and "freedom of expression", a further definition is required. This could be found in Article 10-1 of the Convention for the Protection of Human Rights and Fundamental Freedoms that stated as follows.[7]

"Everyone has the right to freedom of expression. This right shall include the freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises."

Based on the cited article, a possible definition for "freedom of expression" may be as follows.[8]

"Freedom of expression is to be understood as a fundamental right for everyone to be able to express his or her opinion freely and to be allowed to do so. Freedom of expression includes the receipt and dissemination of information and ideas. Freedom of expression is free from interference by public authorities and is not subject to any borders."

Hate speech is understood differently at the national and international level and, because of that, it is very difficult to build up a valid definition of it. The complexity of both terms brings even more difficulties in trying to define them. More questions arise when "freedom of expression", according to Article 10 of the European Convention on Human Rights, is restricted to avoid "hate speech". From here onwards, debates regarding censorship versus the right for everyone to freely express their opinion might pop up.

More detailed information regarding the distinction between "freedom of expression" and "hate speech" will be provided in Chapter 2.

1.3. What are the reasons for the existence of fake news and hate speech?

As mentioned at the beginning, fake news and hate speech are not unknown phenomena in human history. For example, in the Middle Ages, the Jews were blamed for the plague. They were said to have poisoned the wells. This was because they were less frequently affected by the plague. Today we know that they built their wells deeper for religious reasons and thus did not draw the contaminated surface water.[9] However, the question arises as to what reasons still exist today for fake news and hate speech to be used. In this chapter, the reasons for the existence of fake news and hate speech in different subject areas will be shown.

1.3.1. Fake news

There are various reasons for the creation of fake news. It is important here to go back to the distinction made above. As mentioned, fake news can be spread both intentionally and

[7] "European Convention on Human Rights" - Rome, 4.XI.1950, Seite 12, - https://www.echr.coe.int/Documents /Convention_ENG.pdf (last accessed 09.11.2021)

[8] "Freedom of expression and information" - https://www.coe.int/en/web/freedom-expression/freedom-of-expression-and-information (last accessed 09.11.2021)

[9] „Antisemitismus in Verschwörungstheorien" https://www.planet-wissen.de/gesellschaft/psychologie/ verschwoerungstheorien/verschwoerungstheorien-antisemitismus-100.html (last accessed 25.01.2022)

unintentionally. In addition, in case of doubt, it can even be real news that is merely formulated in such a way that it has a negative impact.[10]

1.3.1.1. Conspiracy theories

Often fake news is created in an environment of conspiracy theories. Often, this is one of the cases where the people creating and spreading fake news are not aware that they are doing so. On the contrary, they firmly believe that their worldview is true and the others are not able to see this.[11] A recent example of this is related to the Corona pandemic: from "the coronavirus is a harmless flu" to "when vaccinating, a chip is implanted for monitoring", almost everything is represented.[12]

1.3.1.2. Financial reasons

Financially, various scenarios are conceivable. Competing companies can be weakened by targeted disinformation, which can give one's own company a market advantage. But also, the other way around, by positive reporting for the own company it is possible to profit financially (i.e., if one manages to influence the stock market, he can gain huge profits).[13]

Another financial phenomenon is Clickbait, which is not necessarily subsumed under fake news. It describes the approach of various website operators to lure Internet users to the site with lurid headlines to increase traffic on the site. Through advertisements, higher revenues can also be achieved through more views. It is important to mention here that the news does not necessarily have to be false. It can also be true news but exaggerated or taken out of context.[14]

1.3.1.3. Political motives

Politically motivated fake news pursues the goal of bringing political change. They try to steer the mood in society in one direction. For example, the news is taken out of context or reproduced incompletely.[15] But even in videos, a few cuts are sometimes enough to convey a completely new message with what has been said.

1.3.1.4. For fun or satire

Sometimes fake news is created for fun. In case of doubt, the creator does not assume that anyone could take the message seriously, therefore, there is no malicious intent behind it [1-14].
Of course, some people enjoy deceiving other people; the so-called "trolls" enjoy the resulting attention.[16]

[10] "Was sind Fake News?" https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/was-sind-fake-news/ (last accessed 30.01.2022)

[11] Ibid.

[12] „Die verrücktesten Corona-Verschwörungsmythen - Darum sind sie falsch" https://www.mdr.de/brisant/corona-verschwoerungstheorien-100.html#sprung3 (last accessed 30.01.2022)

[13] "Fake News gefährden Unternehmen" https://www.capital.de/wirtschaft-politik/fake-news-gefaehrden-unternehmen (last accessed 05.01.2022)

