books@ocg.at

Sebastian Brüggemann, Nervin Kutlu,
Robert Müller-Török, Alexander Prosser, Silvia Ručinská,
Tamás Szádeczky, Catalin Vrabie (eds.)

Counterfake

A scientific basis for a policy fighting fake news and hate speech

OESTERREICHISCHE
COMPUTER GESELLSCHAFT
AUSTRIAN
COMPUTER SOCIETY

facultas

COUNTERFAKE

A scientific basis for a policy fighting fake news and hate speech

Sebastian Brüggemann, Nervin Kutlu, Robert Müller-Török, Alexander Prosser,
Silvia Ručinská, Tamás Szádeczky, Catalin Vrabie (eds.)

COUNTERFAKE

A scientific basis for a policy fighting fake news and hate speech

facultas

COUNTERFAKE
A scientific basis for a policy fighting fake news and hate speech

Written by students of the minor "applied e-Government" of the University of Public Administration and Finance Ludwigsburg, Germany

Editors and academic supervisors:

Sebastian Brüggemann
University of Public Administration and Finance Ludwigsburg, DE
Nervin Kutlu
Vienna University of Business and Economics, AT
Robert Müller-Török
University of Public Administration and Finance Ludwigsburg, DE
Alexander Prosser
Vienna University of Business and Economics, AT
Silvia Ručinská
Pavol Jozef Šafárik University in Košice, SK
Tamás Szádeczky
Budapest University of Technology and Economics, HU
Catalin Vrabie
National University of Political Studies and Public Administration, RO

Authors:

Lea Bader	Jochen Bender
Ines Beutel	Anna Funk
Christian Hönig	Konstantinos Katevas
Olga Kirschler	Sabrina Kokott
Christian Munz	Timo Steidle
Timo Vogt	Max Winter
Erik Wurzbach	

Foreword by Dr Andreas Kiefer, General Secretary of the Congress

In 2021, the Congress of Local and Regional Authorities and five academic partners launched a four-month collaborative research project, entitled "Open-Government and Open-Data Against Fake News and Hate Speech". The Congress welcomes the publication of the present research paper which constitutes the main outcome of a successful collaboration and will nurture the future Congress activities supporting local and regional politicians of the Member States of the Council of Europe against fake news and hate speech.

Notwithstanding the fact that there are many opportunities offered by new communication technologies to politicians at all levels of government – allowing for easily accessible and low-threshold communication with citizens – European local and regional politicians, mayors, councilors and elected politicians in regional governments and assemblies have increasingly been exposed to online hate speech and disinformation. The Covid-19 pandemic has further exacerbated the negative impacts of hate speech and fake news on local and regional elected representatives, especially during election campaigns. In addition, information wars, including disinformation campaigns and cyberattacks are becoming commonplace. Online threats by disgruntled citizens or coordinated online campaigns can turn easily into physical attacks, as has been illustrated by the tragic death of the Mayor of Gdansk, Paweł ADAMOWICZ, in 2019 or the assassination attempt on the Mayor of Cologne, Henriette REKER, in 2015 among many others.

The Congress acknowledges the prominence of this issue and held several debates on mayors under pressure in 2018 and 2019 and a debate on "How to preserve democracy in the face of fake news and hate speech" in October 2021. Congress members shared their experiences and the threats they and their families were exposed to. The wide-range of experiences and worries voiced by local and regional politicians during these debates, as well as the variety of ad-hoc solutions explored, highlighted the need for a more complete review of legal and technological frameworks and for a better understanding of how to approach fake news and hate speech. Most of the time, local elected representatives are not prepared to handle such threats or not well aware of legal and technical solutions and tools that are available to them. Some may feel increasingly helpless.

As a result, hate speech and fake news have a paralysing effect on our democracies and political life on local and regional levels. This negative trend creates a toxic atmosphere of fear and confusion among local and regional politicians as well as among citizens of municipalities and regions which disturbs societal cohesion. And it makes interested citizens think twice whether to sacrifice their privacy for serving the people and may put them off standing in elections.

This study shows the concrete impacts of hate speech and fake news on working conditions of local and regional politicians, which were until this report under-evaluated. Better awareness of this negative phenomena was a pre-condition for the Congress to efficiently support local and regional politicians. This study helps to better apprehend how fake news and hate speech work in practice, how they affect the political debate and democracy and how they can be detected. Moreover, the present report clarifies the functioning of social media and the limitations for technical remedies against the platforms allowing for the spread of fake news and hate speech. It explores the benefits of open government and open data, which the Congress has also highlighted in previous resolutions, as a potential tool to improve democracy at local and regional levels, in areas such as budgeting, law making, contracting, policy making, service delivery and in promoting the involvement of citizens in local public life. Furthermore, the evidence gathered in the questionnaire shows that local and regional politicians need not only reliable "hardware", such as rules and law, but also a compatible "software", such as skills, information, resources to address the challenges. This report successfully sheds light on both sides of the challenge.

The Congress collaborated to the "Open-Government and Open-Data Against Fake News and Hate Speech" research project by contributing to the design and inception of the project as well as to the dissemination of the research questionnaire to all its members and youth delegates. Close to 200 members of the Congress took this opportunity to share their experience and ideas of effective remedies against fake news and hate speech. Responding to this questionnaire provided highly valuable input to the research team and highlighted that lack of technical or legal knowledge and skills or confusion vis-à-vis the solutions available were also commonplace.

The Congress would like to thank for their involvement all academic institutions participating in this project including the WU Vienna – Institute of Production Management (Austria), the University of Public Administration and Finance Ludwigsburg (Germany), the Budapest University of Technology and Economics (Hungary), the Faculty of Public Administration at National University of Political Studies and Public Administration Bucharest (Romania), and the Pavol Jozef Šafárik University Košice (Slovakia). The Congress would also like to thank Professor Robert MÜLLER-TÖRÖK, of the University of Public Administration and Finance Ludwigsburg, for facilitating the collaboration between the Congress and the research team. Finally, the Congress would like to express gratitude to all Congress members and youth delegates who devoted some of their precious time to contribute to the questionnaire.

The Congress, as the voice of Europe's municipalities and regions, will work to pursue the conclusions and remedies put forward in this report. These will serve as the basis of a policy report to be adopted by the Congress, providing guidelines and support to local and regional elected representatives. Additionally, future design of Congress cooperation activities in member States most affected by the spread of fake news and hate speech may be informed by the present research report.

Promoting fundamental rights online and offline, the Congress reiterates its attachment to strengthening the protection of the rights of local and regional elected representatives. It is crucial to ensure a meaningful and democratic political life at local and regional levels and to serve the needs of all citizens.

Table of contents

Introduction and Management Summary

Every year, typically during the summer holidays, the minor "applied e-Government" of the Public Administration course programme at Ludwigsburg University chooses a topic where the students learn how to solve complex problems in public administration by means of working in project organization. When looking for an appropriate topic for the winter term 2021/2022 the topic of hate speech and fake news came up, particularly because of an increasingly heated Corona debate. Discussions with politicians, practitioners and researchers led towards the Congress of Local and Regional Authorities and, with great help from the General Secretary and his team, the idea became a real project.

The goal was to provide the member states and delegates of the Congress with

1. Knowledge about the way fake news and hate speech work;

2. Knowledge about the technical and legal foundations of both and, more generally, how the internet works;

3. A probable remedy, which should be verified or at least supported by

4. A questionnaire, collecting the delegates' and youth delegates' opinions and views, especially on assessing their technical and legal knowledge and their views on potential remedies.

The goal was to provide a scientific basis for a discussion in the Congress and probably also in the Parliamentary Assembly of the Council of Europe and, hopefully, to contribute to future policy recommendation.

One of the key findings was that many Congress members are not fully aware of the technical and legal intricacies of the topic, which creates a requirement for more training and education opportunities for Congress members. As shown in the text delivered, things like blocking IP or email addresses, tracing people or computers producing hate speech and prosecuting them are more complex and not so easy to achieve in the "real virtual world" of the internet. State frontiers, which are often not perceived when surfing the web, hinder effective enforcement of national laws and regulations – and often they are simply not applicable.

We drew the parallel between the postal services and the internet, the former being well regulated since 1874 when the Universal Postal Convention took place and the Treaty of Bern came into effect. Whilst the latter is unfortunately still largely unregulated regarding international treaties and enforceable standard.

What would work, both in our opinion and in the views collected from the delegates, is providing more transparency, more open government and more open data – thereby reducing the necessity for "filter bubbles" to produce hate speech according to Niccolò Macchiavelli's famous proverb, that a man who find himself treated unfairly, will always find a way towards fairness – at least in his perception.

We hope to have contributed towards a discussion on the necessity and the limits of a policy recommendation and thank the Congress for this great opportunity.

Budapest, Bucharest, Kosice, Ludwigsburg and Vienna in March 2022

The authors and editors

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

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.

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]

<div align="center">

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."

</div>

---

[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/5937 qd/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

2. How do fake news and hate speech affect political discussion and target persons and how can they be detected?

Authors: Ines Beutel, Olga Kirschler and Sabrina Kokott
Academic supervisor: Robert Müller-Török and Nervin Kutlu

Fake news can have effects, especially in an election or referendum context and so can hate speech. This chapter describes how these effects occur and how internet-based hate speech may turn into real-life physical violence or lead to other consequences in real life.

This chapter also focuses on how fake news and hate speech can be identified, especially in a Social Media context and, given the complexity that, at least to some extent, the line between freedom of expression and hate speech is difficult to identify. Fake news and hate speech may also be used to exercise - likely undue - influence in an organized manner, whether it be astroturfing[58] by lobby groups or influence exercised by both domestic and foreign governments or actors.

This influence is difficult to detect, however, there are means of semantic text and network analysis that may indicate such organized actions. The chapter will provide a survey of existing approaches and their respective applicability.

2.1. The distinction between Freedom of Expression and hate speech and fake news

Freedom of expression is a fundamental human right. On the one hand, it is indispensable in the human rights system, and on the other hand, it is crucial for the functioning of a democratic society. Because of its importance, freedom of expression has been enshrined in the Universal Declaration of Human Rights (UDHR, Article 19) [2-1] and in all major international and regional human rights treaties [2-2].

"Article 19 [UDHR]
Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers." [2-1]

In the European Convention on Human Rights (ECHR), this right is protected by Article 10 [2-2].

"ARTICLE 10 [ECHR]
Freedom of expression
1. Everyone has the right to freedom of expression. This right shall include 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.

---

[58] Astroturfing is, according to Merriam-Webster, "organized activity that is intended to create a false impression of a widespread, spontaneously arising, grassroots movement in support of or in opposition to something (such as a political policy) but that is in reality initiated and controlled by a concealed group or organization (such as a corporation)". Cf. https://www.merriam-webster.com/dictionary/astroturfing last accessed 09.12.2021).

2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary." [2-3, p. 12]

In addition to freedom of expression, Article 10 includes the freedom to receive and impart information without interference by public authorities or other restrictions. However, these are not absolute rights. Nations can restrict these rights if there are legitimate reasons to do so. Possible reasons might include national security, public health, or the protection of other rights [2-2]. However, the legal situation in the context of false or falsified information (fake news) or information containing hate messages (hate speech) is questionable.

2.1.1. Fake news

Regarding fake news, it is difficult to even distinguish whether it is fake news or not, because there is currently no overall or generally accepted definition. According to the general linguistic usage, it is news that has been deliberately spread falsely via the Internet or social networks. However, false statements can be differentiated. In a criminal prosecution, the decisive factor is what type of communication is involved. The dissemination of news can either be facts or personal opinions. The difference between these two types is that facts can be verified or falsified, hence they can be true or false. Contrary to that, expressions of personal opinions cannot be refuted. Correspondingly, they can neither be true nor false. Only if expressions of personal opinions are based on false facts, they can subsequently be ruled false. While personal opinions are protected by the right to freedom of expression, false facts and opinions based on false facts do not fall within the scope of this protection. If an author deliberately writes a false message, he potentially manipulates the reader. False information provided to the reader could lead to the reader taking the actions desired by the author. So, if an author deliberately puts false information into the world, he could even be liable to prosecution under e.g. German law [2-4, pp. 6-8].[59]

2.1.2. Hate speech

However, the right to freedom of expression may be in conflict with other rights. The right to freedom of expression ends, where the protected interests of other persons are violated. To ensure this protection, Article 8 of the European Convention on Human Rights ensures the right to respect for private and family life.

"ARTICLE 8 [ECHR]
Right to respect for private and family life
1. Everyone has the right to respect for his private and family life, his home and his correspondence.

2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others." [2-3, p. 11]

---

[59] https://www.deutschlandfunkkultur.de/fake-news-vorsaetzliche-luegen-muessen-verboten-werden-100.html (last accessed 05.12.2021)

Because hate speech is spread against other groups or individuals and also incites people, it is an abuse of the right to freedom of expression (Article 10 European Convention on Human Rights) and could violate Article 8. However, the use of hate speech is dealt with differently in different nations. For example, the American Bar Association considers hate speech to be legal and protected by the First Amendment, as long as it does not directly incite violence. There are many and various reasons why hate speech should be permitted [2-5, p. 1].

One reason to allow hate speech within the US constitutional system is the theory of the *marketplace of ideas*. According to this theory, all ideas, even the bad ones, should be heard to find the truth. In the free market, the truth should also compete with falseness. It is assumed that in the end, the truth will win this competition. Therefore, each individual should be able to communicate his or her opinions and ideas, so that the best among them can prevail [2-6, pp. 13-14]. Another reason to permit hate speech is the *democratic process*. In a democracy, any expression of opinion should be allowed. Citizens should also have access to all the information they need to educate themselves and make well-thought decisions, such as voting [2-6, p. 15]. Lastly, the theory of *personal liberty* can be applied. According to this theory, every person deserves the right to unrestricted expression, even if it contains hateful or fanatical statements. Expressions of opinion are a human being's freedom and essential for the development of one's potential. Hence, restricting them is a massive interference in human freedom and development [2-6, p. 16].

Unlike the USA, some countries ban hate speech, including Germany, Rwanda and Myanmar. These countries have already experienced that language can have a great effect. It has there historically happened that prejudice against a group has manipulated people and incited them to violence [2-6, p. 16]. In *Germany*, these scenarios happened during the National Socialism with the so-called "Jewish problem". Jewish people were massively attacked with hate speech. Dehumanizing terms were used by calling Jews vermin or snakes. In addition, ethnophobic statements and the attribution of unfavorable characteristics to Jews were used to incite hatred against them. By comparing Jewish people to animals and creating an "us versus them" feeling, mistreatment and violence against members of this group became commonly accepted [2-6, p. 21]. In *Rwanda* in the 1990s, Hutus, who constitute the majority of the population, spread hate speech against Tutsis. This encouraged ordinary citizens and militiamen to carry out mass killings. Between five hundred thousand and one million civilians fell victim, and the Tutsi population was reduced by 75 percent [2-6, p. 21]. Similar to Germany and Rwanda, dehumanization and the creation of an "us vs. them" feeling has led to violence against groups in *Myanmar*. In predominantly Buddhist Myanmar, Muslims belonging to the Rohingya people have been discriminated against for years. Since 2017, the Rohingya have been victims of brutal violence, including rape, murder, and arson. As a result, some seven hundred thousand Rohingya have fled since then [2-6, p. 23].

What these three examples have in common is that the leaders have convinced the citizens to take action against a particular group. The given reason for this incitement was that their own lives and livelihoods were in danger. They portrayed the group as a problem that could cause significant harm. In all cases, mass media were used to spread misinformation and hate speech. In today's world, non-official media, such as social media, also play a particularly important role. Through social media, anybody can create misinformation and spread it at a rapid speed. Because of this possible threat and the past experiences, the mentioned countries now try to prevent the repetition of these events through special laws, e.g. prosecuting those who praise the Holocaust in Germany [2-6, p. 25].

As shown, there are good reasons to legalize hate speech but also to ban it. To decide whether statements are admissible or inadmissible, the European Court of Human Rights follows two

approaches based on the European Convention on Human Rights. First, it examines whether the statement violates the fundamental values of the Convention. The abuse of rights is prohibited by Article 17 of the European Convention on Human Rights [2-7, p. 1]:

“ARTICLE 17 [ECHR]

Prohibition of abuse of rights

Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention.” [2-3, p. 14]

Finally, if the expression constitutes hate speech but does not restrict the fundamental values of the Convention, Article 10 (2) of the European Convention on Human Rights is invoked (see above). This paragraph is finally entitled to restrict hate speech if there is another legitimate interest [2-7, p. 1]. The right to freedom of expression and the right to respect for private and family life are indeed of equal importance, so the margin of appreciation should be equal [2-8, p. 17]. Therefore, it is difficult to decide, if a statement still falls below the protection of free expression or if it is hate speech, hence prohibited. The balance must be struck between the need to protect freedom of expression and the need to protect the individual's rights, respect within society, or public order. For the balances, the European Court of Human Rights has developed extensive case-law on hate speech and incitement to violence [2-2]. In doing so, they defined the following criteria: It has to be weighed up whether the contribution is in the public interest and how high the level of awareness of the person concerned is. Furthermore, the subject of the news report is decisive, as well as the prior conduct of the person. In addition to that, the content, form and consequences of the publication are evaluated and, if applicable, the circumstances under which photos were taken. Furthermore, the Court examines how the information was obtained and its true nature. Finally, the severity of the punishment is put into perspective [2-8, p. 17].

One example in which this balancing was applied is the case of “*von Hannover v. Germany (No. 2)*”. Here, two German newspapers published two photos showing an aristocratic family on vacation. The European Court of Human Rights ruled that these photos violated the right to privacy under Article 8 of the European Convention on Human Rights, as this information did not reflect the interest of the public. A third photo showed a prince in poor health. However, the health condition of the well-known prince is a case of public interest, so Article 8 of the European Convention on Human Rights was not violated.[60] Another example is the case of *“Axel Springer AG v. Germany”*. Here, a magazine published articles about the arrest of an actor for cocaine possession. The actor felt that his right to privacy had been violated, which is why the magazine was fined and prohibited from publishing further articles about the arrest. However, the European Court of Human Rights ruled in this case that these penalties were disproportionate and that the right to freedom of expression had been violated. The reason why freedom of expression outweighs the right to privacy is that the case involved judicial facts about a person known to the public. Also, the person was arrested in a public place, albeit for a minor crime. Even if the punishment was mild compared to the magazine, it still was disproportionate to the legitimate goal pursued.[61]

---

[60] https://globalfreedomofexpression.columbia.edu/cases/von-hannover-v-germany-no-2/ (last accessed 10.12.2021)
[61] https://globalfreedomofexpression.columbia.edu/cases/axel-springer-ag-v-germany/ (last accessed 10.12.2021)

2.2. How fake news can be identified, especially in a social media context

Fake news are usually already structured differently than serious news. Often, fake news can already be recognized by their lurid writing style, their emotionally oriented texts and many exclamation and question marks. Most of the time, these texts are illustrated with spectacular pictures. This eye-catching packaging is designed to attract readers.[62]

The following example shows an article from an online British daily newspaper. It reports that the man in the photo allegedly married a three-meter-long cobra. Allegedly, he would believe that his deceased wife reincarnated into her. With the modified phrase "You may hiss the bride", the author tries to turn the attention to the article already in the headline via the ridiculous writing style. Already here the first skepticism would have to arise. If the source is examined further, the article can be identified as a hoax. When searching for the keywords, such as "man", "cobra" or "Southeast Asia", the true story behind the picture can be found. The man works as a fireman and specializes in catching snakes. The man had posted the photo on his Facebook page, the text is made up. In addition, the source "Daily Mail" is a British tabloid newspaper. These are gossip rags with lurid headlines that always take a dim view of the truth.[63]

---

[62] https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/ (last accessed 20.10.2021)
[63] https://www.geo.de/geolino/magazine/22568-rtkl-falschnachrichten-im-internet-so-erkennt-ihr-fake-news (last accessed 21.12.2021)

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money |
Latest Headlines | The Queen | Royals | Prince Harry | Meghan Markle | World News | Covid 19 | Black Friday

You may hiss the bride! Man marries a 10ft COBRA he believes is his reincarnated girlfriend

- The 10ft cobra was spotted by the unidentified husband in South East Asia
- The heartbroken man said it bore a 'striking resemblance' to his dead lover
- They spend every day together playing board games and going to the gym
- Unidentified man is understood to have taken inspiration from Buddhism

By GARETH DAVIES FOR MAILONLINE
PUBLISHED: 10:36 GMT, 11 November 2016 | UPDATED: 10:36 GMT, 11 November 2016

f Share | 9.2k shares | 51 View comments

A man has married his pet snake because he believes his dead girlfriend came back to life as the cobra.

The snake was spotted by the man in South East Asia, who said it bore a 'striking resemblance' to his former lover.

He and the 10ft serpent now spend every day together watching TV, sharing romantic picnics by the lake, playing board games and going to the gym as a couple.

A man has married his pet snake because he believes his dead girlfriend came back to life as the cobra

Figure 7: Screenshot of the daily mail homepage[64]

---

[64] https://www.dailymail.co.uk/news/article-3926868/You-hiss-bride-Heartbroken-man-marries-PET-SNAKE-believes-dead-girlfriend-reincarnated.html (last accessed 21.12.2021)

Another example shows a Twitter entry with a shark allegedly swimming on a highway. This occurrence should have taken place during Hurricane Harvey in August 2017. The entry is from an unknown person, which is rather unreliable. The better sources are websites of major daily newspapers. The image can be verified by inserting a screenshot, for example, in Google reverse search. The search shows that the shark always appears after hurricanes. So, it is a photomontage composed of several photos.[65]

Figure 8: Twitter entry showing a shark on the freeway[66]

As can be seen from the examples above, some measures can be taken for detection, to be sure that certain information is fake news on the Internet. This chapter will show how fake news can be recognized. Guidelines from various institutions (Landesmedienzentrum Baden-Württemberg, the European Union and the International Federation of Library Associations and Institutions (IFLA)) are used for this purpose.

2.2.1. Structure of the message

First, the structure of the message should be looked at. As described above, fake news often has a comical writing style.[67] Especially in the headline, fake news often uses capital letters and many exclamation marks. Sometimes the language is grammatically incorrect or doesn't fit the type of publication it claims to be [2-9, pp. 42-46]. In addition, the message is very much designed to reach the reader emotionally.[68] More information can be drawn from the formatting. To check the layout of the

---

[65] https://www.geo.de/geolino/magazine/22568-rtkl-falschnachrichten-im-internet-so-erkennt-ihr-fake-news (last accessed 21.12.2021)
[66] https://www.bbc.com/news/blogs-trending-41084578 (last accessed 21.12.2021)
[67] https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/ (last accessed 20.10.2021)
[68] Ibd.

page, it is necessary to check how the layout of the page is organized. The overall layout as well as the fonts, graphic elements and multimedia content should be coherent [2-9, pp. 42-46].

2.2.2. Consider the source

When a source is opened, it is worth taking a look at the page. Information can already be obtained from the URL (Uniform Resource Locator), especially from the TLD (top-level domain) extensions. The URL can be seen in the navigation bar of the browser. It forms the entire link of the particular position on a website. The URL has a uniform structure. First of all, it consists of a scheme, e.g., HTTP (Hypertext Transfer Protocol) or HTTPS (HyperText Transfer Protocol Secure). This is the protocol for transferring data. As the name suggests, HTTPS is the most secure version. The next level of the URL is the third-level domain, e.g., "www.". Third-level domain names are not mandatory unless the user has a special requirement. Usually, only two levels are required. However, using third-level domain names can increase the clarity of domain names and make them more intuitive. In contrast, the second-level domain (SLD) is a mandatory part of the URL and shows the name of the website, e.g., "hs-ludwigsburg". The TLD is the conclusion of the URL. The most commonly used TLD is ".com". However, it can also contain geographical information such as ".de" for Germany [2-10, p. 6].

When investigating a website, it should be noted that fake URLs are often very similar to existing known URLs. Therefore, a close look should be taken at the URL. The SLDs should be popular on the one hand, but also trustworthy. Often a sign of Fake News is when the SLD consists of a large number of digits and hyphens. The TLDs should also be known. Established TLDs like ".com" or ".org" look more trustworthy.[69] [2-10, p. 12]

2.2.3. Author / Imprint

The next step is to check if there is an author or if there is an imprint. Here it is possible to check whether the website is private, institutional or governmental. In this context, it is crucial whether it is also an official account. In addition, information about the author should be available on the homepage. If it is a social media site, the profile can also be examined in more detail. It should be a trustworthy profile, possibly with a picture. In addition, a profile shows the interests of an author and whether he has already posted other articles on this topic. Sometimes there is a self-description of the author. This profile information is at least an indication of how trustworthy the author appears. In addition, other publications by the author on the Internet can be searched, as well as witnesses mentioned in the original article. For a reliable source, the authors should also be generally known on the Internet [2-9, pp. 42-46].

If no author or imprint is found, this probably indicates that the author should be disguised. In Germany, for example, an imprint obligation exists.[70]

2.2.4. Comparison with other sources

To get an overall picture of the situation, it helps to research other sources and compare facts. Particular attention should be paid to the context and time period in which the information and images appear.[71] For

---

[69] https://www.sixclicks.de/blog/domain-endungen-auswahl#Wie%20ist%20eine%20Domain%20aufgebaut (last accessed 09.11.2021)
[70] https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/ (last accessed 20.10.2021)
[71] ibd. (last accessed 20.10.2021)

the keywords such as names of people, places, companies, or products involved, it is necessary to check whether they are related to the real event. When the keywords are entered into a search engine, they should bring up the same event. The source may not be credible as well if the news seems too outrageous. The keywords should be found in other news from credible sources as well. If the message was originally written in another language, the original article should be consulted. Translation errors can also cause disinformation. The content can be checked with the help of own language skills or translation programs, of course with all their limitations. Sometimes a fake message can be unmasked or fact-checking information can be found by adding the term "fake" or "hoax" to the keyword. If one is still not satisfied, an expert such as university lecturers or journalists can be consulted [2-9, pp. 42-46].

2.2.5. Origin of a message

If a message is spread via social media, the originating message should be searched. This works with the help of search engines by entering parts of the message in the search engine field. In this way, contradictory statements can be compared and an overall picture of the situation can be obtained.[72]

2.2.6. Plausible and actual information

In all cases, it is important to weigh for oneself whether the info presented makes sense and could be plausible. Because sometimes it already helps to switch on one's mind and to think one step further to expose possible untruths. During the plausibility check, it also helps to note whether the text, image, video or audio file has a creation date and whether it is up-to-date and plausible.[73] Bear in mind that such a date can be manipulated. When examining the date, it should be noted when exactly an event took place and whether this is correctly stated in the article. A specific date should also be present. Furthermore, it should be checked whether the chronological order of reported events is correct. If a location is specified, it can be examined whether the location of the event is correct [2-9, pp. 42-46].

If studies are cited, the original study can be checked to see if the information given is correct.[74] By inspecting the links provided, it is possible to determine whether the author refers to the original source [2-9, pp. 42-4]).

2.2.7. Images, videos and audio files

Finally, information can be retrieved from photos, videos and other visual cues (including statistics and data) in news items. In the case of videos, images or other multimedia content, it must be questioned whether the visual element is reliable. Particular attention should be paid to signs of manipulation, such as filters, retouching or the like. The image could be a fake. It is necessary to pay attention to whether the medium matches the previous information. e. g. whether the date and time match the event. If there is a credit for the visual element, the authenticity can be investigated by checking the source [2-9, pp. 42-46]. Meanwhile, images, videos and audio files look deceptively real, which is why they are very difficult to identify as fake. In case of mistrust, screenshots can be entered into Google Image Search, for example. The YouTube Dataviewer[75] can also show the exact upload time of a YouTube video and preview images.[76] When using data and charts, it can be questioned whether the numbers and statistics

---

[72] ibd. (last accessed 20.10.2021)
[73] ibd. (last accessed 20.10.2021)
[74] ibd. (last accessed 20.10.2021)
[75] https://citizenevidence.amnestyusa.org/ (last accessed 20.01.2022)
[76] https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/ (last accessed 20.10.2021)

are used plausibly. The figures could be manipulated. Therefore, it is necessary to check whether similar numbers for the same topic can be found elsewhere [2-9, pp. 42-46].

These are the most crucial issues to recognize fake news. But not everything false is also fake. Information can be changed intentionally (disinformation) or unintentionally (misinformation). The spreading of misinformation is therefore not fake news, but may simply be a mistake, bias or some other form of incorrect reporting. As already described in the previous chapter, a distinction must be made between a fact-based report and a *personal opinion*. In the case of an opinion, there is greater freedom due to the freedom of expression. It is therefore important to note whether the content of a medium is a personal opinion or a fact-based report. Some texts may be meant in a humorous sense, e.g., as a joke or satire. *Jokes* are short stories or exposition with a surprising twist or punch line designed to make the reader laugh. Jokes are usually easier to recognize than satire. *Satire* is more serious humor with a type of writing in which circumstances or problems are addressed in an over-exaggerated, ridiculous form. It often works with exaggerations or understatements, with ambiguities or irony. In satires, people's faults and weaknesses are pointed out, often indirectly criticizing the human condition, but mockingly and humorously. Thus, an altered piece of information may be legitimate. In this case, it is particularly worthwhile to look at the source where this message is published, because it may already be a satire page.[77] Furthermore, there are always cases of *incorrect reporting*, which also belong to the term misinformation. This happens when serious sources include manipulated content in their reporting. For example, in the case of the alleged attack in Kongsberg (Norway), a fictitious perpetrator's name was published in the media. The name was originally spread on social media by so-called trolls who deliberately wanted to confuse.[78]

For a brief overview of how to check for fake news, the International Federation of Library Associations an Institution (IFLA) has created an overview that can be viewed in the figure below:
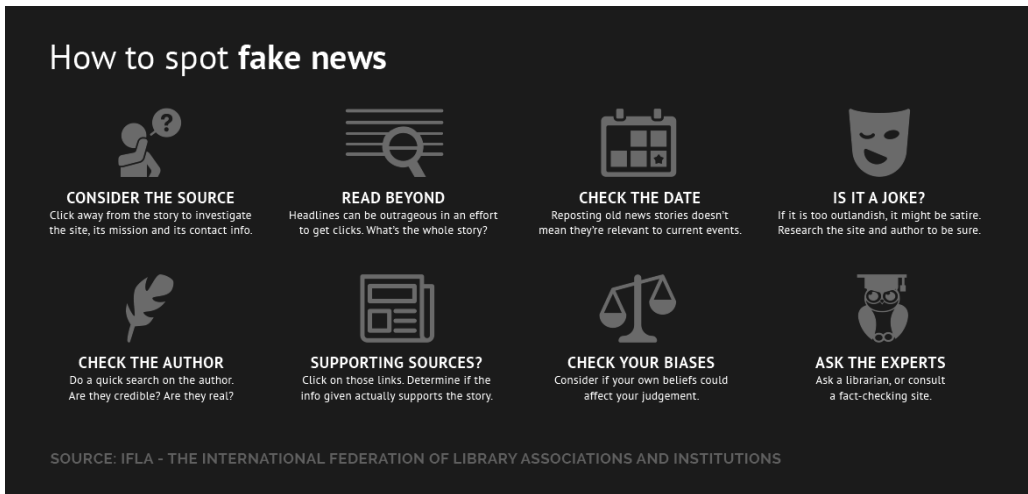
Figure 9: How to spot fake news[79]

[77] https://www.studienkreis.de/deutsch/satire-definition-merkmale/ (last accessed 17.11.2021)
[78] https://www.tagesschau.de/faktenfinder/kongsberg-medien-101.html (last accessed 18.11.2021)
[79] https://www.coe.int/en/web/human-rights-channel/fake-news (last accessed 13.10.2021)

It is important to recognize fake news. Wrong information can influence people's attitudes and actions. The next chapter, therefore, discusses the areas in which this influence can have an impact.

2.3. How can hate speech be identified? Identify the conditions conducive to the use of hate speech

Hate speech can take the form of written or spoken words, or other forms such as pictures, signs, symbols, paintings, music, plays or videos. It also embraces the use of particular conduct, such as gestures, to communicate an idea, message or opinion (cf. [2-11], p. 2). To determine whether a statement is hate speech, it is necessary to define what is meant by hate speech. Aiming to create a consistent understanding of hate speech in this book, a definition is created in Chapter 1. This delimits hate speech as follows:

"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, stigmatisation or threat concerning such person or group of persons, and the justification of any of the foregoing on the grounds of 'race', colour, 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, trivialisation, 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."

Put simply, hate speech involves attacking people in a discriminatory way. It must emerge that discriminatory words were used pejoratively in reference to a population group. Furthermore, hate speech has different characteristics. Laaksonen et al. defined five categories in their study. They define hate speech as messages that (1) incite violent action, (2) calls for discrimination or the promotion of discrimination; (3) attempts to degrade human dignity based on characteristics; (4) involves a threat of violence or the promotion of violent action; (5) or is accompanied by contempt, solicitation, name-calling, or slandering [2-12, p. 7].

However, it is very difficult to detect hate speech in everyday life. Only in the case of swearwords are a clear appearance of hate speech. The discriminatory must be pejoratively used in reference to a population group. But population groups can also be disparaged or denigrated without any form of expression. These manifestations of hate speech are difficult to recognize because they usually appear harmless at first. Restricting it to individual words would therefore not be helpful. Language only becomes hate speech in the context in which it is used. Especially in social media, the manifestation of hate speech is diverse and difficult to detect. Forms such as jokes, satire, or similar appear here, which are not decoded as such without context. Many people do not perform a context analysis and cannot recognize that the post is meant as a joke [2-13, pp. 338-340].

One way of approaching a medium to identify hate speech is presented below. A study by Patton et al. (2020) [2-13] is used for this purpose. They use the Contextual Analysis of social media (CASM) to detect hate speech by specializing in gang violence. The CASM works, in addition to the natural language processing tools, with the differences in geographic, cultural, age-related variance of social media use and communication. To determine the context of a post they use the following steps:

1. Baseline Interpretation

In the first step, a baseline Interpretation must be done. The message can be viewed soberly, just the context is disregarded and only the cover is interpreted. The text, emoji, hashtags, memes, images,

and videos can be used to get a first impression of whether it could be a hate message. Since these interpretations do not yet take the context into account, initial interpretations can still be influenced by prejudices, for example.

2. Examination of all biographical and offline information

To now take a look at the actual message of the message, all biographical and offline information is now to be used. First, the original social media post can be searched for specific mentions of names, communities, groups, schools, streets, local institutions or events. If the message can be assigned to a specific group, it is necessary to look for characteristics relevant to that group, such as words, phrases, emojis and other features. Contextual or cultural features can be located with the help of web-based resources. It should not be forgotten that hate speech can also be a matter of mistranslation or that other cultures have different ways of communicating. Second, the author who wrote the message is considered in more detail. All biographical information can be used for this purpose, e.g., name, date of birth, neighborhood, city. Photos can also be used to gather information on location, gang affiliation, peer network and environment. There may already be other postings where a pattern can already be identified. In this case, it can be compared whether the post matches the original postings. The final step is to take a closer look at the people tagged (@) in the post, like, share, or comment on the post. It can be asked; which relationship the persons have to the author or why these are linked. For the people who reply or comment on the post, there is also the question of the connection to the author and the reason why they are commenting on the post. Perhaps they are attracted to certain content. Possibly an intention can be discerned, namely whether they are trying to escalate or de-escalate the post [2-13, pp. 339-340].

3. Interpretation & Contextual Analysis Assessment

After the extensive context analysis, the original perception can now be reflected. Presumably, some perceptions could be downplayed or exacerbated [2-13, p. 340]. Similar to hate speech, there is a certain framework where hate speech is legitimate. In *satires*, for example, certain topics are presented with a high degree of exaggeration or ironically. This can make this particular type of humor seem like hate speech.[80] But the *time reference* is just as decisive. Over time, various terms have now been portrayed as discriminatory swear words. If older texts are used as sources, they appear as discriminatory with the time reference to today. An example of this is the work of Albert Schweitzer. From 1912 till his death in 1965, Schweitzer worked as a physician in Central Africa, founded the Hôpital Albert Schweitzer in Lambaréné (Gabon) in 1913 and was even awarded the Nobel Peace Prize for his work. Today, however, Albert Schweitzer is, by some people, judged critically because of his paternalistic attitude towards the Africans.[81]

2.4. Effect on political discussion, democracy, economy and society

After now knowing how to identify fake news and hate speech, it is essential to understand why such identification is important.

The role of the press in a democratic society is a vital one. The European Court of Human Rights has repeatedly underlined that the press and other media have a special role in a democratic society as the purveyor of information and public watchdog (cf. [2-14], p.6). Disinformation often highlights

---

[80] https://www.studienkreis.de/deutsch/satire-definition-merkmale/ (last accessed 17.11.2021)
[81] https://www.srf.ch/news/panorama/kritische-stimmen-zum-albert-schweitzer-jubilaeum, 24.03.2013 (last accessed 20.01.2022)

differences and divisions, whether they be between supporters of different political parties, nationalities, races, ethnicities, religious groups, socio-economic classes or castes ([2-15], p. 41). Hate speech can reflect or promote the unjustified assumption that the user is in some way superior to a person or a group of persons that is or are targeted by it. This assumption might lead to certain behaviors and thoughts, in the worst case, it might even end with physical attacks. Those various effects of fake news and hate speech will be discussed in the following paragraphs and shall underline the importance of fighting them.

2.4.1. Effects in general

Before going into deep about what effects fake news and hate speech have on several sectors, the term "effect" must be defined.

In this context "effect" shall be understood as an aimed or unaimed impact on something or someone. Sometimes this impact can't be known right from the get-go of an action. Most of the time people do not think about the outcome of their actions and how they might influence others (how easy it is to influence people and why will be discussed in chapter 2.5).

Coming back to fake news and hate speech, there are various effects and therefore impacts on other people and whole sectors.

Those impacts can be divided into three main groups:

- impacts on supporters/followers

  o fake news and hate speech might convince them of false facts

  o fake news and hate speech might activate them and encourage them to take action, ranging from postings to the use of violence

- impacts on opponents

  o fake news and hate speech might demotivate them in sticking to their opinions, continuing their role as a politician, starting a career as a politician

  o fake news and hate speech might disturb their actions and prevent them from doing what is necessary

- impacts on impartial individuals

  o fake news and hate speech might make them question their opinions and truthful news

  o fake news and hate speech might make them share the information with other people, even if they don't believe it, hence they spread it further

Although fake news and hate speech surely are influencing other sectors too, the following chapters will deal with the effects on political discussion, democracy, economy and society. These effects will be analyzed, based on the three main groups that have been introduced in this chapter (supporters/followers, opponents and impartial individuals).

2.4.2. Effects on political discussion

Nowadays, the world is connected more than ever and slowly but steadily becoming a global village. There is a lot more communication between politicians of other countries but also between politicians in the same country. Without a doubt, the internet and social media made it a lot easier to exchange thoughts, opinions and other information.

Over time, the internet developed and became one of the main sources for people to inform themselves about what is going on in the world and therefore also became a powerful instrument that might cause huge damage when being abused. Fake news and hate speech are perfect examples of such abuse.

Because politicians and their debates on upcoming law and other rules are often very present in modern media (such as television, radio and of course the internet) and ruled as big influencers on society, they often are pulled into the spotlight and become the target of such fake news and hate speech attacks. Political discussions can cause society to split up into two sides and influence the interaction between people in real life. This could be observed during the US Presidential election campaign 2020 and the discussions between Donald Trump and Joe Biden. It appeared that America was divided into Republicans and Democrats, fighting each other and standing up for their vision of how America should continue.[82]

When people are in such a dispute over something, fake news and hate speech can turn into a real weapon. Disinformation campaigns are the policy of "promoting lies, half-truths, and conspiracy theories in the media" [2-16, p.7]. Furthermore, a disinformation campaign can be a non-military measure for achieving political goals. The Russian Minister of Defense for example describes information as "another type of armed forces" [2-15, p.34]. Especially Russia is renowned for internet trolls, who "attack critical articles about Putin or Russian politics in European and U.S. online media, disseminate fake news […] and distort the representation of events on heavily funded Russian export media" [2-16, p. 7]. This demonstrates how fake news are used to influence peoples' minds and opinions on certain topics and by that achieve an advantage in political discussions.

The usage of fake news and hate speech, in particular, their spreading over social media and other news pages, can lead to severe consequences. It might lead to some politicians resigning their political function and status or in a worst-case scenario even to physical violence.

One example for that is the 'Querdenker' movement in Germany, where people are coming together, believing that COVID-19 is all a political setup and used by politicians to make the population bend its' will. Such movements can be seen all over the world right now. People do not accept the measures that are introduced by governments.

One who educates himself/herself by reading truthful articles, researching current hospital figures or talking to diseased people, can easily find out that COVID-19 is definitely not fake, and thousands of people are fighting for their lives daily. On the other hand, there are a lot of theories around COVID-19 and also a lack of appreciation linked to the rules/laws and decisions made during COVID-19.

---

[82] https://newsroom.ucla.edu/magazine/2020-election-trump-biden-divided-america (last accessed 11.12.2021)

Figure 10: The symbol of the 'Querdenken' movement in Germany[83]

2.4.2.1. Impacts on supporters/followers

The three main groups that have been introduced before, can also be transferred to the COVID-19 discussion. Some people are supporting and following those, who spread fake news about COVID-19 and believe that it's just a political setup to restrict their rights. As the number of supporters and followers grew, so did the commitment to fight for their freedom, hence they have been activated. Activated to go on demonstrations against the political measures and also physically defend themselves. This defense is directed against policemen during demonstrations,[84] but also towards individual politicians. The main targets of these physical violent acts during the pandemic are health ministers, virologists and others who are involved in the COVID-19-debate. An example of how far those violent acts can go is a recent parade carrying lighted torches in front of the home of Saxony's health minister.[85] Unfortunately, during the last two years, the number of online death threats against politicians who support pandemic restrictions has increased and put them under enormous emotional and physical.[86]

Figure 11: Example of hate comments against the German epidemiologist Karl Lauterbach on Twitter[87]

---

[83] https://querdenken-711.de/ (last accessed 11.12.2021)

[84] https://www.tagesschau.de/ausland/corona-protest-bruessel-103.html (last accessed 12.12.2021)

[85] https://www.zdf.de/nachrichten/politik/corona-protest-gewalt-verfassungsschutz-warnt-100.html (last accessed 12.12.2021)

[86] https://www.dw.com/en/covid-german-politicians-scientists-face-threats-online/a-56589911 (last accessed 12.12.2021)

[87] https://www.dw.com/en/covid-german-politicians-scientists-face-threats-online/a-56589911 (last accessed 12.12.2021)

Overall, the criminal offenses against public officials are increasing year by year. So do violent offenses. In Germany for example, 1.674 criminal offenses against public officials were registered in 2019, 89 of them were declared as violent offenses. When comparing these numbers to the incidents in 2018 (which were 43), the number of violence offenses more than doubled.[88] Although this might be shocking, there are expected to be a lot more cases that have not been reported, hence the real numbers could be likely higher.

2.4.2.2. Impacts on opponents

In a German study on violence against local politicians by KOMMUNAL (a magazine on local politics) and the opinion research institute Forsa, 2.494 mayors in Germany were asked about their experiences.[89] One of the main realizations is that violence against politicians is no longer only happening in bigger local authority districts but also in small villages. Most of the violent acts are happening at public events and working offices but violent acts are also starting to affect private actions. The affected persons are complaining about being insulted, threatened, and even physically attacked.

Figure 12: Overview of the results of the German study on violence against local politicians[90]

How does that affect politicians and their will to continue with or even start a political career? This question deals with the second main group, the opponents, and what impact fake news and hate speech have on them.

In this context, the study points out that the will to continue a political career or even start one decreases. Nobody wants to deal with violent acts, especially when they are carried out of the career-

---

[88] https://www.bpb.de/apuz/im-dienst-der-gesellschaft-2021/329322/gewalt-gegen-amtstraeger (last accessed 26.10.2021)

[89] https://kommunal.de/kommunalpolitiker-umfrage-2020 (last accessed 26.10.2021)

[90] https://kommunal.de/kommunalpolitiker-umfrage-2020 (last accessed 12.12.2021, own translation)

life into the private-life. This result proves, that fake news and especially hate speech demotivate politicians in their actions and also in their political careers. But not only that, the consequences of fake news and hate speech are going a lot further and affect the private life of politicians and their families so that they have to live in fear for their own lives and the life of their loved ones.

2.4.2.3. Impacts on impartial individuals

Finding out about the impact fake news and hate speech have on impartial individuals is not easy because they're often acting in the background, without anyone noticing it. Compared to supporters of fake news and hate speech, impartial individuals are not as outstanding and present in the media. One impact that all the fake news might have on impartial individuals is that they make them question their point of view and their opinions so that there is a potential threat, they might start to believe them someday and become followers and supporters.

On the other hand, impartial individuals are discussing a lot, because they are standing in between two opinions and try to figure out arguments for and against each side. Therefore, they might share fake news to discuss them with family and friends and that also causes them to spread further.
 Nevertheless, the quick way of communication by messages and sharing links and information online within seconds makes it possible for fake news and hate speech to be spread all over the country and even beyond. Because technology and messengers are constantly being used in our daily lives and becoming more important, even essential in ways of communication, we must expect that the numbers of criminal offenses against public officials will increase further without action against that.