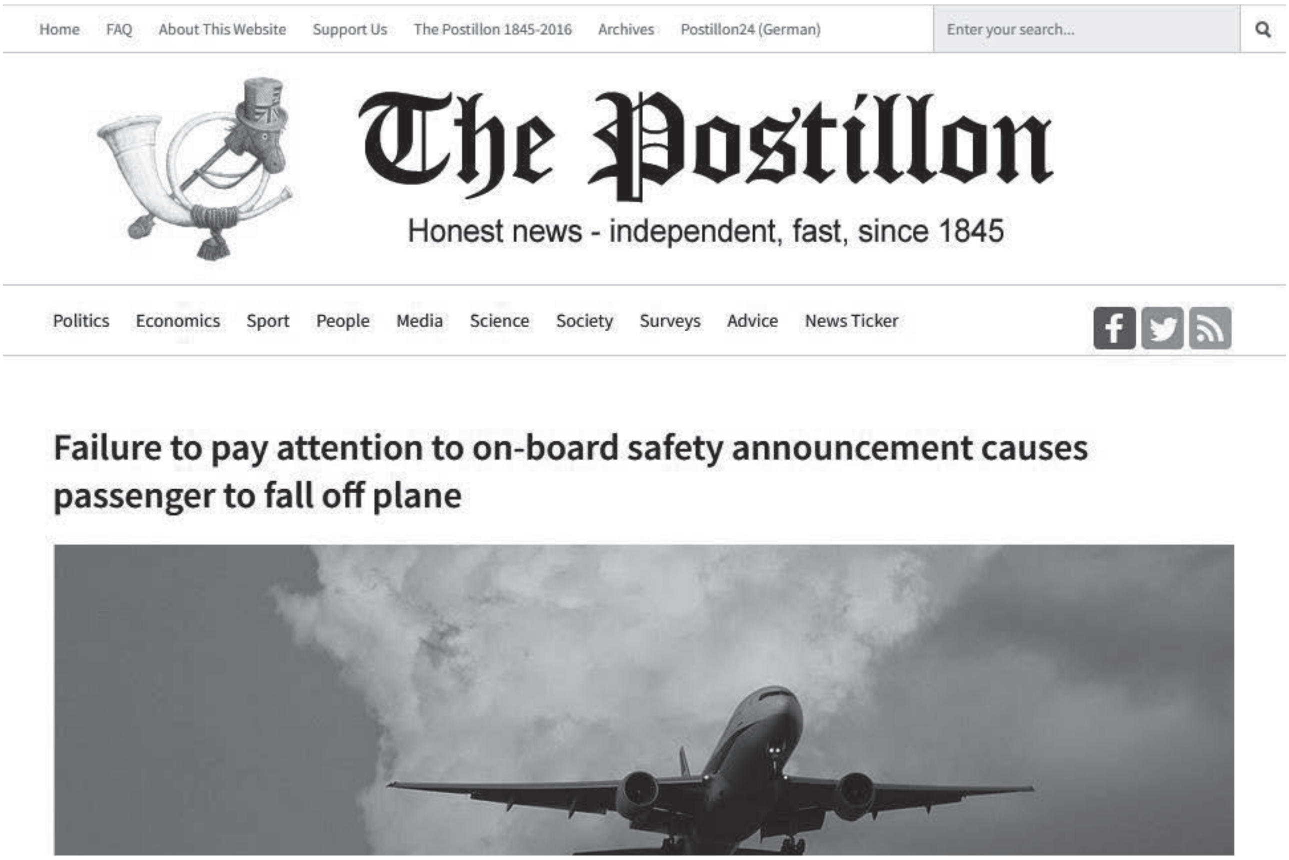
[14] "Clickbaiting – Was ist das?" https://www.ionos.de/digitalguide/online-marketing/verkaufen-im-internet/was-steckt-hinter-clickbaiting/ (last accessed 22.12.2021)

[15] "Was sind Fake News?" https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/was-sind-fake-news/ (last accessed 07.01.2022)

[16] Ibid.

Fake news is also created for satirical reasons. They may not belong to the classic fake news, but they can certainly be misinterpreted if someone does not realize that it is satire.

Satire usually uses humour, irony and exaggeration to point out a grievance or criticize something. This can be a behaviour of politics or even a social problem.[17] Some sites that disseminate satire in written form strongly resemble serious newspapers in their presentation. This can easily lead to misunderstandings.



Failure to pay attention to on-board safety announcement causes passenger to fall off plane

Figure 1: Example of a satirical newspaper[18]

1.3.2. Hate speech

Hate speech, similar to fake news has various reasons for its existence. These reasons will be shown as followed.

1.3.2.1. Social reasons

One reason for the existence of hate speech is that it attempts to achieve social or political change by dividing society. This is most often done by deforming enemy images. These images have at least one of the characteristics from the above definition of hate speech and belong to a single person or a whole group of people. Intending to achieve change, attempts are made to silence these individuals or groups of individuals or to influence their behavior in such a way that they are no longer willing

[17] https://www.lexico.com/definition/satire (last accessed 21.12.2021)
[18] https://www.the-postillon.com/ (last accessed 25.01.2022)

or able to freely express their opinions or continue to perform their jobs. In principle, any person can be attacked by hate speech. As a rule, those attacked are in the public focus of society and usually belong to the group of celebrities and most frequently to the group of politicians at all levels of government, as well as mayors, local councilors or other municipal volunteers [cf. 1-4]. The reason why politicians and municipal volunteers are attacked by hate speech is also that there are lower inhibition thresholds by abusing the perceived anonymity of the Internet and social media. This avoidable anonymity means that people are attacked more quickly and more easily on social media. The mass of attackers would not, or not in the same way, say what they said in social media if they would have the chance to meet the attacked persons in person.

1.3.2.2. Political motives

Another reason for the existence of hate speech is political interests. There is a deliberate attempt to influence political followers by verbally attacking the opposition. No other politician is better known for using social media to directly address his constituents and incite them to support his ideology through hate speech as the former U.S. President Donald Trump. During his term in office, Donald Trump used the news service "Twitter" like no other before him to inspire his supporters and mobilize them against his enemies and the opposition.

Figure 2: Former US-President Donald Trump at a speech[19]

On January 06, 2021, the United States Capitol in Washington was violently stormed by a large group of people and partially taken under their control. The crowd demanded that the U.S. Parliament, which

[19] https://www.brookings.edu/techstream/how-trump-impacts-harmful-twitter-speech-a-case-study-in-three-tweets/ (last accessed 25.01.2022)

was in session at the time, annul the results of the 46th presidential election. During this storming, several people were injured and a woman from the group of attackers was shot. She succumbed to her injuries a short time later.[20] The storming of the Washington Capitol was preceded by a speech and several tweets by former U.S. President Donald Trump. There is an assumption that the speech and tweets were intended to incite resistance and revolt against the newly elected U.S. government, following Donald Trump's election defeat. The suspicion is that it was the incitement in the speech and the tweets with hate speech by Donald Trump that made the mobilization of the group and the storming of the Capitol possible.[21]

Figure 3: Police and supporters of Donald Trump in Washington[22]

1.3.2.3. Personal reasons

Another reason for the existence of hate speech can be that people find pleasure in verbally attacking others. However, it is wrong to refer only to hate speech in this regard. The reason for this is the distinction between hate speech and freedom of expression. Even if the first impression of a text or a statement looks like hate speech, it can be a satire on closer inspection. What is meant here is the generic term "abusive criticism". One example of this is satirist Jan Böhmermann's "defamatory criticism" of Turkish President Recep Tayyip Erdoğan. The discussion about this satirical text was carried to the German Federal Court of Justice and even further to the German Constitutional Court.[23] Here, there is a dispute about when the freedom of expression ends and hate speech begins. As mentioned above, the distinction between hate speech and freedom of expression will be detailed in Chapter 2.