2.4.3. Effects on democracy

As the introductory chapter already pointed out, freedom of expression is an enjoyment required by democracy.[91] Thinking of the small line between freedom of expression and hate speech, it is quite obvious that hate speech and also fake news affect democracy and the standards within. When social media was first implemented in our daily lives, everybody, including politicians, thought that social media would help to make democratic information available and help voters to make more informed choices during an election [2-17, p. 12]. What they did not have in mind is that social media can be misused and therefore affect opinion-building in a negative or simply untrustworthy manner.

One big problem for democracy was mentioned by the US political analyst Charlie Cook: "the wall between real journalism and fake journalism is becoming blurred and sometimes invisible. When people doubt the credibility of legitimate journalism, people are robbed of the facts that underlie our entire democratic process. Elections depend on citizens making informed decisions, but that's impossible if raw sewage is polluting their news feed".[92]

This represents a challenge for democracy, and in particular for the electoral processes throughout Council of Europe member States, affecting the right of freedom of expression, including the right to receive information, and the right to free elections.[93] "While there is no doubt that in a democracy all ideas, even though shocking or disturbing, should in principle be protected […], it is equally true that not all ideas deserve to be circulated" [2-14, p. 9].

---

[91] https://www.legislationline.org/documents/id/8226 (last accessed 20.10.2021)
[92] https://www.coe.int/en/web/human-rights-channel/fake-news (last accessed 13.10.2021)
[93] http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-EN.asp?fileid=28598&lang=en (last accessed 20.10.2021)

While there is fake news, on one hand, spreading untrue information about a topic, there also is news on the other hand that isn't untrue but is promoted to be fake news by politicians. One good example of that is the election campaign and usage of social media by Donald Trump. In his daily tweets, he is purposefully using wrong information to influence another person (Barack Obama and his place of birth), social group (lies about illegal immigrants), country (Mexico and its' population) or organization (World Health Organization) in a negative way [2-18, p. 13].

Figure 13: Example of a tweet made by Donald Trump[94]

Whenever a news magazine posted an article that wasn't supporting his point of view or was even criticizing his way of leading, he would instantly call it "fake news media" and accuse them of trying to tear the country apart [2-18, p. 13].

Professor Tarlach McGonagle, a senior researcher at the Netherlands Network for Human Rights Research, is describing Trump's behavior as a witch hunt where accusations are made in the expectation to inflict public mistrust for media and daily press. In his opinion, this contributes to aggression and hostility towards journalism and media in general [2-18, p. 14].

The main problem is that the opinion of other people in political discussions is often portrayed as wrong, rather than different and accused of destabilizing democracy. As human beings, we are drawn to believe people with the same views and opinions on certain topics. This effect is reinforced when we are hearing a lot about those people in the media.

All societies are experiencing an increasing form of influence by journalism and media. The biggest influence is given by policy and the economy (power and money). Those, who have an excessive amount of political or economic power, are using the media to flaunt it and let it work in their favor [2-18, p. 15].

---

[94] https://www.stiftung-nv.de/de/publikation/kurzanalyse-zu-trumps-crime-tweet-deutschland-viel-aufmerksamkeit-wenig-unterstuetzung (last accessed 12.12.2021)

2.4.3.1. Impacts on supporters/followers

Transferring this cognizance to the impact fake news and hate speech have on supporters and followers, it appears that they can be highly influenced in their opinion-building on democratic votes and views. At the same time, fake news and hate speech might lead them to mistrust the rules and values a democracy is built on. Furthermore, the transparency problem with politics and laws over all is playing a big role in this case. If people could easily find truthful information about current democratic topics and easily understandable articles from the officials directly, not as many people would come across fake news and hate speech or at least would not trust them within seconds.

In contrary to the activation of supporters and followers in the context of political discussion, the activation in the context of democracy is not directed at certain politicians, but at the whole democratic system. Therefore, people are activated to question the system in terms of democratic bases and ask themselves if the politicians are still acting within the boundaries of their legitimation.

Figure 14: Symbol for a democratic election[95]

2.4.3.2. Impacts on opponents

In a democracy, citizens and politicians to a certain degree are dependent on agreeing, on what is real and what is not.[96] Fake news and hate speech can destroy this consensus by offering an alternative reality. Consequently, information-based decision-making is a lot harder.

While supporters and followers of fake news and hate speech are questioning democracy, it is becoming a lot harder for politicians and also democrats, in general, to convince them otherwise. Multiple scandals, for example, the face mask scandal in Germany during the pandemic, where politicians of the CDU and CSU received commission payments for conveying purchase contracts for face masks, have shattered their credibility.[97]

Because fake news and hate speech are spread extremely fast over social media and the internet in general, truthful, and most times less interesting news, are moving into the background and overpowered by disinformation.

---

[95] https://www.osce.org/odihr/463626 (last accessed 12.12.2021)
[96] https://www.dw.com/de/fake-news-sind-eine-bedrohung-f%C3%BCr-die-demokratie/a-37033453 (last accessed 13.12.2021)
[97] https://www.zeit.de/politik/deutschland/2021-03/maskenskandal-cdu-affaere-csu-jens-spahn-georg-nuesslein?
utm_referrer=https%3A%2F%2Fwww.google.com%2F (last accessed 12.12.2021)

As a consequence, Democrats and politicians have to work even harder to regain the trust of all skeptics. How they could do that will be analyzed in chapter 5.

2.4.3.3. Impacts on impartial individuals

Because some true news is ruled fake news by politicians, individuals might get into believing the verdict of the respective politicians and hence believe that these are fake news. This is like the reputation effect under limited knowledge in economics [2-19, pp. 488-500]. This again might lead them to stick to one side or another. If they conclude that actual true news is fake news, the supporters and followers are growing in number. If they remain impartial, it sure will become harder for them to find the difference between fake news and facts, because those lines are blurred out more and more.

2.4.4. Effects on economy

Social media is a big business, orientated on profit and growth rather than sharing truthful information. According to the New York-based Data & Society Research Institute "the financial dependence on Facebook for content distribution has […] weakened the reach of solid journalism. The role of journalism now is to "give people what they want" or "what matters to them" and is embedded in the same logic that drives Facebook's algorithmic personalization and ad-targeting products".[98]

Figure 15: Illustration of multiple social media platforms[99]

An interesting aspect in the context of the economy is "Disinformation as a Service" (DaaS). The difference between disinformation and misinformation was already pointed out in the introductory chapter, hence these terms will not be defined again here.

Nowadays the internet is a powerful and sometimes underestimated instrument to influence people and companies. One way to do so is by using the help of DaaS.[100] As the term already indicates,

---

[98] https://www.coe.int/en/web/human-rights-channel/fake-news (last accessed 13.10.2021)
[99] https://www.som-onlinemarketing.com/2018/05/22/so-knackst-du-den-code-fuer-die-social-media-algorithmen/ (last accessed 13.12.2021)
[100] https://prevency.com/desinformationskampagnen-als-gefahr-fuer-die-wirtschaft/ (last accessed 18.11.2021)

disinformation campaigns are offered as a service to spread fake and harmful news against a certain person, company, etc. These services are offered in the Darknet and pretty much everybody can take advantage of them by spending some money and defining whose reputation shall be damaged. Concrete examples for such DaaS-Campaigns are fake comments on social media, fake recommendations on company websites and their products, or spreading rumors all over the internet and in the daily press.[101]

Before those services were provided, it was quite costly to start a disinformation campaign. They were mainly implemented by criminals and unscrupulous governments underground.[102] Over the years, a lot of DaaS suppliers have emerged and started advertising their "products" routinely to the private sector. Hence, it has become a serious market with lots of different offers and interested customers. What makes this dangerous is that on one hand these campaigns are inexpensive to create and distribute at scale,[103] but on the other hand can cause major damage to those people who are targeted.

The reason why it is inexpensive lies in human nature itself. A 2019 MIT study examined to which extent fake news is spread faster than the truth. The result is that "falsehoods are 70 percent more likely to be retweeted […] than the truth" and reach the first 1.500 people six times faster.[104] Another interesting result of the study is the fact that this phenomenon is more pronounced for false political news rather than false news about science, terrorism, natural disasters or financial information. Somehow people are more interested in stories and scandals in the political environment than in any other environment. Finding out about something that isn't compatible with our expectations or the way we think things should be, makes us talk about it and also share it with friends and family, especially online. As a human, we all have certain expectations and behavioral standards that we have learned and accepted throughout our lives. Whenever news is reporting anything contrary, it instantly makes us think about it, talks about it, and leads to a certain view or opinion about the person that has been involved. As a consequence, it is pretty easy and inexpensive for DaaS suppliers to satisfy their customers' wishes because the news will spread almost automatically after they have been implemented.

While fake news and hate speech are boosting one's finances, they destroy another one. According to current studies, fake news costs the world economy more than 78 billion US dollars annually.[105] The reason for this loss of money is that fake news is more and more used to harm companies and their daily operations. By spreading fake news about one company over the internet, a lot of its customers might read them and question themselves, whether they still want to support the company or not. Not only are these companies losing their customers, but it also makes it a lot harder for them to gain new clients because their reputation has been damaged by the fake news spread. The consequence is a deep cut in their income. Adding upon this, they might have to put in quite a lot of money to rebuild their reputation and gain back some customers, which of course costs money and a well-thought strategy. If this strategy does not work out as planned and the reputation couldn't be rebuilt, the companies might end up in financial ruin.

2.4.5. Effects on society

The effects of fake news and hate speech on society are closely connected with those on democracy and political discussion. People who want a radical change use the internet to spread hate speech

---

[101] Ibid.
[102] Disinformation attacks in corporate sector: PwC (last accessed 24.11.2021)
[103] ibd.
[104] Study: False news spreads faster than the truth | MIT Sloan (last accessed 24.11.2021)
[105] https://prevency.com/desinformationskampagnen-als-gefahr-fuer-die-wirtschaft/ (last accessed 18.11.2021)

and fake news and by that try to convince others from their views and theories.[106] This is in principle what other political activists, parties, etc. also do – the difference being that they do not facilitate hate speech and fake news.

Also political and religious tension is spread more rapidly. Haters are following their goal to exclude specific groups from society and silence their voices. It is kind of contrary to what we expected from the internet: a great basis for communication all over the world and the forming of a global community based on the values of the UDHR, the US constitution and the European Convention of Human Rights. Instead, the internet is drifting in the opposite direction and separates us all into tiny groups of people who are sharing the same opinions within these groups and cutting themselves off of other people with different opinions.

This separation causes more problems that might not be visible at first sight. While one group is actively promoting hate speech about minorities, more and more people are noticing that and joining in. However, on the other side, those social minorities are experiencing severe effects and are excluded even more. Especially hate speech causes a lot of them to question their behavior and ask themselves whether there is something wrong with who they are or what they do. This results in an emotional burden and causes physical symptoms [2-20, p. 29]. Depression, tiredness and insecurity are examples of that. On top of that, a lot of the affected persons are scared that online attacks could be realized and end up in physical violence. As a consequence, they minimize their use of social media, delete their accounts or partly shut off their social life to protect themselves.

Fake news also has a high impact on society. A great example of that is the pizza-gate case. After voices got loud, that Donald Trump had bragged about sexually assaulting women, he pointed out that Bill Clinton had raped women and wanted to focus the attention on those accusations.[107] The leaked E-Mails of Hillary Clinton were said to prove that the rape shall have taken place repeatedly at the "Comet Ping Pong" Pizzeria in Washington D.C. According to the fake news, this Pizzeria was a meeting point of pedophiles, who could order "Pizza", if they wanted to be served a girl, a "Hot Dog", if they wanted to be served a boy and "Sauce" for an orgy.[108] There was no evidence that those meetings took place, but people were very interested in this scandalous story and therefore, the information was spread rapidly online all over America and even across the borders. The backfire of all the shocked citizens even went so far, that the owner of the Pizzeria, James Alefantis, received several death threats and the FBI had to take over this case.

As a consequence of the overwhelming fake news-flood on the pizza-gate case, something terrible happened on the 4th of December 2016. 28-year-old Edgar W. decided to travel from North Carolina to Washington D.C. with the intent of investigating the pedophile meetings by himself. He wanted to help the children that were being held hostage and free them from their suffering. When arriving at the Pizzeria, he brought a gun with him and was overwhelmed by his emotions. In a stroke of anger, he opened the fire and was shooting around with no specific target.[109]

---

[106] https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/hatespeech/folgen-fuer-den-gesellschaftlichen-zusammenhalt/ (last accessed 07.11.2021)
[107] https://www.faz.net/aktuell/feuilleton/debatten/wie-sich-in-amerika-die-herrschaft-der-luege-festigt-14565557.html (last accessed 08.11.2021)
[108] ibd. (last accessed 09.11.2021)
[109] https://www.faz.net/aktuell/gesellschaft/kriminalitaet/pizzagate-fall-mann-kriegt-4-jahre-haft-wegen-selbstjustiz-15073545.html (last accessed 09.11.2021).

Fortunately, nobody was hurt. After police arrived, Edgar W. was arrested and then sentenced to 4 years in prison.

Taking a closer look at this case demonstrates, how dangerous fake news can be and what it can lead to. Here, they were once virtual theories with no truthful evidence and ended up becoming reality for some citizens.

What is special in this scenario is, that Edgar W. didn't want to hurt anybody, he just wanted to free the kids and protect them from more misery. Normally, this would be a heroic act but instead of becoming a hero, this man ended up becoming a criminal and almost a murderer. He has to spend 4 years in prison, might lose contact with various people in his life and might not be able to fully recover (both, emotionally and socially) from everything after he is released from prison.

What took place on the 4th of December 2016 is a perfect example for showing that not only people with bad ideas and criminal thoughts are receptive to fake news, but also people who are willing to help others and make a change for the better. Therefore, it is important to understand how and why some people are more receptive to fake news than others, which will be discussed in the next chapter.

2.5. Why do people fall for fake news?

Why do they not recognize fake news when they see them? Unfortunately, part of the problem with fake news is that people fall for it even when confronted with fact checks. Psychologists and other social scientists are working hard to understand the mechanisms behind the human mindset. There are psychological aspects that explain the mechanisms behind it; because our brain is a very powerful organ that sometimes tries to take efficient shortcuts. These shortcuts are also called heuristics - they are mental strategies and rules of thumb that help us make decisions and judgments with limited knowledge and time. Heuristics and cognitive biases can be dangerous because they can lead us to have unrealistic expectations and make poor decisions. Cognitive biases are gaps in reasoning, remembering, or evaluating something that can lead to false conclusions. They are universal and everyone has them. These aspects contribute to the success of fake news: We tend to avoid cognitive dissonance and prefer information that fits our worldview and social identity. The world and its social and economic interrelationships are complex and therefore simple explanations and/or solutions are sometimes unconsciously preferred. This applies to fake news just as much as to serious news.[110]

> A cognitive bias is a subconscious error in thinking that leads you to misinterpret information from the world around you and affects the rationality and accuracy of decisions and judgments. Biases are unconscious and automatic processes designed to make decision-making quicker and more efficient. Cognitive biases can be caused by several different things, such as heuristics (mental shortcuts), social pressures, and emotions.[111]

This chapter is focused on how cognitive biases work. It attempts to address the question of what goes on in people's minds that makes us more susceptible to falling for fake news and believing misinformation even after it has been corrected? The roles of mass media, as well as social media platforms in the spread of fake news, are also discussed.

---

[110] https://www.cits.ucsb.edu/fake-news/why-we-fall (last accessed 24.11.2021).
[111] https://www.simplypsychology.org/cognitive-bias.html#definition (last accessed 09.11.2021).

2.5.1. Cognitive bias

Numerous cognitive biases influence our thinking in general, but also the way we think about others. They affect the decisions we make, as well as how we use information. The following figure shows the most common cognitive biases to be aware of. As mentioned earlier, everyone has these biases, and the first step to reducing their influence on them is to be aware of them.[112]
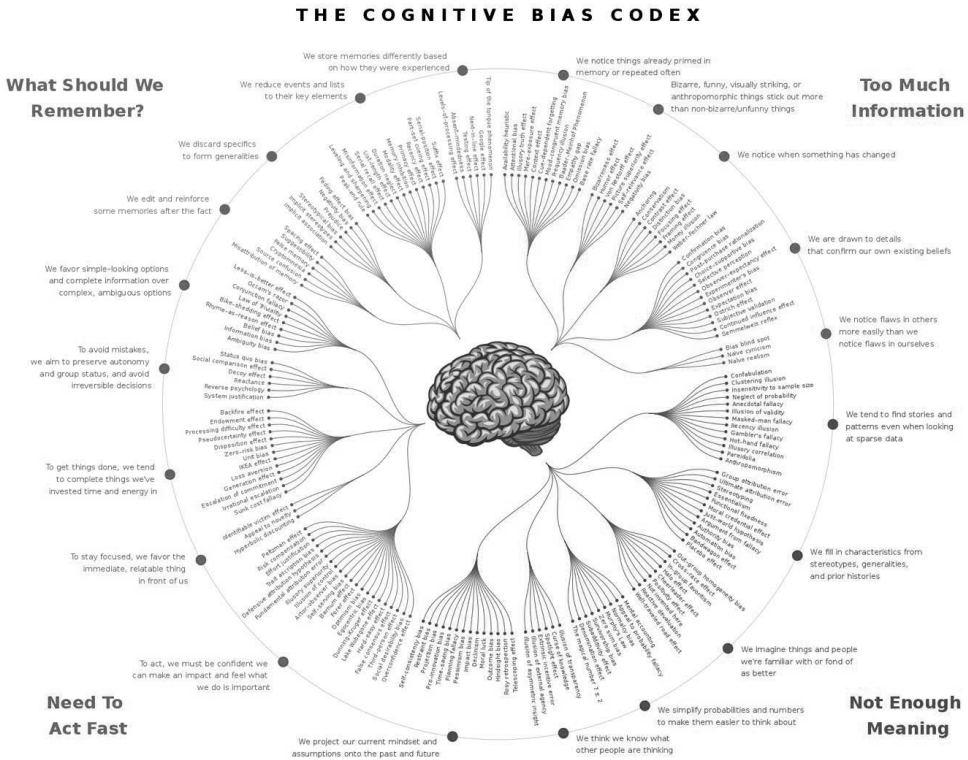
Figure 16: The Cognitive Bias Codex - 180+ biases.[113]

The following describes four types of cognitive biases that are particularly relevant in the context of fake news and its influence on society and politics.[114]

1. Reasoning: Reading and thinking is fundamental

An uncomfortable fact about social media is that people tend to form opinions about the news by paying attention only to the headline, without even reading the accompanying information and without evaluating it thoroughly afterward. A disturbing, if amusing, an example of this is a social experiment conducted by National Public Radio (NPR): "Why Doesn't America Read Anymore?"

---

[112] https://www.cits.ucsb.edu/fake-news/why-we-fall (last accessed 24.11.2021).
[113] https://commons.wikimedia.org/w/index.php?curid=69756809 (last accessed 20.01.2022).
[114] https://www.cits.ucsb.edu/fake-news/why-we-fall (last accessed 24.11.2021).

was the post that NPR jokingly placed on its Facebook page. When users followed the linked-up post, they were directed to the article on the NPR website, which explained that the post was a hoax. Nevertheless, many viewers did not read the clarifying article and responded immediately with comments on the Facebook page.[115]

Figure 17: NPR on Facebook "What has become of our brains?".[116]

Other news providers also report similar experiences. Researchers have studied this problem and examined 2.8 million news articles on Twitter that were shared or commented on by its users. According to the results, more than half of these users never clicked on the shared link and read the entire article. So, people often share or like headlines online without having engaged with the actual content [2-21, 2-22]. This behavior could be very harmful in terms of the spread and influence of fake news. Likewise, it leads to the rise of clickbait as the eye-catching headlines attract attention [2-23]. Thus, people may mistakenly assume that the misleading headlines are true when they do not read the content and do not further explore whether there is any doubt about the story or any other opinion. Sharing headlines without reading the content can also make it appear that they are gaining popularity, i.e., are trending [2-24]. This increases the likelihood that other people will continue to share these headlines without having read them, leading to a kind of socio-cognitive epidemic.

Researchers from the Massachusetts Institute of Technology (MIT), USA, and the University of Regina in Saskatchewan, Canada, studied the analytical reasoning ability of more than three thousand American participants. Their results showed a correlation between a higher score on the intelligence test and a better ability to distinguish fake news headlines from real news headlines. This was the

---

[115] https://www.npr.org/2014/04/01/297690717/why-doesnt-america-read-anymore?utm_medium=facebook &utm_source=npr&utm_campaign=nprnews&utm_content=04012014 (last accessed 24.11.2021).
[116] https://www.facebook.com/NPR/posts/10202059501509428?stream_ref=10 (last accessed 24.11.2021).

case even when the fake news matched the political preferences of the participants. According to the results of the authors, people were more likely to fall for fake news because of laziness of thinking than because of a conscious or unconscious desire to confirm their political preference [2-25]. "People [who] believed false headlines tended to be the people [who] didn't think carefully, regardless of whether those headlines aligned with their ideology," Rand said [2-26]. Confirming this conclusion, a replication study by Rand with colleagues showed that careful thinking had an effect not only on the detection of patently false headlines but also on nonpartisan headlines [2-27]. These results were confirmed in further research. Study participants were shown a series of news stories as they would appear in social media: as screenshots with the headline, source and first sentences of a news story. Some of these stories were false; others were true. Participants were first asked to quickly and visually judge whether the news stories were real or fake. Later, they were asked to judge these stories again, but this time they were asked to take more time and think carefully about the truthfulness of the news. The comparison between the fast and slow procedures showed that participants were better able to discern the truth content when they took more time to think, regardless of the political consistency of the news they evaluate [2-28].

In summary, the (in)ability to distinguish between truth and untruth is related to the ability to reflect on the content read (or heard). Accordingly, more reflective, analytically thinking individuals are less inclined to fall for fake news. Belief in fake news is likewise related to delusionality, dogmatism, religious fundamentalism [2-29], bullshit receptivity, and overclaiming [2-30] – all of which are likewise factors related to analytical thinking.

2. Bandwagon effect: Thousands of "likes" and "shares" cannot be wrong

The apparent popularity of a news story fuels another bias and promotes the acceptance of fake news. The so-called "bandwagon effect" is one of the most researched cognitive biases and explains the influence of popularity on our thinking, e.g. [2-31]. According to this, people tend to adopt a certain behavior, in this case sharing news, simply because others are already doing it. So the more people adopt a particular trend, the more likely it is that other people will also "get on the bandwagon" [2-32]. To this end, the brain uses mental shortcuts (heuristics) designed to facilitate rapid decision-making. It usually takes time to think through an idea or behavior before implementing it. Many people skip the process of individual evaluation by relying on the judgment of others around them whose opinions are perceived as trustworthy - and so the third-party opinion gains popularity.

The bandwagon effect leads us to be more influenced by what has been shared and liked, rather than focusing on what is actually in the content. Related to fake news, this effect in a sense frees us from the responsibility of having to verify the information shared. We simply assume that someone must have verified the information if it has already been shared many times [2-22]. To confirm this assumption, the researchers manually reviewed a sample of 50 articles shared on Twitter, drawn evenly at random from a corpus of articles. The result showed that only a minority of the shared articles contained verified claims:

Figure 18: Verification based on a sample of 50 articles.
Source: [2-22]

The bandwagon effect is a similar phenomenon to "herd mentality" or "groupthink". The cause may be the pressure to conform when it appears that a great many people agree with a message or an opinion [2-33]. Further, the human desire to be right and appear to be an expert may be a reason why many people claim and share information that they have not even read. We all want to be on the winning side and tend to look to other people in our social group to find out what is right and acceptable and adjust our (political) thinking accordingly [2-34]. Likewise, the need to be included can play a role in the acceptance and spread of fake news. Generally, people do not want to be seen as outsiders but want to be liked and go along with what people around them are doing to secure inclusion and social acceptance [2-35]. People want to feel connected to those around them, whether it's members of the same political party, political activists highlighting climate change, members of religious communities, etc.

The bandwagon effect can have several negative effects. For example:

- Researchers have found that polls can influence individuals when a particular candidate is ahead in them. People may tend to then change their vote to be on the winning side. [2-36, 2-37].

- Individuals have been influenced by the anti-vaccination movement and, as a result, were less likely to have their children routinely vaccinated. This movement has been linked to the recent measles outbreak [2-38]. This example can currently be applied to the COVID-19 vaccination campaign as well.

Social media platforms have an important part in the emergence of the bandwagon effect. In relation to social media, the problem of the influence of popularity on the acceptance of fake news is amplified by so-called "social bots". Bots are programs that perform tasks automatically. They can be harmless or even useful or amusing. For example, bots that automatically answer frequently asked questions (FAQ) in a chat are useful. However, bots on social media can also be used with malicious intent to manipulate users by spreading false information or making it appear that certain people, ideas, or products are more popular than they actually are. They can also troll or attack victims and commit many other types of abuse. In the broadest sense, a social bot is a social media account that is controlled, at least in part, by software. The content of posts and the timing of actions usually originate from humans or are closely monitored by humans. Most automated bots simply act (reply, post,

follow, etc.) based on triggers or scripted patterns, such as retweeting all messages from specific accounts or posting from predefined lists.[117]

3. Partisanship: The most common political mental shortcut

Our personality and worldview can also lead us into mental traps. Especially in the case of fake news, our political attitude has a significant part in it, and it is not easy to avoid these traps. Regardless of its orientation, political preferences can affect whether we believe or reject the news; the actual truthfulness does not matter. Research findings supported the thesis that most politically oriented fake news sites were consumed by conservatives. For example, surveys and internet traffic data from the 2016 U.S. presidential campaign showed that Donald Trump's supporters in particular were most likely to visit untrustworthy websites that were frequently shared via Facebook [2-39, pp. 12-18].
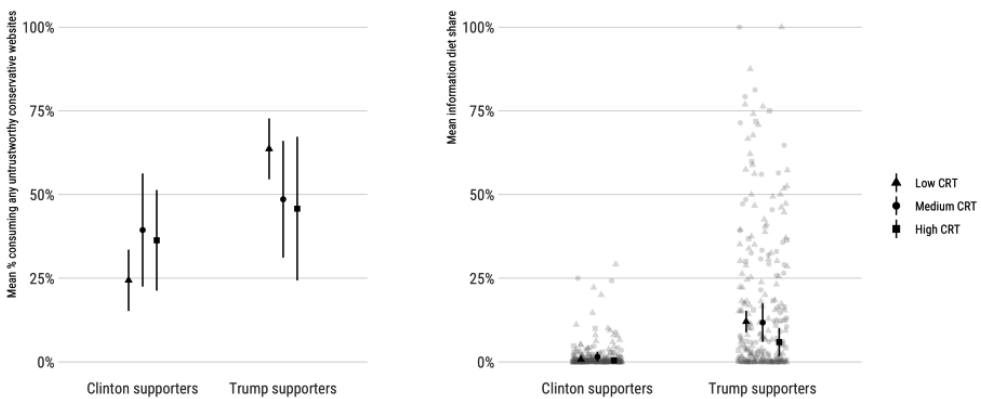
Figure 19: Consumption of untrustworthy conservative websites by CRT score and candidate preference.
Source: [2-39, p. 13]

The collected data also revealed that supporters of presidential candidate Hilary Clinton on average visited fact-checking websites more often and fake news websites less often. The opposite was observed for Trump supporters [2-39, pp. 12-18].

There is a widespread view that the inability to distinguish between true and false news is due to political motivations. When people are confronted with politically valued content, they display an "identity-protective perception." Thus, this results in people believing such content that is consistent with their partisan political identity. In contrast, they in turn give little credence to such content that is inconsistent with their political orientation [2-40, 2-41]. A related theory suggests that people place loyalty to their political identity above truth – and therefore fail to distinguish truth from falsehood and instead simply believe ideologically consistent information [2-42]. These explanations assume that a strong causal influence of political motivation on belief is the dominant factor explaining why people fall for fake news [2-43].

Individuals whose business concept is to make money by spreading fake news on the Internet stated that they tried to target both conservatives and liberals with false news, regardless of their political

---

[117] https://botometer.osome.iu.edu/faq (last accessed 29.11.2021).

preferences. However, people with liberal political views were less likely to click on these stories, which led such websites to stop disseminating pro-liberal fake news:

> "Well, this isn't just a Trump-supporter problem. This is a right-wing issue. Sarah Palin's famous blasting of the lamestream media is a kind of record and testament to the rise of these kinds of people. The post-fact era is what I would refer to it as. This isn't something that started with Trump. This is something that's been in the works for a while. His whole campaign was this thing of discrediting mainstream media sources, which is one of those dog whistles to his supporters. When we were coming up with headlines it's always kind of about the red meat. Trump got into the red meat. He knew who his base was. He knew how to feed them a constant diet of this red meat. We've tried to do similar things to liberals. It just has never worked, it never takes off. You'll get debunked within the first two comments and then the whole thing just kind of fizzles out." [2-44]

Perhaps the clearest evidence that Trump knew exactly how to "feed" his base is the violent storming of Capitol Hill in Washington, D.C., by an estimated 2,000 to 2,500 of his supporters on January 6th, 2021 – just minutes after Trump explicitly asked them to march there. These riots have so far been the culmination of Trump's unsuccessful claims of voter fraud in the November 2020 presidential election and repeated stoking of division in the United States. In a constant barrage of misinformation, Trump repeatedly claimed months before the election that he would only lose if the election was rigged [2-45].

Generally speaking, the political landscape seems to be increasingly populated by actors who spread demonstrably false claims. This problem is exacerbated by foreign efforts, such as Russian propaganda tactics, including on social media, that attempt to influence elections or influence misinformation regarding the COVID-19 pandemic [2-46, pp. 5-6].

4. Belief echoes: The tendency for misinformation to persist

What is worrisome about cognitive bias is that it can persist for a long time, preventing us from eliminating misinformation that has already been internalized. Unfortunately, it is not the case that belief in fake news can be quickly eliminated simply when that news has been disproved or corrected.

Scientific research has confirmed that our memory is not good at remembering what is real and what is fake once we receive the information. Professor Emily Thorson at Boston College has found that even when misinformation has been corrected, "belief echoes" remain. Her research suggests that these echoes can be generated by an automatic or deliberative process: belief echoes occur even when the misinformation is immediately corrected according to the "gold standard" of journalistic fact-checking [2-47].

The "United Kingdom European Union membership referendum" on June 23rd, 2016 – the so-called Brexit – can serve as a vivid example of how the tendency for misinformation to persist can influence political reality. One of the central messages of Brexit supporters in the run-up to the referendum was that Britain was paying 350 million GBP (just under 400 million Euro) each week to fund the budget of the European Union[118]. This claim was circulated by two key figures in the Brexit campaign – Boris Johnson and Nigel Farage – who stated that this money would be better spent on the British healthcare system (NHS National Health Service) after all. Before the referendum, various institutions examined

---

[118] http://www.voteleavetakecontrol.org/briefing_cost.html (last accessed 13.12.2021).

the claims and consistently reported that the amount stated was incorrect.[119] The day after the referendum in which Britons voted for Brexit, Nigel Farage even publicly admitted the "mistake".[120]

Even if some forms of fake news warning pointed people not to believe it, the absence of such warnings can have a greater impact than their presence. Scientists have addressed the question of what can be done to counter political misinformation: they studied how fake news warnings affect people's beliefs. As a result, people were less likely to believe such stories when they came with a warning; but conversely, when a warning was not present, subjects were more likely to believe the stories, whether they were fake or not. Thus, when people are confronted with a warning about misinformation, they are more likely to feel they do not need to be on alert and to rely on the warning. However, in the absence of a warning about misinformation, people are more inclined to believe the information to be credible, even though it is not. Taken together, these results challenge theories of motivated reasoning while highlighting a potential challenge to the politics of using warnings to fight misinformation – a worrying challenge because it is much easier to produce misinformation than to debunk it [2-48].

Twitter-related research has shown that users who posted misleading content often tagged it with a fact-check article [2-22]. The problem, however, was that even when people clicked on the fact check (which doesn't always happen – such as "bandwagon effect"), the facts cited were themselves fake, which is often exploited by clickbait creators to generate clicks and make money that way [2-23].

Doubts should also be raised about the use of warnings, and they should be chosen carefully. If they are formulated too drastically, they could cause a reaction. If they are formulated too weakly, they could be overlooked. Moreover, they could be forgotten over longer periods. Nevertheless, they cannot be dispensed with completely. In particular, warnings against further dissemination, i.e., sharing of false news, appear to be important, since personal recommendation from user to the user represents one of the central mechanisms in the spread of false news on the Internet [2-49].

2.5.2. The role of mass media

As Heinz Bonfadelli of the German Federal Agency for Civic Education (bpb – Bundeszentrale für politische Bildung) sums it up well, mass media such as the press, radio and television, as well as the Internet and social web, make an indispensable contribution to the functioning of democracy. Politicians in general and media professionals in particular, but also the public, assume this. Mass media are supposed to contribute to both the stability and the change of society [2-50].

According to sociologist Niklas Luhmann [2-51], media enable society to observe itself:
– Media as "windows to the world" select and provide relevant topics for the public.
– Media provide citizens with arguments for and against controversial issues.
– Media research the background knowledge necessary for decision-making, prepare it in a comprehensible way and make it widely available.

As a result of this media service, arguments on current issues are exchanged, discussed and critically scrutinized in public. By using the media, the population participates in current social issues and problems. This increases the level of knowledge of all. In addition, it is hoped that minorities such as migrants will also be integrated into society through the media. Through media coverage, social prejudices and perhaps even discrimination against minorities could weaken [2-50].

---

[119] e.g., http://infacts.org/quitting-eu-mean-less-money-nhs-not/, also https://www.bbc.com/news/uk-politics-eu-referendum-36110822 (last accessed 13.12.2021).
[120] https://www.itv.com/goodmorningbritain/articles/nigel-farage-labels-350m-nhs-promise-a-mistake (last accessed 13.12.2021).

Information as a "window on the world" → transparent public arena		
Politics	**Economy**	**Culture – Social**
Creating publicity	Consumer information	Orientation and life support
Articulation of opinions	Circulation of goods	Socialization: values & norms
Control and critique	Securing employment	Integration in society
Early warning function	Value creation, e.g., in the media industry	Education and cultural development
Participation & activation		Entertainment and relaxation

Figure 20: Overview: Functions of the media for the sectors of society.
Source: [2-50]

These expectations of the public mass media are ideal concepts, which are demanded as desirable achievements. In reality, however, they are always only partially realized, which is repeatedly expressed in media criticism and media scolding. Instead of transparent diversity of opinion, the opinion of the government or powerful groups can dominate the media unquestioned as a uniform majority opinion, especially in authoritarian societies with limited media freedom (e.g., Russia or China). But the question also arises for media in democracies as to whether and to what extent they are concretely committed to more or less equality in society. Instead of contributing to integration and solidarity about migrants or other minorities, media can contribute to stereotyping and reinforce discrimination through blanket negative reporting. Finally, there is always the danger of unjustifiably discrediting individuals or social groups through one-sided moralizing portrayals. Content analyses of (German, editor's note) media coverage show that migrants and especially Muslims tend to feature little in the media, and when they do, they are portrayed stereotypically and negatively [2-50, 2-52].

Events of the recent past in the media sector give cause for concern that the quality of media reporting is in danger. Warning voices even speak of a media crisis (cf. [2-53]). In the print sector as well as in broadcasting, a growing media concentration has been underway among media groups for some time: large media groups are becoming increasingly dominant. In parallel, advertising spending is shifting from the press to the Internet, and newspaper use is declining. At the level of media organizations, this has led not least to the layoff of media professionals, the downsizing of newsrooms, and the creation of lower-cost newsrooms. In the newsroom, content is produced jointly for the print edition and the online offering. Journalists thus no longer write an article just for the newspaper, but also create online versions or radio or TV reports at the same time. This has led not least to an increase in the time pressure of journalistic work. But the media crisis is not just a funding crisis; journalism is also affected in terms of content. Commercialization has not only led to a decrease in media diversity but the economic pressure is also expressed in an increased external influence of public relations on reporting, for example as courtesy journalism. The blurring of the boundaries between editorial and advertising sections (keyword: native advertising) endangers journalistic independence. As a result of economization, there is also an increased focus on the audience and its wishes. Information and entertainment, as well as the public and private spheres of politicians, for example, are mixed in reporting to make it more interesting. Criticism of the media focuses on the tabloid press on the one hand and private broadcasting on the other under the headings of "personalization" [2-54] and "infotainment" [2-55]. Both are accused of populism and a lack of independence, as well as a generally low level of quality [2-50].

Infotainment / Personalization

The mixing of informative and entertaining formats of television is called infotainment. The first part of the word comes from "information", the second part is derived from the Anglo-American term "entertainment". As a rule, this describes the tendency, for example, to include more and more "soft" topics such as news about celebrities in news programs. Infotainment also refers to the increasing emotionalization and personalization of news, the latter meaning the focus on a specific person (presenter, "anchorman").[121]

Current crises, such as the COVID-19 pandemic or climate change, have brought about another development that, while not unknown, is putting media ethics to the test. This is the journalistic professional standard of reporting news in as balanced and neutral a manner as possible to also allow for nuanced voice. Journalists adhere to this norm to demonstrate their professionalism and avoid any criticism of their one-sided reports. At the same time, balance can also substitute for plausibility checks, as when reporters do not have enough time for research, or their own skills are not sufficient to assess the validity of certain contradictory statements. When controversial statements are combined with reporters' lack of expertise, the standards of balance become particularly apparent. However, when the voices of dissenting outsiders are copied out of context, it provides them with legitimacy and media prestige that may also enable them to gain political power. Research also shows that ideological bias can also play an important role: For example, right-wing and conservative columnists predominantly allow climate change deniers to have their say in their articles [2-56].

Figure 21: Example of false balance media coverage.[122]

Especially in the context of scientific reporting on the causes and combating of current crises, the term "false balance" has become more popular due to the journalistic practice described above. While balanced and independent reporting is generally considered good journalism, and this balance does have its purpose. The public should always be able to trust that all important aspects will be published in the media, and everyone should be able to find out about all sides of an issue. At the same time,

---

121 [2-50], own translation.
122 https://commons.wikimedia.org/wiki/File:False_balance.jpg (last accessed 21.01.2022).

this should not mean that all sides of an issue deserve equal weight either - e.g., one guest "pro" and one "contra". Science is based on evidence for different hypotheses, which are carefully tested and then built upon those that provide the most evidence. However, when all different opinions and viewpoints are presented side by side and the same proportion in the press or in politics, the impression of equal weight scientific legitimacy is falsely created. In this way, one of the main goals of science is defeated: weighing the evidence.[123]

False balance
is a misleading argument that presents two or more positions as equally valid when the evidence strongly supports one over the others. This may be unintentional faulty reasoning. Alternatively, a false balance may be used to influence and mislead.[124]

Moreover, polarizing topics are particularly favored by the media. For example, in the area of vaccinations against the coronavirus (COVID-19), this seemingly balanced presentation contradicts both the expert consensus expressed in the vaccination recommendations of the European Medicines Agency (EMA)[125] and the national control authorities, such as the German STIKO,[126] and the distribution of opinion in the population [2-57], thus falsely suggesting an equal distribution of opinions. Such false balance leads to uncertainty [2-58] and raises doubts in the population and favors the emergence of fake news, especially in social media. Making the media and their reporting aware of this and calling on them to weigh the evidence rather than the opinions is also an important contribution to strengthening public confidence in vaccination, researchers urge [2-59, pp. 404-405].

The project "Understanding Science" by the University of California Museum of Paleontology, Berkeley, has summarized six steps that can be applied to evaluate scientific messages:[127]

1. Where does the information come from?
   – What is the source of this message?
   – Does that source have an agenda or goal?

2. Are the views of the scientific community accurately portrayed?
   – Who is the expert?
   – Beware of false balance.

3. Is the scientific community's confidence in the ideas accurately portrayed?
   – Science is always ready to revise ideas when new evidence justifies them.
   – Tentativeness does not mean that scientific ideas are untrustworthy … and this is where some media reports on science can mislead, mistaking normal scientific proceedings for untrustworthiness.

4. Is a controversy misrepresented or blown out of proportion?
   – Fundamental scientific controversy: scientists disagreeing about a central hypothesis or theory.

---

[123] https://undsci.berkeley.edu/article/sciencetoolkit_04 (last accessed 04.01.2022).
[124] https://simplicable.com/new/false-balance (last accessed 04.01.2022).
[125] https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-authorised (last accessed 04.01.2022).
[126] https://www.rki.de/DE/Content/Infekt/Impfen/ImpfungenAZ/COVID-19/Impfempfehlung-Zusfassung.html (last accessed 04.01.2022).
[127] https://undsci.berkeley.edu/article/0_0_0/sciencetoolkit_02 (last accessed 04.01.2022).

- Secondary scientific controversy: scientists disagree about a less central aspect of a scientific idea.
- Conflict over the ethicality of methods: disagreement within the scientific community or society at large over the appropriateness of a method used for scientific research.
- Conflict over applications: conflict over the application of scientific knowledge.
- The conflict between scientific ideas and non-scientific viewpoints.

5. Where can I get more information?
   - Find sources with scientific expertise.
   - Avoid ulterior motives.
   - Keep it current.
   - Check for citations.

6. How strong is the evidence?
   - Does the evidence suggest correlation or causation? In other words, do the data suggest that two factors (e.g., high blood pressure and heart attack rates) are correlated with one another or that changes in one cause changes in the other?
   - Is the evidence based on a large sample of observations (e.g., 10,000 patients with high blood pressure) or just a few isolated incidents?
   - Does the evidence back up all the claims made in the article (e.g., about the cause of heart attacks, a new blood pressure drug, and preventative strategies) or just a few of them?
   - Are the claims in the article supported by multiple lines of evidence (e.g., from clinical trials, epidemiological studies, and animal studies)?
   - Does the scientific community find the evidence convincing?

2.5.3. The role of social media platforms
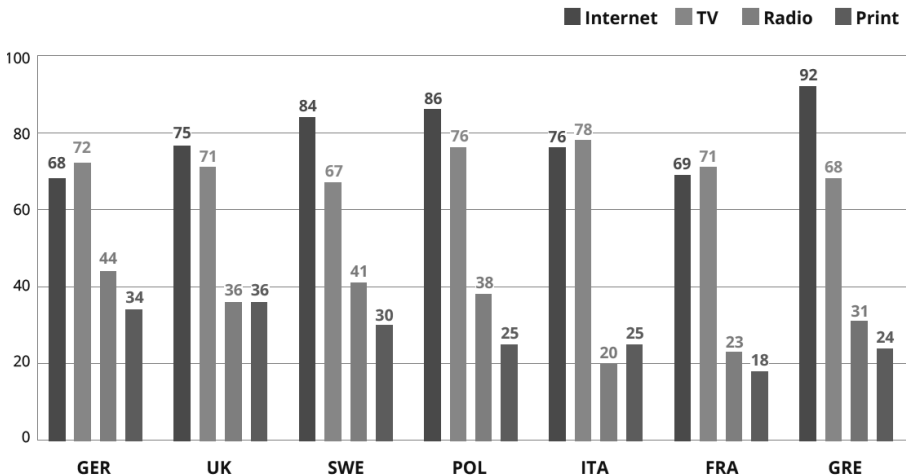
Source: Reuters Institute Digital News Survey 2019

Bases: GER: n = 2022; UK: n = 2023; SWE: n = 2007; POL: n = 2009; ITA: n = 2006; FRA: n = 2005; GRE: n = 2018

Figure 22: Regularly used news sources 2019 (Percentage).
Source: [2-60, p. 18]

Unlike in the past, when information was only obtained from mass media such as newspapers, television, and radio, today we are constantly flooded with information. Social media platforms (Twitter, Facebook, Instagram, YouTube, etc.) and messenger services (Telegram, WhatsApp, etc.) have grown immensely in importance and now play a key role in the spread of information. But not only news is shared on social media - regardless of its truth content - but also opinions. The problem with this is that opinions often do not necessarily match the facts. No matter how convincing facts may be, they do not always influence the opinion once it has been formed. Scientists have known about the phenomenon of unbending personal opinion for decades (e.g. [2-61, 2-62]).

This phenomenon of unbending personal opinion is known as *confirmation bias*. Confirmation bias is "the tendency to gather evidence that confirms preexisting expectations, typically by emphasizing or pursuing supporting evidence while dismissing or failing to seek contradictory evidence"[128]. In other words, it describes the fact that we generally regard information that supports our worldview as credible and relevant, while we ignore or dismiss as nonsense information that does not fit into our worldview.

The developers of social media platforms and online search engines are well informed about how human cognitive biases work. With this knowledge, personalized technologies - called algorithms - are created to determine what people should see online. These personalization technologies are built to pick only the most interesting and relevant content for each user, and can thus reinforce users' cognitive and social biases and make them more susceptible to misinformation and fake news. The detailed advertising tools built into many social media platforms, for example, allow disinformation activists to exploit confirmation bias by tailoring messages to people who are already inclined to believe them. If a user frequently clicks on Facebook links from a particular news source, Facebook tends to show that person more content from that source. This creates what is known as the "filter bubble" or "echo chamber" effect [2-63]. This effect can isolate people from other perspectives and reinforce confirmation bias [2-64].