[20] "Vier Tote nach Sturm auf Kapitol" - https://www.tagesschau.de/ausland/kapitol-gestuermt-119.html (last accessed 14.01.2022)

[21] "How Trump impacts harmful Twitter speech" - https://www.brookings.edu/techstream/how-trump-impacts-harmful-twitter-speech-a-case-study-in-three-tweets/ (last accessed 27.10.2021), "Sturm auf das Kapitol und Trumps Twitter-Sperre" - https://jura-online.de/blog/2021/01/14/sturm-auf-das-kapitol-und-trumps-twitter-sperre/ (last accessed 27.10.2021)

[22] https://www.tagesschau.de/ausland/kapitol-gestuermt-119.html (last accessed 25.01.2022)

[23] "Schmähkritik" - https://www.zeit.de/gesellschaft/zeitgeschehen/2019-12/jan-boehmermann-affaere-bundesverfassungsgericht-schmaehkritik-gedicht (last accessed 27.10.2021)

1.3.2.4. Financial gains

It is difficult to earn money with hate speech and to cite it as a reason for existence. However, it can be argued that the operators are not interested to delete posts or tweets with hate speech. Basically, the operators want to have as many users as possible on their portals. Users are exposed to advertisements and those companies involved in the process are paying the operators to show the ads on their portals. It is more economical for the merchandisers if the portal where they post ads has large numbers of users. If there are comments with hate speech on the operator's portal and the operator would delete these comments or even ban the authors of the comments with hate speech or delete their accounts, the operator would reduce its number of users and make itself uninteresting for companies that want to show advertisements on social media portals. Therefore, it could be assumed that operators of social media platforms may find it annoying to delete hate speech comments in order not to lose their users. This view is shared by the former CDU member of the German Bundestag, Ruprecht Polenz, in an interview with Daphne Wolter, media policy officer at the Konrad Adenauer Foundation, entitled "The business model of Facebook and Twitter prevents a sensible debate culture" on 21 July 2021. In the opinion of Ruprecht Polanz, the algorithm of the platform would have to be changed so that users are kept on the platform, but not by the filter bubble created through the algorithm, but by pointing out other opinions and statements. The variety of opinions and statements would make a discussion possible again, and one's point of view would not be limited to just one path.[24]

1.3.2.5. Propaganda

Another connection for the use of hate speech and a related business model can be drawn with so-called troll factories. One such troll factory was uncovered by journalist Andrej Soschnikow in St. Petersburg, Russia.[25]

Figure 4: "Troll factory" in St. Petersburg[26]

[24] „Das Geschäftsmodell von Facebook und Twitter verhindert eine vernünftige Debattenkultur" - https://www.medienpolitik.net/2021/07/das-geschaeftsmodell-von-facebook-und-twitter-verhindert-eine-vernuenftige-debattenkultur/, (last accessed 24.11.2021)

[25] "Russische Trollfabrik" - https://www.spiegel.de/netzwelt/netzpolitik/russische-trollfabrik-eine-insiderin-berichtet-a-1036139.html (last accessed 10.11.2021)

[26] https://www.spiegel.de/netzwelt/netzpolitik/russische-trollfabrik-eine-insiderin-berichtet-a-1036139.html (last accessed 25.01.2022)

In an inconspicuous building, several hundred people are said to purposefully spread false news and generate more credibility through their comments. Trolls are individuals who attempt to disrupt discussions on a topic or comment on social media platforms with their comments, or to influence them in such a way that the desired reaction of other users is achieved. For this purpose, false or propagandistic postings, comments, pictures and videos are also posted. Trolls, in the analyzed context, have nothing to do with the creature from Norse mythology. The term comes from the English-speaking world and means "trolling with bait", meaning fishing with bait, which is pulled through the water to attract fish, snap and thus catch. According to the same principle, the trolls in social media try to attract attention with their comments and fake news and to influence other users through targeted manipulation. The NDR reporters also have documents showing that this troll factory in St. Petersburg belongs to a businessman who is very close to President Vladimir Putin.[27] It can be assumed that it would be a lucrative business model for him to employ trolls specifically for the current Russian government. It must also be worthwhile because different former trolls from this troll factory reported that they receive between 40,000 and 50,000 rubles (up to 800 euros), depending on the area of operation in the troll factory. English-speaking trolls in this troll factory are said to receive as much as 1,000 euros per month.[28] This would strengthen the assumption that hate speech can also be used as a business model.

1.4. What are the harmful impacts of "fake news" and "hate speech"?

Fake news and hate speech can harm various areas of life. These areas might refer to the society in general, but also the mental and physical health of individuals. Likewise, fake news and hate speech can hurt politics and government work. On the following pages, we want to show the harmful impacts fake news and hate speech can have.

The effects of fake news and hate speech on political discussion, democracy, economy, health and society is to be detailed more in chapter two.

1.4.1. Fake news

Fake news can have dramatic effects on various areas, such as politics or the economy, especially if they are used in a targeted manner. The following paragraph briefly shows the impact of fake news on different areas.