Another important component of social media is information trending on the platform based on what gets the most clicks. Authors and researchers Giovanni Luca Ciampaglia and Filippo Menczer refer to this as popularity bias. In their research, they found that an algorithm that aims to promote popular content can have a negative impact on the overall quality of information on the platform. This similarly impacts existing cognitive biases, as what appears to be popular is promoted regardless of its quality. Such algorithmic biases can be manipulated by the effect of the social bots [2-65]. These computer programs interact with humans via social media accounts and most, such as Big Ben[129] from Twitter, are harmless. However, there are also social bots in use that hide their true nature and serve malicious purposes, such as spreading disinformation or falsely creating the appearance of a grassroots movement[130], also called "astroturfing"[131] [2-64].

To study how the structure of online social networks makes users vulnerable to disinformation, the Hoaxy system was developed.[132] Hoaxy tracks the spread of content from sources with low credibility visualizes it and shows how it competes with fact-checking content. Analysis of data collected by Hoaxy during the 2016 U.S. presidential election found that Twitter accounts that spread misinformation were almost completely cut off from fact-checkers corrections [2-22]. Accounts that

---

[128] https://dictionary.apa.org/confirmation-bias (last accessed 27.10.2021).
[129] https://twitter.com/big_ben_clock (last accessed 04.01.2022).
[130] e.g., https://www.businessinsider.com/astroturfing-grassroots-movements-2011-9 (last accessed 04.01.2022).
[131] https://lobbypedia.de/wiki/Astroturfing or https://www.merriam-webster.com/dictionary/astroturfing (last accessed 04.01.2022).
[132] https://hoaxy.osome.iu.edu (last accessed 04.01.2022).

spread misinformation were examined more closely. The authors found a very dense core group of accounts that retweeted each other almost exclusively - including several bots [2-64].
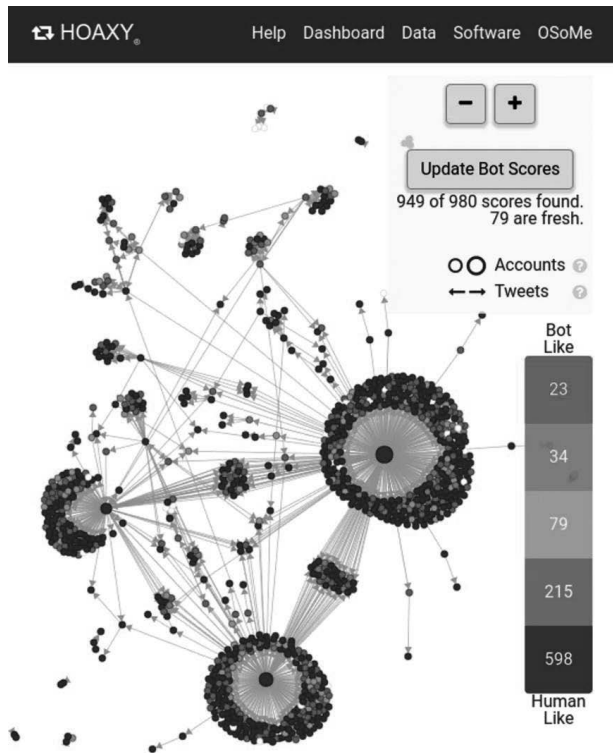
Figure 23: A screenshot of a Hoaxy search shows how common bots – in red and dark pink – are spreading a false story on Twitter.[133]

As a collaborative project of the Observatory on Social Media (OSoMe, pronounced "awesome")134 at Indiana University, USA, the tool "Botometer"135 was also developed. This tool helps to detect social bots on Twitter and investigate the manipulation strategies behind them. Botometer uses machine learning (a branch of artificial intelligence – editor's note) to detect bot accounts by examining thousands of different characteristics of Twitter accounts, such as the timing of their posts, the frequency of their tweets, and the accounts they follow and retweet. It turned out that up to 15 percent of active Twitter accounts showed signs of bots [2-66]. In combination with Hoaxy, many bots have been discovered that exploit their victims' cognitive, confirmation, and popularity preferences as well as Twitter's algorithmic preferences. The social bots enable filter bubbles to be built around vulnerable users and submit fake news to them. The bots first attract the attention of users who, for example, support a particular politician by tweeting that candidate's hashtags or mentioning/retweeting the person. In addition, the bots can reinforce false claims and slander

---

[133] https://theconversation.com/misinformation-and-biases-infect-social-media-both-intentionally-and-accidentally-97148 (last accessed 04.01.2021).
[134] https://osome.iu.edu (last accessed 04.01.2022).
[135] https://botometer.osome.iu.edu (last accessed 04.01.2022).

opponents by retweeting articles from less credible sources that match certain keywords. In this way, the algorithm also highlights false stories that are frequently shared with other users [2-64].

Figure 24: A screenshot of the Botometer website, which checks the followers of German Foreign Minister Annalena Baerbock for possible bot accounts.[136]

As already discussed above, misinformation on the Internet has a direct impact on our society and its political, economic, and ideological orientation. The essential role of social media in the spread of fake news is currently receiving a lot of attention, and the political developments of recent years illustrate the pressing problem: Facebook and Cambridge Analytica, which significantly influenced the election of Donald Trump as president of the USA and likewise the Brexit referendum in 2016 [2-67] or the widespread skepticism about human influence on climate change,[137] for example.

In December 2021, it was announced that Rohingya who had fled the genocide in Myanmar was suing Facebook's parent company, Meta, for 150 billion US-Dollars in damages. They allege that Facebook's algorithm, which promoted extremism and violence, was knowingly and willfully used,

---

[136] Screenshot Botometer (last accessed 04.01.2022).
[137] https://www.encyclopedia.com/environment/energy-government-and-defense-magazines/climate-change-skeptics (last accessed 09.11.2021).

and that the corporation was complicit in the 2017 genocide. They claim that racist hatred of the Rohingya was deliberately fomented on the platform, misinformation was spread, and calls were made for violence against the predominantly Muslim ethnic group. "For example, the military called us 'animals', 'monkeys', 'donkeys', 'rapists' - this spread further and further on Facebook", recalls Rohingya Ambia Perveen. An independent expert report confirms Facebook's complicity.[138]

Already before the Rohingya lawsuit, Facebook's involvement in spreading propaganda and fake news was highlighted by the revelations of former Facebook employee and whistleblower Frances Haugen. On the 4th of October 2021 hearing before the U.S. Senate, Haugen called for strict regulation of the global enterprise: It needs pressure, monitoring and oversight because so far Facebook has been a black box. Facebook's products harm children, stoke division in society, and weaken democracy. The company's leadership, she said, knows how to make Facebook and Instagram safer, but does not want to make the necessary adjustments because it puts its immense profits above the common good and the truth. "This has to change", Haugen told the U.S. Senate, and for that reason, she decided to let the public know about it.[139]

What does it mean that Facebook knows how to make its service safer? The answer to this question is hidden behind the algorithm used by Facebook. Facebook and every other social media service work with their algorithms. There is no clear definition of algorithms in social and cultural studies research. They are often understood as hidden and powerful mechanisms that have a great influence on our digital lives. The problem with algorithms stems from their lack of transparency, which is difficult for outsiders to comprehend - their detailed mode of operation is a closely guarded secret of social media providers [2-68, pp. 181-182]. Tarleton Gillespie describes the algorithm as a recipe assembled in programmable steps. Applied to social media providers, this, therefore, means that developers first define a model in which the problems to be solved and the goals to be achieved are formulated. Once such a model has been formulated, the recipe for it, i.e. the algorithm, is developed [2-69, p. 19].

From the perspective of social media users, this lack of transparency is particularly problematic because they do not know what content is being hidden from them by the algorithms. The systematic hiding of more differentiated content and the display of one-sided content can lead to a distorted view of the world, as described by the metaphors of the filter bubble and the echo chamber. This applies to political content as well as to ideologies and extreme viewpoints of all kinds [2-60, pp. 12-13].

---

[138] Ambia Perveen in conversation with Michael Borgers (text by Mike Herbstreuth): https://www.deutschlandfunk.de/rohingya-klage-facebook-100.html (last accessed 05.01.2022).
[139] statement of Frances Haugen: https://www.washingtonpost.com/context/facebook-whistleblower-frances-haugen-s-senate-testimony/8d324185-d725-4d99-9160-9ce9e13f58a3/ (last accessed 09.11.2021).

Main Findings:

- Algorithms largely govern the selection, sorting, and presentation of information on the Internet.

- The logics that underpin algorithmic gatekeeping differ from the logics of human gatekeeping.

- Algorithms customize content to the interests and preferences of each user (personalization).

- Due to their business model—advertising—tech companies like Facebook or Google try to maximize the amount of time people spent on their platforms.

- Algorithms, like Facebook's news feed values, emphasize personal significance to increase audience engagement with particular types of content, whereas journalistic gatekeeping emphasizes social significance—oriented toward the public tasks of journalism.

- A deeper understanding of the contemporary gatekeeping process requires a detailed examination of the generally opaque algorithmic systems.

Figure 25: Main findings of Algorithmic Information Filtering.
Source: [2-60, p. 13]

Laura Chinchilla, former president of Costa Rica and chair of the Kofi Annan Commission on Elections and Democracy in the Digital Age, summed up the not-so-simple truths about digital social media platforms:

[...] On the one hand, digital technologies have played a vital role in providing free access to government data and information; encouraging citizen participation in public decision making; introducing new voices to the public debate; fostering the transparency and scrutiny of administrative actions; knitting global advocacies together on issues affecting human rights, the rule of law and democracy; and mobilizing new actors eager to find alternative avenues for political participation. The Arab Spring almost a decade ago, the pro-democracy protests in Hong Kong this summer and the toppling of Puerto Rico's governor in July are only a few examples.
On the other, the alarming number of episodes involving the use of social media platforms to manipulate elections and public debates, as well as the surge of extremist groups using the internet to incite hatred and violence, clearly warns us that the adverse relationship between those platforms and democracy is no longer just anecdotal.
Fake news is as old as news, and hate speech is as old as speech. But the digital age has provided a ripe environment for the virulent reproduction and visibility of both. To be clear, the promise of the betterment of the human condition held by new technologies is beyond question. But the risks have become just as apparent. [...] [2-70]

In summary, social media networks in particular can contribute to the polarization and radicalization of their users by reflecting a distorted picture of news as well as opinions and by reinforcing cognitive biases. Moreover, fake news spreads much faster than true news. The rapid spread was confirmed by an MIT study, whose results showed that *"the top 1% of false news cascades diffused to between 1000 and 100,000 people, whereas the truth rarely diffused to more than 1000 people"* [2-71].

References Chapter 2

[2-1]   UNITED NATIONS, Universal Declaration of Human Rights. Available at https://www.un.org/en/about-us/universal-declaration-of-human-rights (last accessed 06 January 2022).

[2-2]   EUROPEAN COURT OF HUMAN RIGHTS, Council of Europe's work on Hate Speech. Available at https://www.coe.int/en/web/committee-on-combatting-hate-speech/council-of-europe-work-on-hate-speech (last accessed 13 October 2021).

[2-3]   EUROPEAN COURT OF HUMAN RIGHTS, European Convention on Human Rights, Strasbourg 2013. Available at https://www.echr.coe.int/documents/convention_eng.pdf (last accessed 01 December 2021).

[2-4]   DEUTSCHER BUNDESTAG, Fake-News, Definition und Rechtslage, 2017. Available at https://www.bundestag.de/resource/blob/502158/99feb7f3b7fd1721ab4ea631d8779247/wd-10-003-17-pdf-data.pdf (last accessed 05 December 2021).

[2-5]   MACAVANEY, S., HAO-REN, Y., YANG, E., RUSSELL, K., GOHARIAN, N., FRIEDER, O., Hate speech detection: Challenges and solutions, in: PLoS ONE 14(8) (2019). Available at https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0221152 (last accessed 20 January 2022).

[2-6]   CARLSON, C. R., Hate Speech, Cambridge, Massachusetts: The MIT Press, 2021. Available at https://doi.org/10.7551/mitpress/12491.003.0001 (last accessed 20 January 2022).

[2-7]   EUROPEAN COURT OF HUMAN RIGHTS, Hate speech, 2020. Available at https://www.echr.coe.int/documents/fs_hate_speech_eng.pdf (last accessed 18 October 2021).

[2-8]   EUROPEAN COURT OF HUMAN RIGHTS, Guide on Article 8 of the European Convention on Human Rights, 2021. Available at https://www.echr.coe.int/documents/guide_art_8_eng.pdf (last accessed 01 December 2021).

[2-9]   COUNCIL OF EUROPE, Digital Resistance, 2020. Available at https://rm.coe.int/digital-resistance-handbook-21012021/1680a1269d (last accessed 27 September 2021).

[2-10]   MAZZEO, V., RAPISARDA, A., GIUFFRIDA, G., Detection of Fake News on COVID-19 on Web Search Engines, 2021. Available at https://www.frontiersin.org/article/10.3389/fphy.2021.685730 (last accessed 20 January 2022).

[2-11]   EUROPEAN COMMISSION AGAINST RACISM AND INTOLERANCE, ECRI General Policy Recommendation No.15 on Combating Hate Speech: key points, 2016. Available at https://rm.coe.int/16805d59ee (last accessed 20 January 2022).

[2-12]   LAAKSONEN, S.-M., HAAPOJA, J., KINNUNEN, T., NILIMARKKA, M., PÖYHTÄRI, R., The Datafication of Hate: Expectations and Challenges in Automated Hate Speech Monitoring, 2020. Available at https://www.frontiersin.org/article/10.3389/fdata.2020.00003 (last accessed 20 January 2022).

[2-13]   PATTON, D. U., FREY, W. R., MCGREGOR, K. A., LEE, F.-T.; MCKEOWN, K., MOSS, E., Contextual Analysis of Social Media: The Promise and Challenge of Eliciting Context in Social Media Posts with Natural Language Processing, in: Proceedings of the 2020

AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), New York 2020. Available at https://doi.org/10.1145/3375627.3375841 (last accessed 20 January 2022).

[2-14] COUNCIL OF EUROPE, Compilation of Venice: Commission opinions and reports concerning freedom of expression and media, in: European Commission for democracy through law, Strasbourg 2020. Available at https://www.venice.coe.int/webforms/documents/?pdf=CDL-PI(2020)008-e (last accessed 13 October 2021).

[2-15] COUNCIL OF EUROPE, Information Disorder: Toward an interdisciplinary framework for research and policy making, in: Council of Europe report DGI (2017)09, Strasbourg 2017. Available at 0900001680766412 (coe.int) (last accessed 13 October 2021).

[2-16] MEISTER, S., Isolation and Propaganda: the roots and instruments of Russia's disinformation campaign, in: Transatlantic Academy Paper Series 2015-16 No.6, Washington DC 2016. Available at https://dgap.org/system/files/article_pdfs/meister_isolationpropoganda_apr16_web_1.pdf (last accessed 11 December 2021).

[2-17] DOUBLET, Y-M., in: Disinformation and electoral campaigns, Council of Europe, Strasbourg 2019. Available at https://book.coe.int/fr/droit-international/7985-disinformation-and-electoral-campaigns.html (last accessed 20 October 2021).

[2-18] TURČILO, L., OBRENOVIĆ, M., Fehlinformationen, Desinformationen, Malinformationen: Ursachen, Entwicklungen und ihr Einfluss auf die Demokratie, in: Heinrich Böll Stiftung, Demokratie im Fokus #3, 2020. Available at https://www.boell.de/sites/default/files/2020-08/200825_E-Paper3_DE.pdf (last accessed 03 November 2021).

[2-19] AKERLOF, G.A., The Market for "Lemons": Quality Uncertainty and the Market Mechanism, in: The Quarterly Journal of Economics, Vol. 84 No.3, 1970.

[2-20] DR. STAHEL, L., in: Status quo und Massnahmen zu rassistischer Hassrede im Internet: Übersicht und Empfehlungen, Soziologisches Institut, Universität Zürich, 2020. Available at https://www.edi.admin.ch/dam/edi/de/dokumente/FRB/Neue%20Website%20FRB/T%C3%A4tigkeitsfelder/Medien_Internet/Stahel_2020_Status%20quo%20und%20Massnahmen%20zu%20rassistischer%20Hassrede.pdf.download.pdf/Stahel_2020_Status%20quo%20und%20Massnahmen%20zu%20rassistischer%20Hassrede.pdf (last accessed 09 November 2021).

[2-21] GABIELKOV, M., RAMACHANDRAN, A., CHAINTREAU, A., LEGOUT, A. (2016). Social Clicks: What and Who Gets Read on Twitter? ACM SIGMETRICS Perform. Eval. Rev. 44, 1 (June 2016), pp. 179–192, https://doi.org/10.1145/2964791.2901462.

[2-22] SHAO, C., HUI, P.-M., WANG, L., JIANG, X., FLAMMINI, A., MENCZER, F., et al. (2018). Anatomy of an online misinformation network. PLoS ONE 13(4): e0196087, https://doi.org/10.1371/journal.pone.0196087.

[2-23] SILVERMAN, C., ALEXANDER, L. How Teens In The Balkans Are Duping Trump Supporters With Fake News. BuzzFeedNews, November 3, 2016 (online), https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo#.nfGBdzv3rN (last accessed November 24, 2021).

[2-24] KING, G., PAN, J., & ROBERTS, M. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. American Political Science Review, 111(3), pp. 484-501, doi:10.1017/S0003055417000144.

[2-25] PENNYCOOK, G., RAND. D.G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, Volume 188, 2019, pp. 39-50, https://doi.org/10.1016/j.cognition.2018.06.011.

[2-26] WEIR, K. Why we fall for fake news: Hijacked thinking or laziness? American Psychological Association, 11.02.2020 (online), https://www.apa.org/news/apa/2020/fake-news (last accessed December 4, 2021).

[2-27] ROSS, R. M., RAND, D. G., PENNYCOOK, G. (2019). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. PsyArXiv, 13 November 2019, https://doi.org/10.31234/osf.io/cgsx6.

[2-28] BAGO, B., RAND, D. G., PENNYCOOK, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. Journal of Experimental Psychology: General, 149(8), pp. 1608–1613, https://doi.org/10.1037/xge0000729.

[2-29] BRONSTEIN, M.V., PENNYCOOK, G., BEAR, A., RAND. D.G., CANNON, T.D. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. Journal of Applied Research in Memory and Cognition, Volume 8, Issue 1, 2019, pp. 108-117, https://doi.org/10.1016/j.jarmac.2018.09.005.

[2-30] PENNYCOOK, G., RAND, D.G. (2019). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality. 2020, 88, pp. 185– 200, https://doi.org/10.1111/jopy.12476.

[2-31] SUNDAR, S.S., KNOBLOCH-WESTERWICK, S., HASTALL, M.R. (2007). News cues: Information scent and cognitive heuristics. Journal of the American Society for Information Science and Technology, 58, pp. 366-378, https://doi.org/10.1002/asi.20511.

[2-32] SCHMITT-BECK, R. (2015). Bandwagon Effect. In The International Encyclopedia of Political Communication, G. Mazzoleni (Ed.), https://doi.org/10.1002/9781118541555.wbiepc015.

[2-33] LEVITAN, L.C., VERHULST, B. (2016). Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change. Political Behavior, 38, pp. 277–315, https://doi.org/10.1007/s11109-015-9312-x.

[2-34] MALLINSON, D.J., HATEMI, P.K. (2018). The effects of information and social conformity on opinion change. PLoS ONE 13(5): e0196600, https://doi.org/10.1371/journal.pone.0196600.

[2-35] INSKO, C.A., SMITH, R.H., ALICKE, M.D., WADE, J., TAYLOR, S. (1985). Conformity and Group Size: The Concern with Being Right and the Concern with Being Liked. Personality and Social Psychology Bulletin, 1985, 11(1), pp. 41-50, https://doi.org/10.1177/0146167285111004.

[2-36] KISS, Á., SIMONOVITS, G. (2013). Identifying the bandwagon effect in two-round elections. Public Choice, 160, 2014, pp. 327–344, https://doi.org/10.1007/s11127-013-0146-y.

[2-37] MORWITZ, V.G., PLUZINSKI, C. (1996). Do Polls Reflect Opinions or Do Opinions Reflect Polls? The Impact of Political Polling on Voters' Expectations, Preferences, and Behavior. Journal of Consumer Research, Volume 23, Issue 1, June 1996, pp. 53–67, https://doi.org/10.1086/209466.

[2-38] BENECKE, O., DEYOUNG, S.E. (2019). Anti-Vaccine Decision-Making and Measles Resurgence in the United States. Global Pediatric Health. January 2019, https://doi.org/10.1177/2333794X19862949.

[2-39] GUESS, A.M., NYHAN, B., REIFLER, J. (2018). Exposure to untrustworthy websites in the 2016 U.S. election. Available at https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf (last accessed 04.12.2021).

[2-40] KAHAN, D.M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. Cultural Cognition Project Working Paper Series No. 164, Yale Law School, Public Law Research Paper No. 605, Yale Law & Economics Research Paper No. 575, Available at SSRN: https://ssrn.com/abstract=2973067 or http://dx.doi.org/10.2139/ssrn.2973067.

[2-41] KAHAN, D.M. (2013). Ideology, motivated reasoning, and cognitive reflection. Judgment and Decision Making, Vol. 8, No. 4, July 2013, pp. 407-424. Available at http://journal.sjdm.org/13/13313/jdm13313.pdf.

[2-42] VAN BAVEL, J.J., PEREIRA, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. Trends in Cognitive Sciences 22 (3), pp. 213–224, https://doi.org/10.1016/j.tics.2018.01.004.

[2-43] PENNYCOOK, G., RAND, D.G. (2021). The Psychology of Fake News. Trends in Cognitive Sciences 25 (5), pp. 388–402, https://doi.org/10.1016/j.tics.2021.02.007.

[2-44] SYDELL, L. We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned. NPR, November 23, 2016 (online), https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs?t=1638618145394 (last accessed December 4, 2021).

[2-45] GABBATT, A. 'Incited by the president': politicians blame Trump for insurrection on Capitol Hill. The Guardian, January 7, 2021 (online), https://www.theguardian.com/us-news/2021/jan/06/donald-trump-politicians-insurrection-capitol-hill, (last accessed December 4, 2021).

[2-46] U.S. DEPARTMENT OF STATE (2020). GEC Special Report: Pillars of Russia's Disinformation and Propaganda Ecosystem. Available at https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia's-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf.

[2-47] THORSON, E. (2015). Belief Echoes: The Persistent Effects of Corrected Misinformation. Political Communication, 33 (3) 2016, pp. 460-480, https://doi.org/10.1080/10584609.2015.1102187.

[2-48] PENNYCOOK, G., BEAR, A., COLLINS, E.T., RAND, D.G. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. Management Science 66 (11), pp. 4944–4957, https://doi.org/10.1287/mnsc.2019.3478.