1.4.1.1. Society

Most at risk, of course, is society as a whole and, as a result, politics and democracy. First, people who do not recognize fake news as such inadvertently contribute to its spread. Older people are particularly affected, partly because there is a lack of educational offerings for these generations.[29] Young people are also at risk as they are less likely to inform themselves about multiple news sources and are more likely to end up in a so-called filter bubble. Such a filter bubble is created by the

[27] "Die Trolle" - https://www.ndr.de/fernsehen/sendungen/panorama_die_reporter/Die-Trolle,sendung524970.html (last accessed 10.10.2021)

[28] "Informationskrieg in der Ukraine-Krise" - https://www.focus.de/politik/ausland/propaganda-auf-bestellung-so-funktionieren-putins-troll-fabriken_id_4592188.html (last accessed 10.10.2021)

[29] "Desinformation: Experten sehen große Gefahr für Gesellschaft" https://www.br.de/nachrichten/deutschland-welt/desinformation-experten-sehen-in-fake-news-eine-grosse-gefahr-fuer-die-gesellschaft,SeHdE6w (last accessed 03.01.2022)

interaction of algorithms on social media platforms.[30] Chapters two will provide a closer look at how social media works and the impact of fake news on society.

1.4.1.2. Politics

Fake news can influence political events. For example, it is suspected that fake news may have contributed to Donald Trump's election, as 115 pro-Trump fake stories were shown to have been shared on social media around 30 million times. In comparison, only 41 pro-Clinton fake stories were shared about 7 million times.

There was also fake news in Germany related to the 2017 federal election, which was mainly spread by right-wing extremists and dealt with refugees and crime. However, this probably had less of an impact because, unlike in the U.S., social media in Germany plays a rather subordinate role in information gathering [1-11].

A deeper insight into the effects on political discussions will be given in Chapter two.

1.4.1.3. Economy

Fake news can also have a serious impact on the economy. Companies can be weakened by being targeted by disinformation campaigns. For example, attempts are made to prevent the recruitment of new specialists or to discredit the company management as well as badmouthing a product or influencing the share price – all these are possible points of attack. Sometimes, attacks are even launched against entire industries. The damage can run into millions, and because of the lack of traceability, no one can be held liable. [31]

One example in which an entire industry is attacked is that it is repeatedly propagated that electric cars have a worse environmental balance than internal combustion engines. The pollutants in the production of electric cars are supposedly to blame for this. However, this is backed up by outdated study data.[32]

1.4.1.4. Health

Although it may not be immediately obvious, fake news can certainly have a negative impact on the health of individuals. There are several examples of this, particularly in the context of the Corona pandemic. Be it people who drink bleach because they believe it helps against the virus.[33] Or people who do not get vaccinated for fear that the vaccination will make them infertile or could have other long-term consequences.[34]

[30] “Fake News und Verschwörungstheorien - Von Querdenkern, Social Bots und alten Säugetieren“ https://www.uni-ulm.de/universitaet/hochschulkommunikation/presse-und-oeffentlichkeitsarbeit/unimagazin/online-ausgabe-uni-ulm-intern/uni-ulm-intern-nr-354-dezember-2020/schwerpunkt-wissenschaftskommunikation/fake-news/ (last accessed 03.01.2022)

[31] “Fake News gefährden Unternehmen” https://www.capital.de/wirtschaft-politik/fake-news-gefaehrden-unternehmen (last accessed 03.01.2022)

[32] “Sind E-Autos doch Klima-Killer? – Der Faktencheck“ https://www.swr3.de/aktuell/fake-news-check/faktencheck-sind-e-autos-doch-klima-killer-co2-bei-herstellung-problematisch-100.html (last accessed 25.01.2022)

[33] “Familie verkauft Bleichmittel als Medikament gegen Corona – mehrere Tote“ https://www.rnd.de/panorama/familie-verkauft-bleichmittel-als-medikament-gegen-corona-mehrere-tote-M7FBC7I5CRDNHOUH5QDZDJFEYE.html (last accessed 04.01.2022)

[34] https://www.zusammengegencorona.de/impfen/basiswissen-zum-impfen/impfmythen/ (last accessed 04.01.2022)

But fake news can also have an impact on mental health. In a study that examined the effects of fake news on young women, almost half said they had already experienced anxiety, stress, sadness or depressive moods as a result of fake news.[35]

1.4.2. Hate speech

Nowadays, hate speech probably has the greatest impact and can cause a lot of damage. It has the power to divide and to create "friend-foe thinking" been called "intellectual arson" in the past.[36] Hate speech prevents a diverse way of looking at issues, as it only refers to one's social class or affiliation.[37]

1.4.2.1. Society

By dividing society, those affected by hate speech are forced into stereotypes, enemy images and groups. The classification of those affected and the resulting social stress and pressure can lead to physical and mental damage. Children and young people, in particular, suffer from hate speech in the form of cyberbullying. The stress factor of hate speech also increases for adults, depending on how much the hate posts hurt. In some cases, hate speech or cyberbullying, in all ages and social groups, leads to severe physical, psychological and social damage. Unfortunately, it is not uncommon for these afflictions to end in the suicide of the individual.[38]

1.4.2.2. Politics

Hate speech can also affect politics and governments. It becomes a problem in this area when no political opinion is expressed, but only insults. When these insults are also directed straight to politicians and members of parliament, a dangerous combination is created. The combination of "dividing society" and "physical and mental damage" has increasingly led to politicians and MPs deleting their social media accounts in the past due to the persistent occurrence of hate speech. Worse, hate speech and hostility towards politicians and MPs on social media have also led to an increase in physical assaults, bodily harm, and murders of politicians in the past [1-5].

Hate speech can also influence negatively the electoral behavior of voters. Democratic elections are characterized by free and fair competition in the power struggle for political office. Censoring hate speech in the election campaign can lead to dissatisfaction among the population, with the knowledge that not all opinions represented in the political competition are being shown. Censoring hate speech during election campaigns may be seen by citizens as a desirable means of protecting democracy, but the censored politician may not be able to express his or her opinion freely and may see this as unjustified [1-9].