[2-49] DEL VICARIO, M., BESSI, A., ZOLLO, F., PETRONI, F., SCALA, A., CALDARELLI, G. et al. (2016). The spreading of misinformation online. Proceedings of the National Academy of Sciences of the United States of America 113 (3), pp. 554–559, https://doi.org/10.1073/pnas.1517441113.

[2-50] BONFADELLI, H. Medien und Gesellschaft im Wandel. Bundeszentrale für politische Bildung, December 9, 2016 (online), https://www.bpb.de/gesellschaft/medien-und-sport/medienpolitik/236435/medien-und-gesellschaft-im-wandel (last accessed December 29, 2021).

[2-51] LUHMANN, N. (1996). Die Funktion der Massenmedien. Die Realität der Massenmedien, VS Verlag für Sozialwissenschaften, Wiesbaden, https://doi.org/10.1007/978-3-663-01103-3_13.

[2-52] SCHIFFER, S. Der Islam in deutschen Medien. Bundeszentrale für politische Bildung, May 10, 2005 (online), https://www.bpb.de/apuz/29060/der-islam-in-deutschen-medien (last accessed December 29, 2021).

[2-53] JARREN, O., KÜNZLER, M., PUPPIS, M. (2012). Medienwandel oder Medienkrise? Folgen für Medienstrukturen und ihre Erforschung. Nomos Verlagsgesellschaft, pp. 9–25, https://doi.org/10.5771/9783845236735-9.

[2-54] HOFFMANN, J., RAUPP, J. (2006). Politische Personalisierung. Pub 51, pp. 456–478, https://doi.org/10.1007/s11616-006-0240-y.

[2-55] BERNHARD, U., SCHARF, W. (2008). „Infotainment" in der Presse. Pub 53, pp. 231–250, https://doi.org/10.1007/s11616-008-0077-7.

[2-56] BRÜGGEMANN, M., ENGESSER, S. (2017). Beyond false balance: How interpretive journalism shapes media coverage of climate change. Global Environmental Change 42, pp. 58–67, https://doi.org/10.1016/j.gloenvcha.2016.11.004.

[2-57] HORSTKÖTTER, N., MÜLLER, U., OMMEN, O., PLATTE, A., RECKENDREES, B., STANDER, V., LANG, P., THAISS, H. (2017). Einstellungen, Wissen und Verhalten von Erwachsenen und Eltern gegenüber Impfungen – Ergebnisse der Repräsentativbefragung 2016 zum Infektionsschutz. BZgA-Forschungsbericht. Köln: Bundeszentrale für gesundheitliche Aufklärung. Available at https://www.bzga.de/fileadmin/user_upload/PDF/studien/infektionsschutzstudie_2016--f4f414f596989cf814a77a03d45df8a1.pdf (last accessed January 23, 2022).

[2-58] DIXON, G. N., CLARKE, C. E. (2013). Heightening Uncertainty Around Certain Science: Media Coverage, False Balance, and the Autism-Vaccine Controversy. Science Communication, 35 (3), pp. 358–382. https://doi.org/10.1177/1075547012458290.

[2-59] BETSCH, C., SCHMID, P., KORN, L. et al. (2019). Impfverhalten psychologisch erklären, messen und verändern. Bundesgesundheitsblatt 62, pp. 400–409, https://doi.org/10.1007/s00103-019-02900-6, also available at https://www.rki.de/DE/Content/Kommissionen/Bundesgesundheitsblatt/Downloads/2019_04_Betsch.pdf?__blob=publicationFile (last accessed January 23, 2022).

[2-60] STARK, B., STEGMANN, D., MAGIN, M., JÜRGENS, P. (2020). Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse. AW

AlgorithmWatch. Available at https://algorithmwatch.org/en/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf (last accessed January 23, 2022).

[2-61] ROSS, L., LEPPER, M. R., HUBBARD, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 32 (5), pp. 880–892, https://doi.org/10.1037/0022-3514.32.5.880.

[2-62] NICKERSON, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology, 2 (2), pp. 175–220, https://doi.org/10.1037/1089-2680.2.2.175.

[2-63] SAXENA, R. The social media "echo chamber" is real. ARS Technica, March 13, 2017 (online), https://arstechnica.com/science/2017/03/the-social-media-echo-chamber-is-real/ (last accessed January 04, 2022).

[2-64] CIAMPAGLIA, G.L., MENCZER, F. Misinformation and biases infect social media, both intentionally and accidentally. The Conversation, June 20, 2018 (updated January 10, 2019) (online), https://theconversation.com/misinformation-and-biases-infect-social-media-both-intentionally-and-accidentally-97148 (last accessed January 04, 2022).

[2-65] FERRARA, E., VAROL, O., DAVIS, C., MENCZER, F., FLAMMINI, A. (2016). The Rise of Social Bots. Communications of the ACM, July 2016, Vol. 59 No. 7, pp. 96-104, 10.1145/2818717. Available at https://cacm.acm.org/magazines/2016/7/204021-the-rise-of-social-bots/fulltext (last accessed January 04, 2022).

[2-66] VAROL, O., FERRARA, E., DAVIS, C.A., MENCZER, F., FLAMMINI, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. AAAI Publications, Eleventh International AAAI Conference on Web and Social Media. Available at https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587 (last accessed January 25, 2022).

[2-67] ADAMS, T. Facebook's week of shame: the Cambridge Analytica fallout. The Guardian, March 24, 2018 (online), https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence (last accessed November 09, 2021).

[2-68] HERDER, J. (2018). Regieren Algorithmen? Über den sanften Einfluss algorithmischer Modelle. (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft), pp. 179-203, ISBN: 978-3-9818892-5-3, https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57547-2.

[2-69] GILLESPIE, T. (2016). Algorithm. In B. Peters (Ed.), Digital Keywords: A Vocabulary of Information Society and Culture, pp. 18–30. Princeton University Press, https://doi.org/10.2307/j.ctvct0023.6.

[2-70] CHINCHILLA, L. Post-Truth Politics Afflicts the Global South, Too. The New York Times, October 15, 2019 (online), https://www.nytimes.com/2019/10/15/opinion/politics-global-south.html, also https://www.kofiannanfoundation.org/supporting-democracy-and-elections-with-integrity/annan-commission/post-truth-politics-afflicts-the-global-south-too/ (last accessed December 09, 2021).

[2-71] VOSOUGHI, S., ROY, D., ARAL, S. (2018). The spread of true and false news online. Science (New York, N.Y.) 359 (6380), p. 1146–1151, https://www.science.org/doi/10.1126/science.aap9559.

3. Technical Foundations – how the Internet works and why technical remedies are of limited use

Authors: Anna Funk, Christian Hönig and Christian Munz
Academic supervisor: Tamás Szadeczky

3.1. Introduction

A probable cause why remedies applied by political and administrative bodies are not satisfactorily successful is that administrative staff and political deciders have limited knowledge of how both hate speech and fake news work from a technical perspective. One needs a sound understanding of how posting and disseminating hate speech and fake news or, generally spoken, the whole internet works. The goal of this chapter is to describe how it works. This does of course not include any judgment on whether this is good or bad – it is simply a description of mechanisms and the status quo.

The main message of this chapter is, that the internet is not, absolutely not comparable to any other media we know and to the legal setting we know from these other media. The main reasons for this are that the internet has no proprietor, organizer, authority whatsoever (1) and that a person or computer program which produces hate speech and fake news or any internet content is not located in the same country as the person affected (2) and hence not subject to the jurisdiction of this country. More than that, it is neither sure nor likely that the internet platform used is in either the country of the person producing hate speech and fake news or in the country of the person affected (3).

These three phenomena are different from a classic media setting like newspapers, radio, or television.

3.2. The Internet: A world without much Governance

The whole nature of the internet is that of a non-governed, self-administrated organism. As Leiner et al. have shown in their "A brief history of the internet" [3-1], the internet was not intended to host political discussions, hate speech, etc. "There would be no global control at the operations level." was one of four fundamental principles (called ground rules) stated by Robert E. Kahn [3-1, p. 24]. In 1969 the Requests for Comments (RFCs) were introduced by S. Crocker of UCLA, which were de facto standards but formally totally informal [3-1, p.28].

"The IETF now has more than 75 working groups, each working on a different aspect of Internet engineering. Each of these working groups has a mailing list to discuss one or more draft documents under development. When consensus is reached on a draft document it may be distributed as an RFC" [3-1, p. 28].

This quote describes why the internet itself cannot be measured by standards of sophisticated legislation like e.g., on press affairs: Because it was never intended nor "organized" to host such a quantity and quality of load like it does today. If those few academics from different universities over the USA, all distinguished scientists, everyone would have imagined that hate speech battles by millions of users would be fought on social media, they would likely have chosen a fundamentally different design.

3.2.1. The Postal Union and the Treaty of Bern on International Postal Services – A different approach and regime

The Treaty of Bern was signed on 9 October 1874. The treaty intended to standardize postal services and regulations to exchange international mail freely. It is the basis of global postal services. The signatories of the treaty form one postal area, to exchange shipments. Besides the definition of terms, the conditions are written down like costs, general shipping terms and more. The Treaty of Bern only regulates general conditions of the different types of shipments but details about the constitution and the handling. Another part of the treaty is instructions of shipping limitations, regulations of the post exchange with countries that have not signed the treaty and more. At the very end a special regulation for some signatories states, that no signatory country is bound to deliver a sending to an area, where the shipper can ship postal items into another member country and profit from their lower fees [3-2].

Article three of the treaty of Bern states that every member county has to make sure that every user has access to universal postal service. This service has to be area-wide and affordable. Another important article is article 12: the member countries have to take care of the acceptance, processing, transport and delivery of letters. Therefore, there are specific rules that every member country has to deliver every letter from any member country.

In 1874 the Universal Postal Union was founded and since then they regulate the international cooperation of the postal services and the basic parameters of the cross-border sending and the upcoming costs for the delivery. The main task of the Universal Postal Union is to secure worldwide, timely delivery of letters and packages.[140]

To ensure the delivery of letters there are four different bodies at the Universal Postal Union. First the Congress. It's the supreme authority and gathers every four years. Diplomats from all 192 member countries are in Congress. They make decisions about the future of the postal sector; they also agree on new rules or policies for the international exchange. Another body is the Postal Operations Council, they are the technical and operational part. Its members are 40 member countries, which were elected by the last Congress. The main task of the Postal Operations Council is to help postal services around the world to modernize and upgrade postal products and services. This body also makes recommendations on standards for technological operational or further processes. The Council of Administration consists of 41 member countries and meets every year at the Universal Postal Union headquarter. This body has to ensure the continuity of the work between Congresses, conducts the activities and studies regulatory, administrative, legislative and legal issues. To be able to react in time to changes in the postal sector, the Council of Administration can approve proposals by the Postal Operations Council until the next Congress session. The promotion and coordination of all aspects of technical assistance among member countries is also a responsibility of the council. The last body, the International Bureau has a secretariat function. It supports the other bodies of the Universal Postal Union logistically and technically. It has taken a stronger leadership role in certain activities, like the application of postal technology through its Postal Technology Centre, the development of postal markets through potential growth areas such as direct mail and electronic mail services (EMS)[141], and the monitoring of the quality of service on a global scale. Through the Postal Technology Centre, the Universal Postal Union has entrenched some regional support centers all over the world to support information technology activities.[142]

---

[140] https://www.upu.int/en/Universal-Postal-Union (last accessed 17.12.2021)

[141] These services comprise, among others, sending electronic letters to a postal authority which prints it, puts it into an envelope and delivers it to the recipient. These services shall not be confused with e-Mail.

[142] https://www.upu.int/en/Universal-Postal-Union/About-UPU/Bodies (last accessed 17.12.2021).

Due to this treaty the signatories agreed to these standards and to establish an administration to control the postal services. Also, there is the Universal Postal Union with its bodies to ensure that agreements of the treaty are kept by the signatories. There are no rules like the Treaty of Bern on the internet, nor someone who can enforce any rules. So, there are no universally enforceable legal consequences on the internet.

3.2.2. IETF Recommendations[143]

Worldwide broadcasting and distribution of information is possible through the internet. It's also a medium for collaboration and communication between individuals [3-1, p. 22]. The internet can be divided into four historically developed aspects. First into the technological aspect which started with research on packet switching and the Advanced Research Projects Agency Network (ARPANET). The research also expands the current limits like scale and performance. The next is the social aspect, this formed a community of Internauts who work together to design and expand the current technology. The operations and management aspects are important for a global and complex operational infrastructure. And not to forget about the commercialization aspect where the research results and accessible information can be transitioned highly efficiently [3-1, p. 23].

Documentation is a very important part of the internet. It began with the constant growth of the internet through free and open access to basic documents like the specifications of the protocols. The academic traditions for open publication of ideas and results got promoted for the ARPANET and the Internet in the academic research community. But the usual way of an academic publication was too formal and slow to create a dynamic exchange. In 1969 S. Crocker established the Request for Comments (RFC) series of notes. The idea of the RFCs was to share ideas with other network researchers in an informal and fast way. In the beginning, the RFCs were printed documents, which were sent by snail mail. As soon as the File Transfer Protocol (FTP) was created and applied, the RFCs became online files and could be accessed through the network at many sites all around the world. RFCs should create a positive feedback loop through ideas or suggestions of another RFC with further ideas. The relevant ideas which belong together will be summed up in a specification document. Documents like this will then be used as a basis for the implementations of different research teams. Although the RFCs started as informational documents, they are more focused on protocol standards and "documents of record" now. The Internet is among other things constantly because of open access to the RFCs. Open access allows their usage for example in classes and for the development of new systems. Emails also changed the development of protocol specifications, technical standards, and Internet engineering. While at the beginning of RFCs the researchers from one place presented their ideas to the community, email enables joint authors with specific knowledge from all around the world to work together and come up with new and innovative ideas. Therefore, the IETF works with many mailing lists in each working group. In this mailing lists, the draft documents in progress can be discussed and improved. As soon as the working group has enough consensus researched the draft could be distributed as an RFC [3-1, p. 28].

Unlike a legal document, a draft of law or such an RFC is never voted on, never formally closed and hence its status can be described as "informal and permanently pending". The internet works because all actors design and develop their products like browsers, file services, etc. under strict observation of the RFCs on a voluntary basis.

---

[143] https://www.ietf.org/ (last accessed 25.10.2021)

The IETF (Internet Engineering Task Force) is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of Internet architecture and the smooth operation of the Internet. The IETF operates in working groups to do the technical work. These groups are specialized in different areas, these are managed by Area Directors (ADs). The groups are supported via many programs of the Internet Society, another community "governing" the internet. A lot of the work gets done through mailing lists. There are IETF meetings and events three times a year, like the IETF Hackathons, which show practical implementations of IETF standards. The IETF aims to make the Internet work better by producing high quality, relevant technical documents that affect how people use and manage the internet. To reach this goal, the IETF established the following principles.[144]

- Open process: because of this principle the documents, the WG mailing lists, or attendance lists and the Meeting minutes of the IETF are publicly available on the internet. Therefore, any interested person can participate, inform themselves and make his or her voice heard.

- Technical competence: the IETF is willing to listen to technically competent input from any source. The IETF expects its output to be developed after strong network engineering principles, also called "engineering quality".

- Volunteer Core: people who work for the IETF want to make the internet work better.

- Rough consensus and running code: the standards are developed based on engineering judgement and real-world experience in applying their specifications.

- Protocol ownership: the IETF takes ownership of some protocols, so it accepts full responsibility, even if some aspects may not be seen on the internet. But if the IETF does not take the responsibility for a protocol it does not try to get control over it, even if it touches the Internet.

Note that the IETF is not responsible for the internet nor owns it. It is of limited legal nature, namely subpoenas and similar legal papers can be served under US law. Requests for authenticated documents and other information directly from the IETF may be made either informally or formally through a third-party subpoena.[145]

The most important documents from the IETF are the above-mentioned RFCs. There are more than 9,000 individually-numbered documents in the series. The RFCs address many aspects of computer networking, like technical foundations of the internet, addressing, routing and transport technologies. The RFCs furthermore specify protocols that are for services used by billions of people daily, like real-time collaboration, email and the domain name system.

RFCs may have different statuses, depending on their level and what they cover: Internet Standard, Proposed Standard, Best Current Practice, Experimental, Informational, and Historic.

RFCs start as Internet-Drafts (I-Ds). These are normally going to be improved and revised in different working groups. When RFCs get published, they are freely available. The described technical specifications are implemented and adopted voluntarily by software developers, hardware manufacturers, and network operators from around the world.

---

[144] https://www.ietf.org/about/mission/ (last accessed 24.01.2022)

[145] See https://www.ietf.org/about/administration/legal-request-procedures/ for the procedure (last accessed 24.1.2022).

Please note again that these documents are not binding and, if not obeyed, do not have legal consequences, but if everyone on the internet acts as they suggest, the internet is working better.
So RFCs are more like a recommendation for the users on what to do or not to do. The whole internet can be compared to a highway without any police or roadside assistance services. There are no binding rules which have to be followed because they can't be enforced or controlled. The RFCs recommend not to hack somebody but it's done despite this, simply because it's possible.[146]

"As the current rapid expansion of the Internet is fueled by the realization of its capability to promote information sharing, we should understand that the network's first role in information sharing was sharing the information about its own design and operation through the RFC documents. This unique method for evolving new capabilities in the network will continue to be critical to the future evolution of the Internet.". [3-1, p. 27]
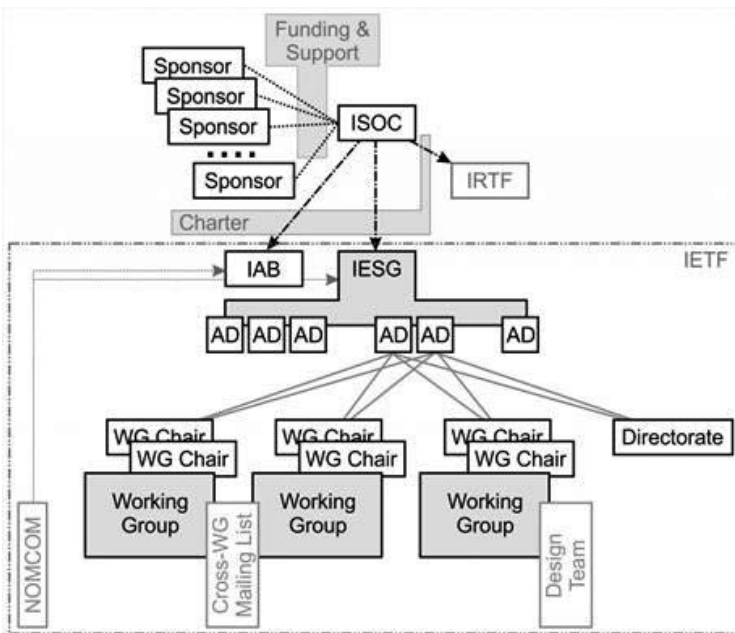

Figure 26: Organizational structure within IETF[147]

3.3.   A brief introduction to how the internet works

3.3.1. Global Internet accessibility

Today, internet access is available in almost every country. Around 51 percent of the world's population is expected to have internet access as of 2019. Only the access rates may vary by country and region, depending on the local infrastructure. Even in areas that currently do not have the infrastructure to access the internet, there will most likely be the possibility to get a satellite connection in the near future. The technical availability of access to the internet also directly leads to steady growth of active users of social media, with an estimated 4.2 billion active users as of January 2021.[148]
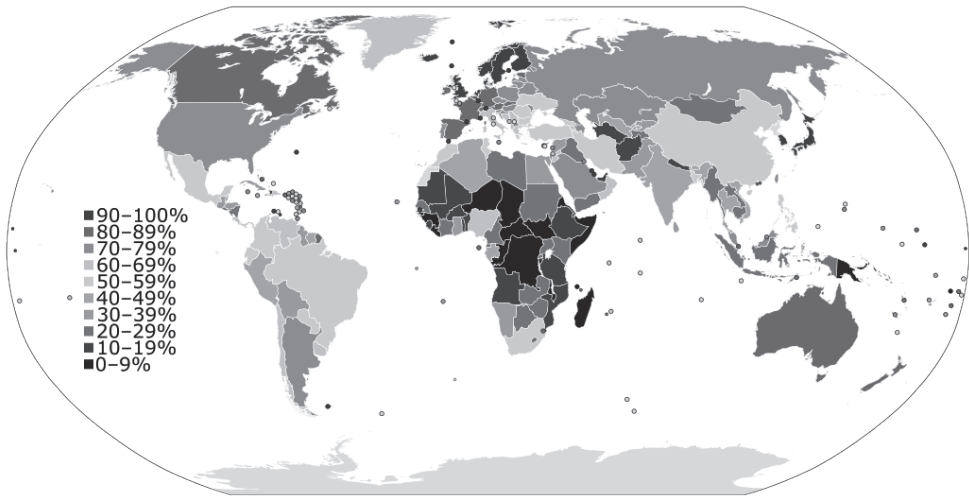
Figure 27: A world map colored to show the level of Internet penetration, by Jeff Ogden is licensed under CC-BY-SA 3.0[149]

3.3.2. Network Architecture Types

The internet itself is not one certain network but rather a collection of different networks and systems which are interconnected. In the early days of computing, the so-called mainframe architecture was one type of simple, local network. Multiple terminals were connected to a single mainframe system, that did all the calculations for all connected clients. In modern days it comes down to two major types of network architectures: Client-Server and Peer-to-Peer.

Most services on the internet (e.g., websites with social media functionality) are based on the client-server concept. On the client side, usually, the computer or smartphone of a user mainly shows data provided by a server. None of this data is stored permanently locally on the client. Instead, all the resources, information and data are stored and processed on the server. The owner of the server therefore technically owns the entire data stored on the server and may change or cancel the service he provides at any time. It is also possible to provide different versions or alternatives of data and services to different clients (e.g., for different countries or regions) ultimately leading to the possibility to create

---

[148] https://www.statista.com/statistics/617136/digital-population-worldwide/ (last accessed 09.01.2022)
[149] https://commons.wikimedia.org/wiki/File:InternetPenetrationWorldMap.svg (last accessed 08.02.2022).

different realities for each client. Those modifications to data by the server can only be noticed by users if they use multiple clients with different versions of data provided and do a side-by-side comparison. Despite the hierarchic composition, many sublayers of the ISO/OSI model, and connections between subnets are possible, which makes their analysis problematic [3-6, p. 52].

Another and less frequently used network architecture, is the so-called Peer-to-Peer. All clients work together on the same protocol level and each peer works in the network both as a server and as a client. Without one central server that provides certain services, it is not possible to simply shut down a Peer-to-Peer network via a single point. Additionally, it is not a single person alone responsible for the network. It is also not easy to provide certain functionalities like commenting on content which is a very important function in social media. However, Peer-to-Peer networks are perfectly fit for sharing large amounts of data (e.g., videos, music and archives). With the data being available redundant over multiple peers (users), it is almost impossible to stop the spread of peer-to-peer shared data.