After showing various possible reasons why hate speech could exist and what harmful impacts it might have, it is now important to show which people are behind that. Basically, hate speech occurs

[35] "Ihre Angst, unser Auftrag" https://www.zeit.de/zeit-magazin/leben/2021-10/fake-news-frauen-einfluss-falschinformationen-social-media-verunsicherung-vertrauensverlust-medien?utm_referrer=https%3A%2F%2Fwww.google.com%2F (last accessed 25.01.2022)

[36] "Geh sterben!" - https://www.amadeu-antonio-stiftung.de/w/files/pdfs/hatespeech.pdf (last accessed 26.11.2021)

[37] "Folgen für den gesellschaftlichen Zusammenhalt" - https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/hatespeech/folgen-fuer-den-gesellschaftlichen-zusammenhalt/ (last accessed 25.10.2021)

[38] "Formen von Cybermobbing" - https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/cybermobbing /formen-von-cybermobbing/ (last accessed 25.10.2021), "Betroffen sind Gruppen und Gruppenzugehörige" - https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/hatespeech/betroffen-sind-gruppen -und-gruppenzugehoerige-aber-auch-kinder-und-jugendliche/#footnote-1 (last accessed 25.10.2021)

in all social classes and all age groups.[39] In addition to the trolls already mentioned, there are also so-called "haters" and "faith-warriors". Haters and faith-warriors differ from trolls in that they consider their own opinion or worldview to be the only true one.

Haters are quick to use insults and personal verbal attacks as legitimate means. The hater does not want to understand the point of view of the counterpart or it lacks the fundamental understanding of it. The hater also feels safe due to the supposed anonymity of the Internet, which lowers his inhibition threshold to cover a fellow human being with mockery and insults.

The faith-warrior extends these characteristics of the hater in that he wants to stand up for an ideal or a conviction. The faith-warrior feels threatened, is driven by fear that a change in their existing and beloved worldview or view of humanity is imminent. To defend his opinion or worldview to others, the faith-warrior also insults his counterpart and is completely receptive to the opinion of others who think differently or to facts. The faith-warrior is firmly convinced of his ideology and sees his mission in saving the world and converting those who, according to his view, think wrongly, with all means. All of these three types use hate speech as a tool to hurt, insult or manipulate.

In addition to real people, technical aids also fuel the problem. Where troll factories are populated with humans, bots and software agents are executed as computer programs or as algorithms. Bots and software agents are the most commonly known terms in connection with fake news and hate speech. While bots perform repetitive tasks automatically without the need for interaction with a human user, software agents are capable of autonomous and self-dynamic behavior.[40] This means that no further external signal needs to be delivered to the software agent to make it respond. These two technical tools are often used in social media to respond to specific tweets or hashtags and to send prefabricated posts.[41] The distinction and potential impact of bots and software agents will be discussed in more detail in chapter three.

1.5. Ethics in social media

In an effort to understand why certain people, spread fake news or use hate speech in social media, it is necessary to take a look at the moral motivations; what political, monetary, or sociological reasons there might be and what are the effects. Also, very important is to know what typology of fake news and what types of haters we face. However, the question arises again as to which profound, individual character traits drive a person to spread fake news and hate speech. What morals do these people have? And is it these people and their moral concepts alone that make it possible for fake news and hate speech to be spread and to continue to be present in social media? Aren't the platform operators, such as Facebook (Meta) or Twitter, also significantly involved in the fact that fake news and hate speech continue to find their way into social media and are not consistently deleted? We will try to clarify this issue in the present chapter.

First of all, it is important to say, we start from the assumption that the interactions between users on social media platforms should be polite, respectful and constructive. This forms the basis for understanding what ethics is and how this term could be defined.

[39] ”Die Täter: von Trollen und Glaubenskriegern” - https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/hatespeech/die-taeter-von-trollen-und-glaubenskriegern/#/medien-und-bildung/jugendmedienschutz/hatespeech/die-taeter-von-trollen-und-glaubenskriegern/#c62149 (last accessed 10.11.2021)

[40] ”Social Bots” - https://wirtschaftslexikon.gabler.de/definition/social-bots-54247 (last accessed 18.11.2021)

[41] ”Agent • Definition” - https://wirtschaftslexikon.gabler.de/definition/agent-28615?redirectedfrom=42033 (last accessed 18.11.2021)

Ethics is the doctrine or theory of action according to the distinction between good and evil. The subject of ethics is morality.[42] Morality refers to the normative orientations to ideals, values, rules and judgments that determine or should determine the actions of people.[43] Media ethics deals both with ethics and morality of the media and with ethics and morality as applied in the media, i.e., in the content of the media.[44] Information ethics is concerned with the morality of the information society and morality in the information society. It examines how we, offering and using information and communication technologies and new media, behave or should behave in moral terms.[45]

These moral orientation values and the fundamental distinction between good and evil determine the actions of all people, at all times and everywhere. Every action or omission is based on the learned and internalized values, norms, rules and judgments from childhood to old age. These guiding values are not final and they may change by either strengthening or weakening over a lifetime. The changes can improve into good or deteriorate into evil. If these ethics and the moral concepts they contain exist in real life and the orientation values are almost always observed, why do these ethics not also lead to equal, respectful and constructive interaction in social media? The response to these behaviors of users in social media must be considered individually and is accompanied by different moral concepts and motivations for spreading fake news and using hate speech. In addition to the individual behavior of users, it is also up to the providers of the various platforms to implement their own community guidelines and to demand and implement compliance with these guidelines. One way to prevent fake news and hate speech in social media is to comply with netiquette.

> "Netiquette - as the term, a contraction of "net" and "etiquette", suggests – regulates behavior in computer networks and on the Internet. In a sense, it is the "etiquette" for communicating, interacting, and dealing with one another in communities, discussion forums, chats, and e-mail correspondence, and it aims to promote responsible behavior in the virtual realm as a whole."[46]

However, there is no binding basis in society for compliance with these netiquettes. Companies, like users in communities, can be asked to comply with these rules of conduct, and in some cases forced to do so. Facebook (Meta) presents the Facebook Community Standards in their Transparency Center. Here, Facebook (Meta) describes their approach to how to treat each other in the community and what types of information should be shared. It is divided into the following areas.[47]

> AUTHENTICITY
> "We want to ensure that the content users see on Facebook is authentic. We believe authenticity creates a better environment for sharing content. That's why we want to prevent people from using Facebook to misrepresent themselves or their actions and activities."

> SECURITY
> "Our goal is to make Facebook a safe place. Content that threatens users has the potential to intimidate, exclude or silence others. That's why it's not allowed on Facebook."

[42] "Ethik • Definition" - https://wirtschaftslexikon.gabler.de/definition/ethik-34332 (last accessed 20.12.2021)
[43] "Moral • Definition" - https://wirtschaftslexikon.gabler.de/definition/moral-38236 (last accessed 20.12.2021)
[44] "Medienethik • Definition" - https://wirtschaftslexikon.gabler.de/definition/medienethik-53884#panel-compact (last accessed 21.12.2021)
[45] "Informationsethik • Definition" - https://wirtschaftslexikon.gabler.de/definition/informationsethik-53486 (last accessed 21.12.2021)
[46] "Netiquette • Definition" - https://wirtschaftslexikon.gabler.de/definition/netiquette-53879 (last accessed 25.10.2021)
[47] "Facebook-Gemeinschaftsstandards | Transparency Center" - https://transparency.fb.com/de-de/policies/community-standards/?from=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards (last accessed 21.12.2021)

DATA PROTECTION
"We are committed to protecting the privacy and personal information. Within a framework of secure privacy, our users can be themselves, decide how and when to share things on Facebook, and connect with others more easily."

DIGNITY
"We believe that all people are equal in dignity and rights. Therefore, we expect them to respect the dignity of others and not harass or humiliate others."

To enforce these community standards, Facebook (Meta) provides a team of 15,000 reviewers in over 50 different languages, worldwide, to assess potential violations on their platform.[48] If a violation is detected, it will be removed, the violation will be listed and counted. In case of more frequent violations of the same account, it might get restricted. Finally, such an account can also be disabled, up to the removal of entire pages and groups.[49] In India for example, there are always lapses in adhering to these self-imposed community standards. Facebook (Meta) has been used for fake news and hate speech to incite hatred between Hindus and Muslims. India has the largest number of users of Facebook (Meta) and the budget allocated to India by the company to fight fake news and hate speech seems to be very small. This uneven distribution of budget has resulted in injuries and deaths due to Facebook's (meta) failure to strictly delete or flag fake news or disable accounts in India.[50]

Twitter also has community guidelines that are divided into the areas of security, authenticity and privacy. Twitter sees itself as a platform on which communication between people is to be promoted. Twitter wants to prevent any influence on this communication, no matter what kind. This is to ensure that all people can communicate freely and safely with each other.[51]

"Twitter exists to promote the public conversation. Violence, harassment, and other similar behaviors discourage people from expressing their opinions, ultimately harming the global public conversation. Our rules are designed to ensure that everyone can freely and safely participate in the public conversation."[52]

Twitter relies on a different way of approaching violations of its policies. Users report violations, either when they are personally affected or when they think a tweet violates the guidelines. Twitter checks these reports and weighs up whether action needs to be taken according to the context.

"So, *it depends on the context.* When deciding whether to take enforcement action, we may consider a number of factors, including:

- whether the conduct is directed against an individual, a group, or a group of people in need of special protection,

- whether the report was made by a data subject or by an uninvolved person,

[48] "Ermittlung von Verstößen | Transparency Center" - https://transparency.fb.com/de-de/enforcement/detecting-violations/ (last accessed 21.12.2021)
[49] "Ergreifen von Maßnahmen | Transparency Center" - https://transparency.fb.com/de-de/enforcement/taking-action/ (last accessed 21.12.2021)
[50] "FÜR GEWALTAUFRUFE MISSBRAUCHT - Facebook soll zu wenig gegen Hassbotschaften in Indien getan haben" - https://www.faz.net/aktuell/wirtschaft/digitec/facebook-soll-hassbotschaften-in-indien-ignoriert-haben-17601128.html (last accessed 21.12.2021)
[51] "Die Twitter Regeln" - https://help.twitter.com/de/rules-and-policies/twitter-rules (last accessed 21.12.2021)
[52] Ibid.

- whether the user has previously violated our policies,
- how serious the violation is,
- whether the content may be an issue of legitimate public interest."[53]

Twitter's enforcement actions are included at various levels. At the tweet level, for example, an account can be notified that its tweet does not comply with the guidelines. This is intended to prevent minor policy errors from being accompanied by severe penalties. Tweets can be flagged, turn invisible, hidden or with a request for removal by the user. At the direct message level, the potential violators of the policy can be blacklisted. Thus, there is no communication between the originator and the reporting party. The account level includes a temporary write block for the account or the permanent blocking of the account. Probably, the best-known account blocking in recent times was that of the former President of the USA, Donald Trump. In his case, Twitter blocked the account due to the assumption that further incitement to violence would occur in connection with the storming of the Washington Capitol.[54]

Unfortunately, the active reporting of hate speech on Twitter does not always seem to run smoothly either, as the incident in front of the German Twitter headquarters shows. Here, some still active hate speech was sprayed on the street. The person who painted these tweets on the street wanted to draw attention to Twitter's inconsistency in deleting fake news and hate speech.