Another reason why sharing platforms are rather peer-to-peer-organized is quite simple: If user A shares directly with user B and more clients, then there is no need for an actual platform hosting the content. Consequently, no one has neither a say nor any legal liability. The most used peer-to-peer network is BitTorrent.[150]

3.3.3. The IP-Protocol

A uniform language is required to communicate between client and server or between peers. For these networks, this is the TCP/IP (Transmission Control Protocol/Internet Protocol) protocol family. Due to the setup of the Internet, instead of circuit-switching (direct, continuous communication), it uses packet-switching (sending smaller amounts of data, slicing the communication to fixed-sized packs). Large amounts of data such as videos and images are packed into small data packets and transmitted over the network. The receiver of the data packets reassembles them and processes the data and then displays the requested image in the client's browser, for example.

Similar to the address in the postal system, each participant in the network (whether client, server or peer) has its IP address via which it can be reached and is needed to send and receive the data packets defined by the internet protocol. An IP address is unique within each network and assigned only once at a time but might be reassigned any time to a different network device.

The assignment of an IP address to the client is performed by an Internet Service Provider. Internet Service Providers are usually private companies providing the network architecture for internet access to their customers. If the user does not log in with his data when using an Internet service, only the Internet service provider can say for sure which customer is behind which IP address since only the provider has the billing information of the user. Often, the Internet service provider does not permanently assign an IP address to a customer but instead assigns a new IP address from the Internet service provider's address pool to the customer after some time. If the Internet Access Provider does store the information which customer had which IP address at a certain time, it is possible to trace back certain actions on the internet (e.g., postings in social networks) to one exact customer or person.

However, only the owner of the connection is accessible in this way. In practice, several users are often connected within a local network and use the Internet connection offered by the Internet service provider. This works via so-called Networks Access Translation (NAT). If a computer in the local

---

[150] https://www.bittorrent.org/introduction.html (last accessed 24.01.2022)

network initiates a connection to a computer on the Internet, the data packets with the request are first transmitted to the router of the local network. This router performs the address translation of the sender address, i.e., exchanges the address of the internal computer with its IP address that gets routed through the internet, then transmits the request. The router thus presents itself to the Internet as the sender of the request. At the same time, the initiation of this network traffic is dynamically stored in a NAT translation table in the router to process the response from the Internet. When the response arrives, the table entry is used to determine the original initiator, the client in the local network. The computer in the local network now receives the data packets from the router and can process them.
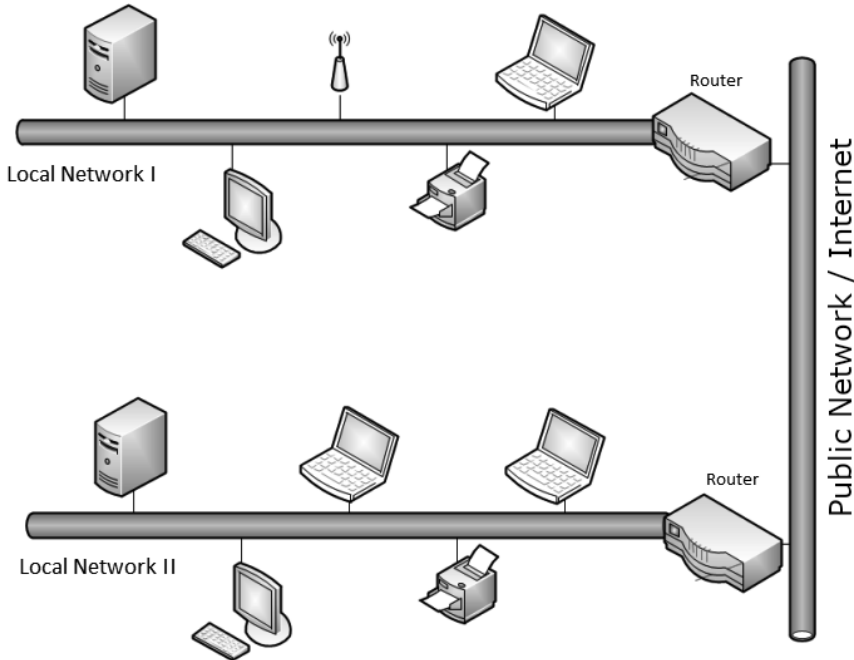
Figure 28: Structure of the internet, Alexander Prosser (2013)[151]

To assign data to individual clients on the local network, each client is assigned a local IP address. For this purpose, certain IP address ranges are not routed via the Internet but are only accessible within the local network.

Each network device has a unique MAC address (Media Access Control). The MAC address is often referred to as the physical address. It is assigned by the manufacturer of the network device and is unique in theory, but for many devices, it can be changed by the user through software manipulation. The router of the local network uses the MAC address of the network devices to assign them a local IP address, which is then used to handle the data traffic. If a user is located within a local network, the exact assignment to a person is only possible with the cooperation of the operator of the local network and only if the user does not regularly change the MAC address, e.g., of his laptop.[152]

---

[151] https://www.wu.ac.at/fileadmin/wu/o/evoting/Folien/LLM2013_01.pdf (last accessed 24.01.2022).
[152] https://standards.ieee.org/products-services/regauth/oui36/index.html (last accessed 24.01.2022.)

3.3.4. IP-Address assignment

An Internet service provider does not arbitrarily assign IP addresses to its customers. Instead, IP addresses are assigned hierarchically. The highest authority is the Internet Assigned Numbers Authority (IANA, iana.org), which controls the entire IP address space. It allocates IP address blocks to regional IP address allocators, the so-called Regional Internet Registries (RIR), which are responsible for IP address allocation in certain continents and regions. Those are:

- African Network Information Centre (AfriNIC) for the African continent.[153]

- Asia-Pacific Network Information Centre (APNIC) for the Asia-Pacific region.[154]

- American Registry for Internet Numbers (ARIN) for North America, North Atlantic and some Latin America & Caribbean Network Information Centre (LACNIC) for Latin America and the Caribbean islands.[155]

- Latin America & Caribbean Network Information Centre (LACNIC) for Latin America and the Caribbean[156]

- Réseaux IP Européens Network Coordination Centre (RIPE NCC) for Europe, Central Asia and the Middle East[157]
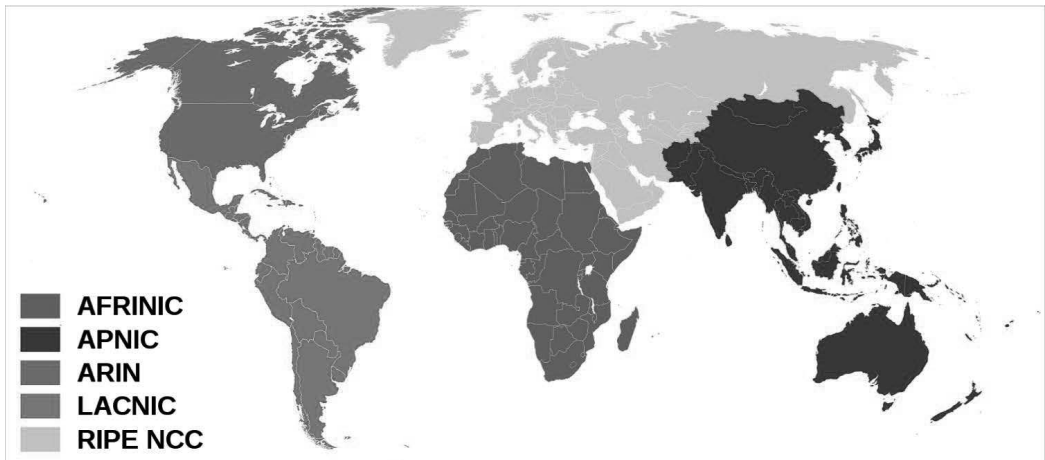
Figure 29: Regional Internet Registries world map, by Dork, Canuckguy, Sémhur is licensed under CC-BY-SA 3.0[158]

All Council of Europe member states are within the territorial jurisdiction of the RIPE NCC. Qualified Internet providers in the relevant regions can now apply to the Regional Internet Registries

---

[153] https://afrinic.net/about (last accessed 09.01.2022)
[154] https://www.apnic.net/about-apnic/organization/ (last accessed 09.01.2022)
[155] https://www.arin.net/about/welcome/region/ (last accessed 09.01.2022)
[156] https://www.lacnic.net/631/2/lacnic/coverage-area (last accessed 09.01.2022)
[157] https://www.ripe.net/about-us/what-we-do/ripe-ncc-service-region (last accessed 09.01.2022)
[158] https://de.wikipedia.org/wiki/Regional_Internet_Registry#/media/Datei:Regional_Internet_Registries_world_map.svg (last accessed 08.02.2022).

for their own IP address blocks, provided they meet the technical and administrative requirements and can demonstrate a need. Internet service providers are entitled to manage the address range allocated to them largely autonomously, but the allocation of IP addresses by these authorities to other providers and customers is nevertheless subject to extensive conditions: The request for necessary IP addresses must be precisely planned and justified, and the provider must ensure that accurate records are kept of the internal transfer of IP addresses. These two conditions in particular generate an enormous amount of administrative work for providers, even though the actual allocation of IP addresses by the authorities is traditionally free of charge.

Nevertheless, this documentation of IP address allocation is important so that it can be determined at any time to which provider or user a particular address block is assigned and so that future demand can be estimated.

Note that the IANA is also of a very limited legal nature, it is headquartered in California/USA and acts as a coordinator to make sure that neither an IP address nor a domain name is used more than once. For the concrete domains and addresses, the individual local authorities like the DENIC in Germany who controls the .de top-level domain are responsible. Of course, each of these individual authorities is liable to local law and law enforcement, e.g., the DENIC to German law only.

3.3.5. Tracing an IP-Address

Since the allocation of IP addresses to customers by the Internet service provider must be documented at the responsible regional Internet registry, the IP address of a homepage, for example, indicates the country in which it is hosted. The query is possible for everyone through IP tracking services, as these simply access the WHOIS services of the Regional Internet Registries.

Basic Tracking Info	
IP Address:	193.196.151.202
Hostname:	www.hs-ludwigsburg.de
Internet Protocol:	IPv4 - Version 4
Types:	Public
IP Classes:	Class C Range (192.0.0.0 to 223.255.255.255)
Reverse DNS:	202.151.196.193.in-addr.arpa
Blacklist Check:	Not Blacklisted (Clean) [193.196.151.202 Blacklist Check]
NS (Nameservers):	dns1.belwue.de >> 129.143.2.10
	dns3.belwue.de >> 129.143.253.133
	dns5.belwue.de >> 129.143.4.5
Location Details	
Continent:	Europe (EU)
Country:	Germany (DE)
Capital:	Berlin
State:	Baden-Wurttemberg
City:	Ludwigsburg
Postal:	71642
ISP:	Universitaet Stuttgart

Figure 30: Own image, Information from ip-tracker.org.
The IP-address shown provides the website of the „Hochschule für öffentliche Verwaltung und Finanzen Ludwigsburg".

Based on the IP address of a client, the operators of a website can determine, for example, from which country and region the client is accessing the website. In this way, services are partially tailored to the respective user and relevant information is displayed first and foremost. Localization accurate to within a few meters based on the user's IP address is not possible. However, region or state can be determined with around 80 percent certainty.[159]

---

[159] https://www.if-so.com/geo-targeting/ (last accessed 08.01.2022)
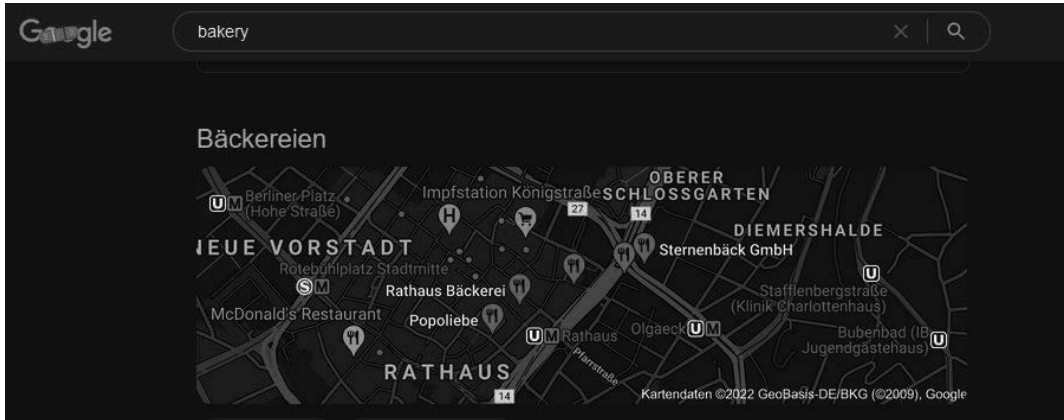
Figure 31: Own image, Google search for a bakery.
Except for the keyword, no other information was passed. Only the IP address was used for geolocation. The map extract shows a part of the city center of Stuttgart, Baden-Württemberg. The user's actual location is only about five kilometers away on the outskirts of the city.

3.3.6. Evade IP tracking

The easiest and fastest way to disguise one's own IP address and thus location when using Internet services is to use proxies. A proxy is a kind of intermediary between the client or user and an Internet resource, such as a website. If the proxy is appropriately configured not to forward the user's IP address to the website, but replaces it with its own and receives the website's data packets vicariously and forwards them to the user, the user's IP address is effectively disguised. The user needs to know that the data transmitted via a proxy can not only be read, stored, and evaluated by the proxy, but can even be manipulated. In addition, the loading time of websites is usually noticeably increased, since many proxy servers are used not only by one, but by hundreds of users, and all data packets of the website must take the "detour" via the proxy server.[160,161]

A proxy can also be used to filter data packets. Many companies use proxies to protect their employees' computers and other clients from dangerous websites. Web filters for pornographic, violent or similar content can also be set up for employees in this way. Another security aspect is the restriction of unwanted remote access to the client which goes beyond the response packets, since the contacted target system from the public network does not send its response packets directly to the client, but sends them to the proxy, which can actively control the connection.[162] End-to-end secure encryption between the website and the client (see 2.7) is essential when using a proxy to ensure the security and integrity of your data.

3.3.7. Website encryption and trust

The monitoring of network traffic by a wide variety of countries, institutions, companies, and others (see Section 3) creates the need for widely deployed, secure, encryption of websites and online services.

---

[160] https://www.ionos.de/digitalguide/server/knowhow/was-ist-ein-reverse-proxy/ (last accessed 09.01.2022)
[161] https://it-service.network/it-lexikon/proxy (last accessed 09.01.2022)
[162] https://tarnkappe.info/tarnkappe-guide-was-ist-ein-proxy/ (last accessed 09.01.2022)

In addition to personal data on social networks, the security of the ever-growing use of online banking is particularly at risk. Almost two-thirds of citizens in the euro area already use online banking.[163]

The Transport Layer Security (TLS) protocol, often also known by its predecessor name as Secure Sockets Layer (SSL), ensures encrypted transmission of data on the Internet. It is a hybrid encryption protocol that combines asymmetric and symmetric encryption (see paragraph 4.2). The encryption is intended to protect the transmitted data from unauthorized access by third parties and manipulation or forgery. In addition, TLS enables authentication of the communication participants and verification of identities of receiver or sender. TLS is often used for secure connections between a client with an Internet browser and a web server via HTTPS. But other protocols such as SMTP (Simple Mail Transfer Protocol), POP3 (Post Office Protocol) or FTP (File Transfer Protocol) can also use Transport Layer Security. Communication via TLS can be divided into two phases. First, a connection is established in which the client and server prove their identity to each other. Once a trusted connection has been established, the data is transferred using an encryption algorithm.

The so-called Transport Layer Security Record Protocol plays a central role in Transport Layer Security. Four other protocols of the standard build on this. These four protocols are:

- the Handshake Protocol

- the Alert Protocol

- the Change Cipher Spec Protocol

- the Application Data Protocol

The Handshake Protocol is responsible for negotiating a session and its security parameters. Among other things, the Handshake Protocol negotiates the cryptographic algorithms and key material used and authenticates the communication partners. The Alert Protocol is responsible for the error and alarm handling of TLS connections. It can initiate the immediate termination of a connection. The Application Data Protocol is used to split application data into blocks, compress, encrypt and transmit them. Finally, the Change Cipher Spec Protocol informs the receiver that the sender is changing to the cipher suite previously negotiated in the Handshake Protocol.

When a client establishes a connection to a server, the server authenticates itself with a certificate (see 2.7.1). The client verifies the trustworthiness of the certificate and that it matches the server name. Optionally, the client can authenticate itself to the server. In the next step, the communication partners derive a cryptographic session key with the help of the server's public key, which they then use to encrypt all messages to be transmitted. The authentication and identification of the communication partners are thus based on asymmetric encryption methods and public-key cryptography. The actual session key is a one-time-use symmetric key that is used to both decrypt and encrypt the data. Besides the TLS version, the exact used symmetric and asymmetric cryptographic algorithms, and also protocol settings determine the resultant security level [3-7, p. 128].

---

[163] https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_bde15cbc&lang=en (last accessed 09.01.2022)
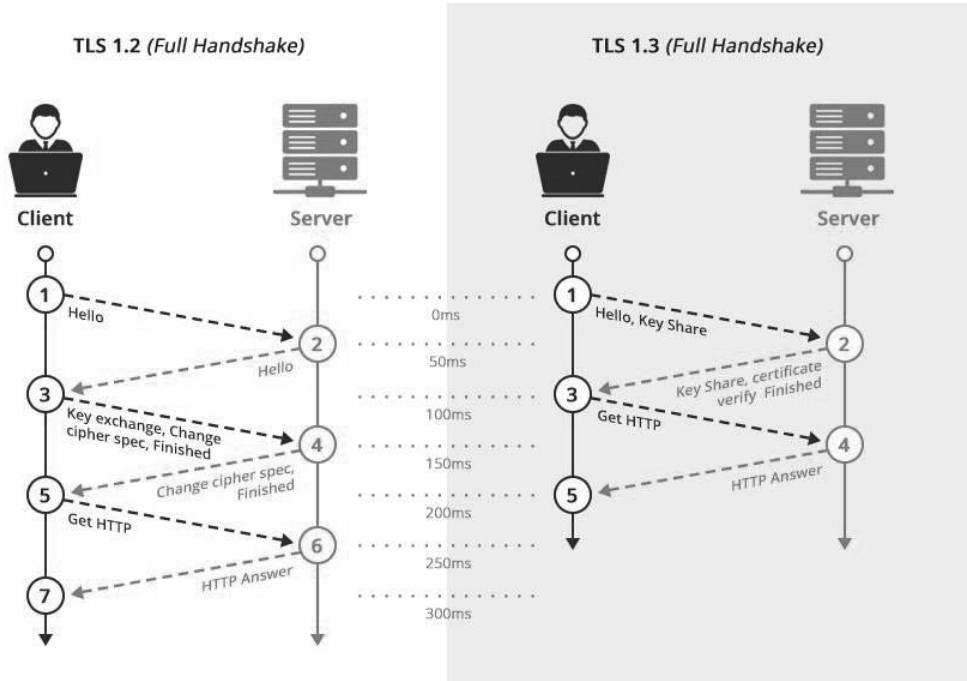
Figure 32: TLS 1.2 and 1.3 comparison by SSL2Buy.com[164]

3.3.7.1. Trustworthy certification authorities

The digital certificate is an electronic proof of authenticity issued by a certification authority (CA). On the Internet, certificates have the comparable function of an ID card in the offline world. With the help of a certificate, a public key can be securely assigned to a specific owner. The contents of the certificate include information about the name of the owner and the issuer of the certificate as well as about the validity period and the use of the certificate.

Together with the public key infrastructure (PKI), certificates enable information to be transmitted securely and encrypted on the Internet. The encryption is based on asymmetric cryptographic processes with private and public keys. The certificate reliably confirms to whom the public key belongs. Browsers and operating systems keep a list of trusted certification authorities. If a certificate is issued by such a certification authority, the computer considers it to be genuine.

The X.509 standard specifies what content must be included in a certificate and what form. Some information is mandatory others are optional. X.509 certificates are used, for example, to encrypt websites using the HTTPS protocol or to sign and encrypt e-mails using the S/MIME standard.

Important information in an X.509 certificate includes:[165]

- the version number

- the serial number

- the algorithms used to create it

- the name of the issuer

- the name of the holder

- the validity period

- information about the public key of the holder

- information about the intended use of the certificate

- the digital signature of the Certification Authority

Certification authorities (CA) or trust centers play an important role in the public key infrastructure and certificates. They check the details and identity of an applicant for a certificate and issue it if the details are correct. They can also take care of publishing the certificates and storing them in public directories. Other tasks of the CA include managing and publishing certificate revocation lists and recording all certification activities of the Certification Authority.

Figure 33: Own image, Screenshot of an SSL test by qualys[166] for the domain hoenig.online.
The shown certificate was issued by Let's encrypt, a certificate authority provided by Mozilla, EFF and others.[167]

3.3.8. Domains and the Domain Name System

On the internet, data is always transferred between devices addressed by so-called IP addresses, which are the actual address instances of the internet protocol (for which the abbreviation IP stands). However, numbers are more difficult for people to remember than letters put together in a meaningful way, which is why a naming concept was devised for the internet that was based on the existing internet protocol, the Domain Name System (DNS).

The DNS has a hierarchical structure in order not to have to manage and process all DNS queries centrally, which would simply fail due to the number of domains and queries, but to be able to provide the DNS information in a distributed and heavily redundant manner. The so-called root servers form the top hierarchy. They contain information about the name servers that are responsible for the top-level domains on the Internet.[168]

The next hierarchy level is represented by the name servers of the top-level domains. Each of these top-level domains has its own name servers on the Internet, which contain information about the

---

[166] https://www.ssllabs.com/ssltest/ (last accessed 09.01.2022)
[167] https://letsencrypt.org/about/ (last accessed 09.01.2022)
[168] https://www.cloudflare.com/learning/dns/what-is-dns/ (last accessed 21.01.2022)

domains registered under the respective top-level domain. These name servers are administered by registries through which domain names can be registered within the respective top-level domain. As of January 2022, there are more than 2500 by ICANN officially accredited registrars over the world.[169] For example, GoDaddy, LLC is located in the US under US law or IONOS SE is located in Germany under EU and German law.

$$ \text{https://www.coe.int} $$

Third Level   Top Level

Second Level

Figure 34: Domain levels

If we read this address from right to left, we first have the top-level domain "online", the first visible DNS hierarchy level. From a technical point of view, this means that the domain name to the left was registered below the top-level domain "online". The second part, i.e., "coe", is the so-called second-level domain, i.e., the second visible DNS hierarchy level. In the case of the top-level domain "int", domain names can be registered directly under the top-level domain, which is not possible for all top-level domains. The domain "int" is also the only domain that is administered exclusively by IANA itself.[170] Within some other top-level domains, there is another level of hierarchy that identifies the category. For example, in the United Kingdom, domain names cannot be registered directly under the national top-level domain "uk", but there are further levels of hierarchy, such as "co.uk" for commercial addresses or "ne.uk" for network-specific addresses. If there is such a further hierarchy level, this second hierarchy level is the second-level domain (for example, "co" in "co.uk") and the actual domain name is the third-level domain, i.e., the third hierarchy level. Everything that now follows on the left after the actual domain name (in the example, "www" after "coe.int") is the responsibility of the person who registered the domain name and can be used for individual computers. In the zone file for "coe.int", therefore, an entry is created for the name "www" in the form which ensures that when "www.coe.int" is requested, the response is the IP address of the webserver on which the homepage of the Council of Europe is located.