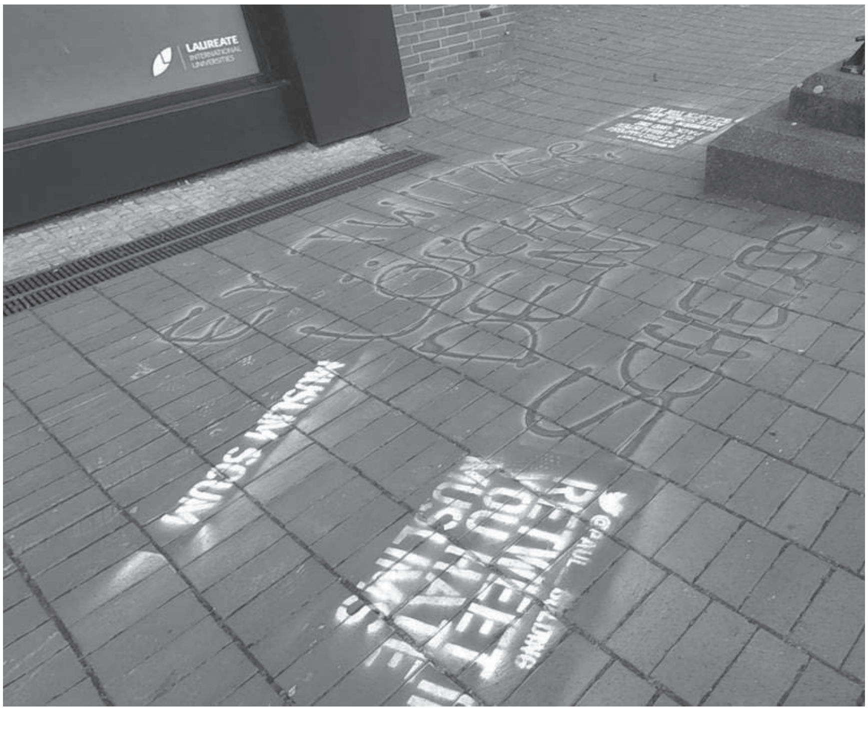


Figure 5: Racist tweets as graphite in front of Twitter headquarters Germany[55]

[53] "Die Vorgehensweise von Twitter bei der Entwicklung von Richtlinien und bei deren Durchsetzung" - https://help.twitter.com/de/rules-and-policies/enforcement-philosophy (last accessed 21.12.2021).

[54] "Konto des US-Präsidenten: Twitter sperrt Trump "dauerhaft"" - https://www.tagesschau.de/ausland/amerika/twitter-sperrt-trump-101.html (last accessed 21.12.2021).

[55] https://www.vice.com/de/article/5937qd/jemand-hat-rassistische-tweets-vor-die-deutsche-twitter-zentrale-gespruht (last accessed 25.01.2022).

One of the tweets was also deleted only after this action by the author himself, but only by the threat of criminal charges by another user, not because of the company's actions.[56]

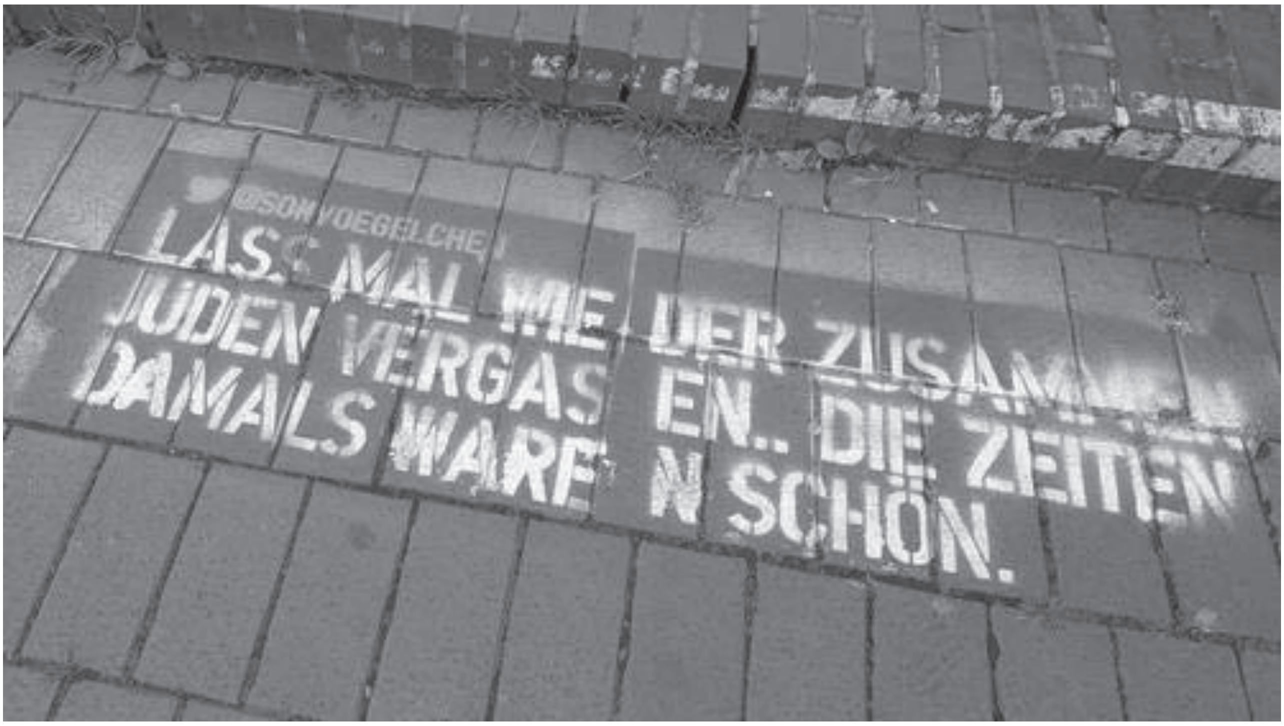


Figure 6: Racist tweets as graphite in front of Twitter headquarters Germany[57]

In the end, it's up to us how we treat each other and whether we report or fight fake news and hate speech on social media. More than ever, platform providers must implement their community guidelines.

1.6. Legal requirements and problems

To prevent fake news in social media, there is a need for clear legal principles, which, however, are very different in the EU and other countries. It depends on the prevailing political system in the respective country. The more Western-oriented the system, the more weight freedom of expression and freedom of information have in relation to fake news.

Aiming to prevent hate speech on the Internet, it would be sufficient if the aforementioned principles of the ethics of social media were adhered to. However, it is clear that these principles lose their meaning as soon as trolls, haters or faith-warriors come into play. A good example of legislation against hate speech is the mixture of criminal and civil law in Australia. This is based on a variety of criminal offenses. It will classify conduct as unlawful if it is reasonably believed to incite hatred, serious contempt or serious ridicule against a person based on their race. However, these criminal and civil law offenses have rarely if ever been applied [1-10].

The problem with fake news and hate speech, however, is not limited to who uses them but how and for what reason they are doing so. It also depends on the country, the social media platform on which

[56] ”Jemand hat rassistische Tweets vor die deutsche Twitter-Zentrale gesprüht“ - https://www.vice.com/de/article/5937qd/jemand-hat-rassistische-tweets-vor-die-deutsche-twitter-zentrale-gespruht (last accessed 21.12.2021).

[57] https://www.vice.com/de/article/5937qd/jemand-hat-rassistische-tweets-vor-die-deutsche-twitter-zentrale-gespruht (last accessed 25.01.2022)

the fake news or hate speech is used and whether a legal violation has been committed. It is important to show where the legal problems are and what a possible solution could be to combat fake news and hate speech and prosecute them legally. This question will be dealt with in detail in chapter four.

1.7. What can be a possible solution to encounter fake news or hate speech?

In our opinion, it is necessary to resolutely counter fake news and hate speech. In addition to the reasons for the existence and the negative influences of fake news and hate speech on various areas of society and politics, this paper also points out the technical and legal possibilities and problems. However, the technical and legal foundations make it difficult to counteract fake news and hate speech. A possible and practicable solution is the philosophy of Open Government in connection with Open Data. Through the disclosure of freely accessible data and the resulting transparency of the administration, it allows people, companies and organizations to inform themselves freely, independently and without biases. Chapter five will provide a more detailed insight into what Open Government is and how Open Data can help to combat fake news and hate speech.

References Chapter 1

[1-1] Council of Europe, Committee of Experts on Combating Hate Speech (ADI/MSI-DIS), Background Document, 25 May 2020, page 2 of 8

[1-2] ECRI General Policy Recommendation No.15 on Combating Hate Speech: key points, March 2016, page 2, Topic B

[1-3] Faßrainer, W. and Müller-Török, R.: „Der Wahrheitsbegriff der Rechtswissenschaften im Lichte der Philosophie":in: Tagungsband des 13. Internationalen Rechtsinformatik Symposions – IRIS 2010, ISBN 978-3-85403-226-3, S. 535-540, 25.-27. February 2010, Salzburg.

[1-4] Geschke Daniel, Klaßen Anja, Quent Matthias, Richter Christoph - „ #HASS IM NETZ: DER SCHLEICHENDE ANGRIFF AUF UNSERE DEMOKRATIE", Institut für Demokratie und Zivilgesellschaft, ISBN: 978-3-940878-41-0, Juni 2019, Kapitel 4.2, Seite 24, Abbildung 14, last accessed 27.10.2021

[1-5] Geschke Daniel, Klaßen Anja, Quent Matthias, Richter Christoph - „ #HASS IM NETZ: DER SCHLEICHENDE ANGRIFF AUF UNSERE DEMOKRATIE", Institut für Demokratie und Zivilgesellschaft, ISBN: 978-3-940878-41-0, Juni 2019, Kapitel 5.2, Seite 28, last accessed 26.10.2021

[1-6] Digital Resistance - An empowering handbook for teachers on how to support their students to recognise fake news and false information found in the online environment, 2020, Council of Europe, ISBN: 978-92-871-8715-4

[1-7] Stahel Lea - Status quo und Maßnahmen zu rassistischer Hassrede im Internet: Übersicht und Empfehlungen, Soziologisches Institut, Universität Zürich, August 2020, Kapitel 3.1, Seite 5, last accessed 20.10.2021

[1-8] A multi-dimensional approach to disinformation, 2018, European Commission, ISBN: 978-92-79-80420-5

[1-9] International Journal of Public Opinion Research, Vol. 26 No.4 2014 - "The Way Democracy Works: The Impact of Hate Speech Prosecution of a Politician on Citizens' Satisfaction With Democratic Performance", Oxford University, Press on behalf of The World Association for Public Opinion Research

[1-10] Gelber Katharine and McNamara Luke, Australian Journal of Human Rights - "Changes in the expression of prejudice in public discourse in Australia: assessing the impact of hate speech laws on letters to the editor 1992-2010", ISSN: 1323-238X

[1-11] Doublet, Yves-Marie: Disinformation and electoral campaigns, 2019, Council of Europe, 978-92-871-8911-0

[1-12] ZANKOVA, B. (2019). Smart society "Fake analytica" style? Smart Cities and Regional Development (SCRD) Journal, 3(1), 63–78. Retrieved from https://scrd.eu/index.php/scrd/article/view/47

[1-13] Are Computers Already Smarter Than Humans? LANCE WHITNEY, Time. https://time.com/4960778/computers-smarter-than-humans/ last accessed 28.01.2022

[1-14] Wardle, Claire/Derakshan, Hossein: Information Disorder - Toward an interdisciplinary framework for research and policymaking, 2017, Council of Europe