In addition to the purely technical data, administrative information is also required during registration, which makes the ownership of the domain name clear. This necessary information is divided into the description and the contacts.

The description usually contains information about the owner of the domain and possibly other administrative information that is required during registration.

The contacts are divided into the different areas of responsibility that exist in the administration of a domain name:

- administrative contact ("admin-c")

The administrative contact is the official owner of the domain name and the general contact for questions regarding the domain name or all entries under it.

- technical contact ("tech-c")

The technical contact is responsible for the technical handling of the domain name. Usually, a person from the IT department of the company concerned or the Internet provider is indicated here.

- billing contact ("billing-c")

The billing contact is found at registries of top-level domains, where a registered domain name must be paid directly to the registry. Usually, a person from the accounting department of the company concerned or also the Internet provider is specified here.

- zone contact ("zone-c")

The zone contact is required for some top-level domains and specifies a person who is responsible for entries within the zone file of the domain name. Usually, this is also a person from the IT department of the company concerned or the Internet provider.

Note that all these contacts are normally email addresses only. A physical address, where writs and civil papers can be served, is neither mandatory nor common.

3.3.8.1. Name resolution

Name servers have two tasks in the DNS: On the one hand, they can be authoritative for certain domains and hold zone information on the Internet, but on the other hand, they are also needed for name resolution. Domain names must be resolved if, for example, a user has entered a domain name in his web browser and the browser first needs the IP address of the target computer to contact it.

Name resolution on the Internet is also hierarchical:

In the introductory step of a name resolution, a client that needs the IP address of a certain resource on the Internet requests the name server responsible for it. The user usually does not need to worry about the address of the name server, since this data is usually supplied with the access parameters of the Internet access. If the name server has already recently performed a corresponding name resolution for the same domain name, it will immediately return the searched IP address. Otherwise, it will perform the following steps to be able to provide an answer.

The name server will resolve the domain name from right to left. So first it will determine which name server is responsible for the top-level domain "int". The root servers are responsible for the information about the top-level domains, so the name server will contact a corresponding root server and ask if it knows which IP address "www.coe.int" points to. This is not authoritatively responsible for "www.coe.int" and will answer him, according to the DNS hierarchy, with the address of the name server for the top-level domain "int", which can give more detailed information.

In the next step, the name server will contact the name server of the top-level domain "org" to obtain the information it is looking for. It will also send the same request to this server as to the root server. In this case, the name server for the top-level domain "int" is already authoritatively responsible for

"coe.int" and will respond with the addresses of the name servers that are responsible for the domain. Normally, for other domains, these are the nameservers of the Internet provider in the next hierarchy level with which the domain is registered.

After the request to the name server responsible for the domain "coe.int", this looks in its zone file to see which IP address was registered and returns it.

After the name server has now determined the IP address of the desired query, it passes this on to the client. At the same time, the name server stores the result of this query in its cache for a certain period to be able to provide the answer immediately in the event of a possibly identical query.[171, 172]
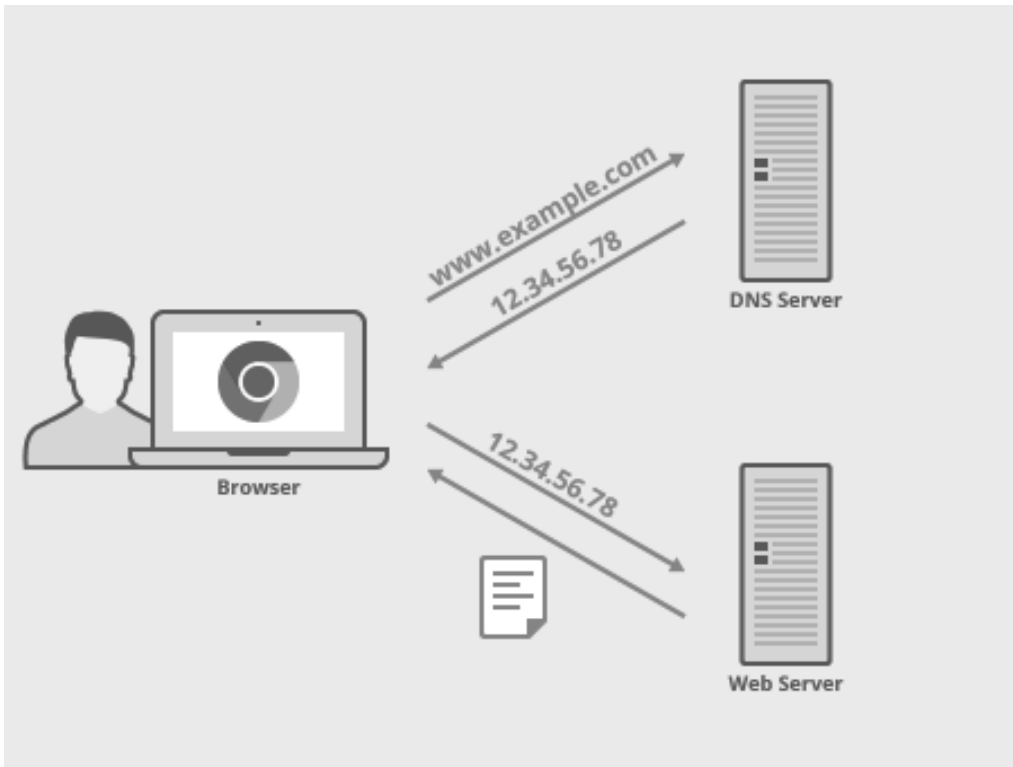
Figure 35: DNS-Server, by Seobility is licensed under CC-BY-SA 4.0[173]

3.3.8.2. Domain Blocking and Censoring

There are recurring demands to block certain websites with content that is prohibited by law in the respective country. These are often implemented with DNS blocks, as the website is hosted on servers abroad and the operators are not tangible. The name servers, which are operated by the customer's Internet service provider in the customer's own country, must then be configured accordingly so that

---

[171] https://www.techtarget.com/searchnetworking/definition/domain-name-system (last accessed 21.01.2022)
[172] https://www.seobility.net/de/wiki/DNS-Server (last accessed 21.01.2022)
[173] https://www.seobility.net/de/wiki/DNS-Server (last accessed 08.02.2022)

certain websites are not resolved. The customer will then not know the corresponding IP address of the website and will not be able to access it.[174,175]

However, this only affects the name servers in the customer's own country. The name servers that are located in the corresponding country of the website operator will continue to resolve the IP address of the website that is known to them.

If a customer is affected by the DNS blocking of his Internet service provider, he only has to adjust his configuration in such a way that a DNS server abroad is addressed for the resolution of a website domain. The time required for such a configuration change is limited to a few minutes.

Figure 36: Own image, DNS server settings of the author.
Instead of the "recommended" DNS servers of the Internet service provider, which would implement DNS blocks of the legislators in Germany, the DNS servers of Quad9, based in Switzerland[176], and Google, based in the United States[177], are addressed. Both outside the German legal jurisdiction.

3.4. Surveillance of network traffic

Data traffic on the Internet is continuously monitored by various, mostly governmental, agencies. The best-known example is probably the surveillance programs of the U.S. intelligence services, the nature and scope of which Edward Snowden reported to the Council of Europe in 2014.[178]

---

[174] https://tarnkappe.info/vodafone-muss-library-genesis-sperren/ (last accessed 21.01.2022)
[175] https://tarnkappe.info/boerse-to-teilweise-von-vodafone-gesperrt/ (last accessed 21.01.2022)
[176] https://www.quad9.net/about (last accessed 21.01.2022)
[177] https://developers.google.com/speed/public-dns (last accessed 21.01.2022)
[178] https://pace.coe.int/en/news/4960 (last accessed 25.01.2022)

The United States is mentioned first here because it enacted the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act (USA Patriot Act) in response to the terrorist attacks of September 11, 2001. It provides U.S. law enforcement and intelligence agencies with extensive investigative, intercept, and surveillance capabilities aimed at deterring foreign terrorists and detecting and apprehending those in the country [3-3]. Part of the Patriot Act is the massive simplification for the issuance of a "National Security Letter". [3-4, p. 448] This enables law enforcement agencies to query personal data of users from banks, telecommunications providers or financial service providers, among others, without the need for judicial review of the order. The form in which intelligence agencies within the United States and around the world monitor telecommunications was revealed to the world in early summer 2013. Around 1.5 million previously secret documents were copied by whistleblower Edward Snowden, who had previously been an employee of various companies working for the National Security Agency intelligence service for 4 years, and made available to the press and thus to the world public. The impression was quickly created that the intelligence services of the United States were accessing, storing, and processing electronic data on a massive and warrantless scale [3-5, p. 36].

One of the first major surveillance programs of the U.S. intelligence services is "PRISM." With PRISM data was collected from U.S. companies Microsoft, Yahoo, Google, Facebook, PalTalk, AOL, Skype, YouTube, and Apple, including emails, chats, photos, telephone and video conferences, and even login data. While it is denied by the companies involved, it can be assumed that the National Security Agency has direct access to the servers of the affected U.S. companies. Depending on the exit interface through which the National Security Agency receives its data, real-time access to e-mails or chats is even possible.[179]

Whether this type of surveillance is necessary and justified by the threat of terrorist attacks does not need to be answered here. Other countries also have systems for monitoring the Internet traffic of the respective countries, be it at network nodes in Russia[180] or also at DE-CIX in Germany[181], one of the largest network nodes worldwide.

3.4.1. Implementing backdoors and weakening encryption

In addition to directly accessing unencrypted connections, a popular way for intelligence agencies is to deliberately implement vulnerabilities or backdoors in encryption software. Security solutions involving intelligence agencies have already been marketed with a backdoor or weakened encryption to eavesdrop on supposedly secure communications.[182]

However, the danger of weakened encryption algorithms or backdoors in network interfaces must not be underestimated under any circumstances. It can never be guaranteed that these access options to supposedly secure means of communication will not be abused, whether by governments of different countries or by criminal hacker groups that could misuse these options for their purposes. For example, the access data for a backdoor in the network software of a major manufacturer of

---

[179] https://www.washingtonpost.com/wp-srv/special/politics/prism-collectiondocuments/ (last accessed 25.01.2022)
[180] https://www.faz.net/aktuell/politik/ausland/russland-internet-wird-ab-jetzt-vom-staat-kontrolliert-16462733.html (last accessed 25.01.2022)
[181] https://netzpolitik.org/2015/klaus-landefeld-de-cix/ (last accessed 25.01.2022)
[182] https://www.zdf.de/nachrichten/politik/cryptoleaks-bnd-cia-operation-rubikon-100.html (last accessed 25.01.2022)

networking hardware was used by unknown third parties. With this login data, it was possible to monitor and read supposedly secure connections, such as VPN (see 4.2), in real-time.[183]

Note that the internet surveillance performed by US agencies is far more transparent than the surveillance probably and highly likely performed by agencies of other nations, who lack e.g. a Congress with public hearings, a Freedom of Information act and independent courts as well as civil society with powerful entities like the American Civil Liberties Union (ACLU).

We may assume that the whole internet is subject to far more surveillance, let aside the fact that big players like Facebook, Twitter, Google etc. are subject to US laws and jurisdiction and also to interventions from numerous US agencies.

3.5. Cryptographic basics and ways to remain anonymous in the net

For data being sent between two internet nodes via several unknown internet nodes where every single node can read – and probably alter – the data, the necessity arises, to encrypt the data sent. Sensitive data like credit card details or simple love letters do not need to be read by an unknown intermediary. The following chapter describes how basic cryptographical techniques work and how one of the most used tools, a so-called Virtual Private Network (VPN) works.

3.5.1. Basic cryptography

The usage of cryptography can be traced back to Gaius Iulius Caesar and his military campaigns in nowadays France and Belgium. He used a very simple but to that time a very effective way of cryptography. He put instead of the letter 'a' the letter 'c' and incremented the alphabet twice to encrypt his message (from the original word 'apple' to 'crrng'). The decryption was also very easy for those who know the code. They decremented the alphabet twice backward (from the original word 'crrng' to 'apple') and had now the message from the original sender. The enemies on the way can't read the message if they don't get the encryption. [184]

Today's methods of encrypting messages are a lot more secure than the "old" one described above. The process to break encryption today by a brute force attack (a method who the hacker tries out all combinations that are possible) can take up to 7.5 million years. [185]

Encryption today is divided into two main parts. The so-called "symmetric encryption" and the "asymmetric encryption". These have the same principle in encrypting the message with a key, but the accessibility of the key is handled differently.

3.5.1.1. Symmetric encryption

The encryption in the symmetric part works with an encryption key which is used to encrypt the message (plain text). Because of the encryption by the key, the text is now encrypted and for people who don't have the key to decrypt it is illegible (cypher text). To decrypt the message from the cypher text the receiver uses the same key as the sender. The problem is: How does the sender send this key to the receiver without any risk of detection?

---

[183] https://www.rapid7.com/blog/post/2015/12/20/cve-2015-7755-juniper-screenos-authentication-backdoor/ (last accessed 25.01.2022)
[184] https://www.wu.ac.at/fileadmin/wu/o/evoting/Folien/LLM2013_02.pdf (last accessed10.12.2021)
[185] https://www.password-depot.de/know-how/brute-force-angriffe.htm (last accessed 13.12.2021)
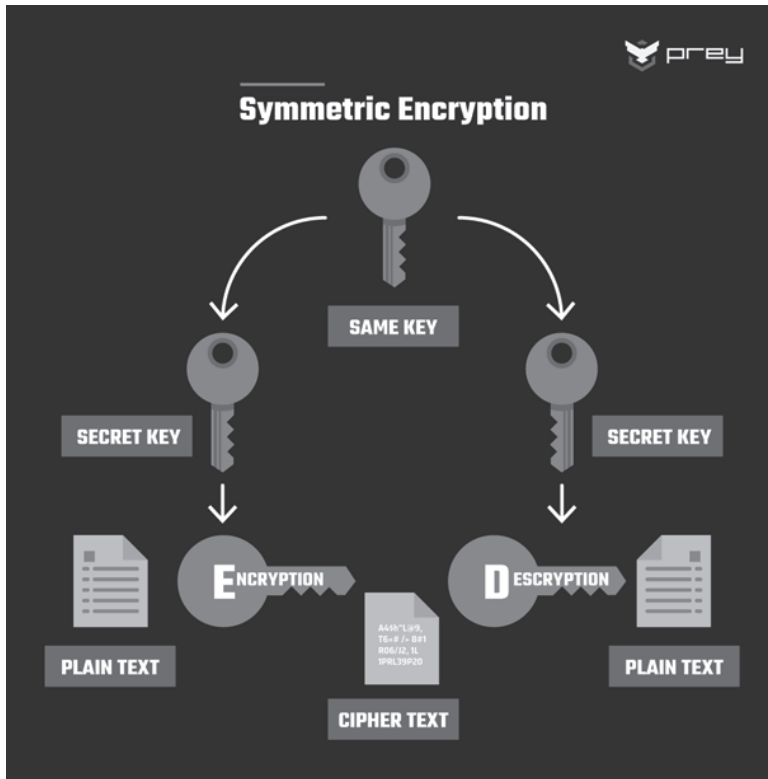
Figure 37: Symmetric encryption[186]

This is the only weakness of symmetric encryption. Some encryption programs solve the issue by assigning the two users that are communicating the keyway before the message is sent. Some others do it in an analog way by saving the key on an e.g., USB drive and deleting the key after the key file is implemented in the receiver's program.

In the following encryption method, this problem is non-existent because the keys to encrypt and decrypt are not the same.

3.5.1.2. Asymmetric encryption

The encryption in the asymmetric part works with a public encryption key which is used to encrypt the message (plain text). This key can be shared with anyone. It is, as its name says, accessible by the public. Because of the encryption by the key, the text is now encrypted and for people who don't have the key to decrypt it is illegible (cypher text). To decrypt the message from the cypher text the receiver uses a private key that only the receiver knows and owns. Only with that key, the message can be decrypted. From public key can't be inferred to the public key in any way. Now the receiver has the message with no risk that the private key is reviled in the transaction.

---

[186] https://preyproject.com/uploads/2020/09/rrss_01.png?resize=1024%2C1024&ssl=1 (last accessed 14.01.2022)

This method is today the mainly used one, because of its high standard of security.
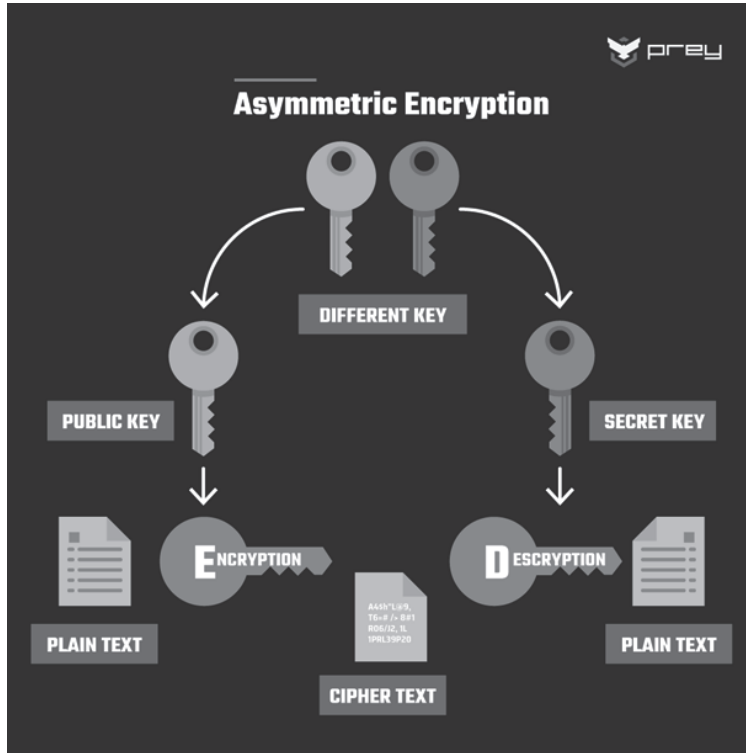
Figure 38: Asymmetric encryption[187]

3.5.2. Circumventing censorship by VPN

First of all, we need to determine what censorship is about. Censorship is an act of controlling or hiding a piece of information.[188] The topic of censorship is divided into two parts.

On the one hand, the pre-censorship is an occurrence of the publication mostly by media (books, movies, etc.) where the media is controlled by a governmental office. The governmental office decides whether there has to be a modification or the media is ready to publish it. This part of censorship is in Germany written in the constitution in Article 5.[189]

On the other hand, post-censorship is a mechanism that controls after the information is published. Everybody can have a free opinion but if it violates a law the person who breaks the law can be punished.[190]

---

[187] https://preyproject.com/uploads/2020/09/rrss_02.png?resize=1024%2C1024&ssl=1 (last accessed 14.01.2022)
[188] https://www.collinsdictionary.com/de/worterbuch/englisch/censorship (last accessed 23.11.2021)
[189] https://www.dwds.de/wb/Vorzensur (last accessed 10.12.2021)
[190] https://www.researchgate.net/publication/348871122_Wiemker-Dalg_-_Censorship (last accessed 10.12.2021)

The lack of uncensored information is perceived as a problem by the majority of the people, so they solve it by circumventing censorship.

The most common way to bypass censorship today is probably the VPN (virtual private network). A VPN is a technical application that disguises the data from the computer that the person uses by encrypting it and builds on that way a protected network connection. Mainly this function is a way to make it for third parties difficult to follow or to nab the data from the person's computer.[191]

The technical part of the VPN works as in the next paragraphs explained:

"A VPN hides your IP address by letting the network redirect it through a specially configured remote server run by a VPN host. This means that if you surf online with a VPN, the VPN server becomes the source of your data. This means your Internet Service Provider (ISP) and other third parties cannot see which websites you visit or what data you send and receive online. A VPN works like a filter that turns all your data into "gibberish". Even if someone were to get their hands on your data, it would be useless."[192]
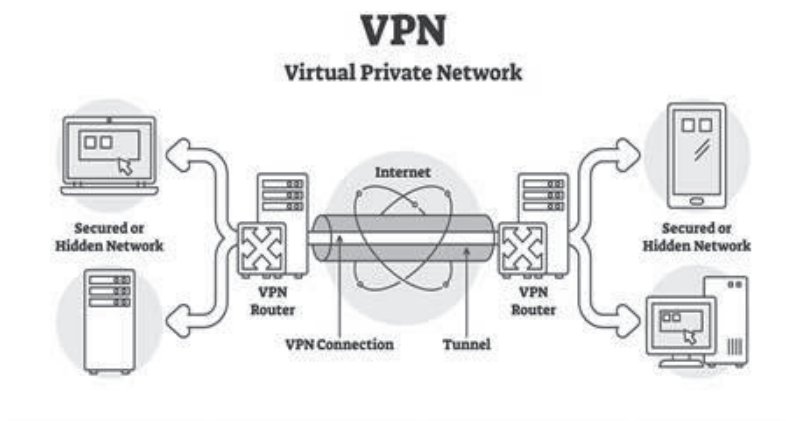
Figure 39: Virtual Private Network[193]

To prove its simplicity and usefulness, the following paragraphs describe how to connect to a VPN and one example of its value. The example is from the HVF Ludwigsburg and details the connection to the school's online library database.

The advantages of using the library database are the free access to a variety of academic/scientific sources and other types of literature. The first step is to download a VPN client, called 'OpenVPN'. In a browser of your choice search "OpenVPN", go to their website and download the VPN Client.

---

[191] https://www.kaspersky.com/resource-center/definitions/what-is-a-vpn (last accessed 08.11.2021)
[192] https://www.kaspersky.com/resource-center/definitions/what-is-a-vpn (last accessed 17.11.2021)
[193] https://10rgcev9tbx3hzifb27uulgw-wpengine.netdna-ssl.com/wp-content/uploads/2021/06/VPN-in-Online-Casinos.jpg (last accessed 14.01.2022)

When the download is completed, you open the OpenVPN and go to profiles to set up the connection to your university. As you see in the following picture below, only 2 fields have to be filled in: the name you assign to the connection (nr. 1) and the server-hostname (nr. 2).
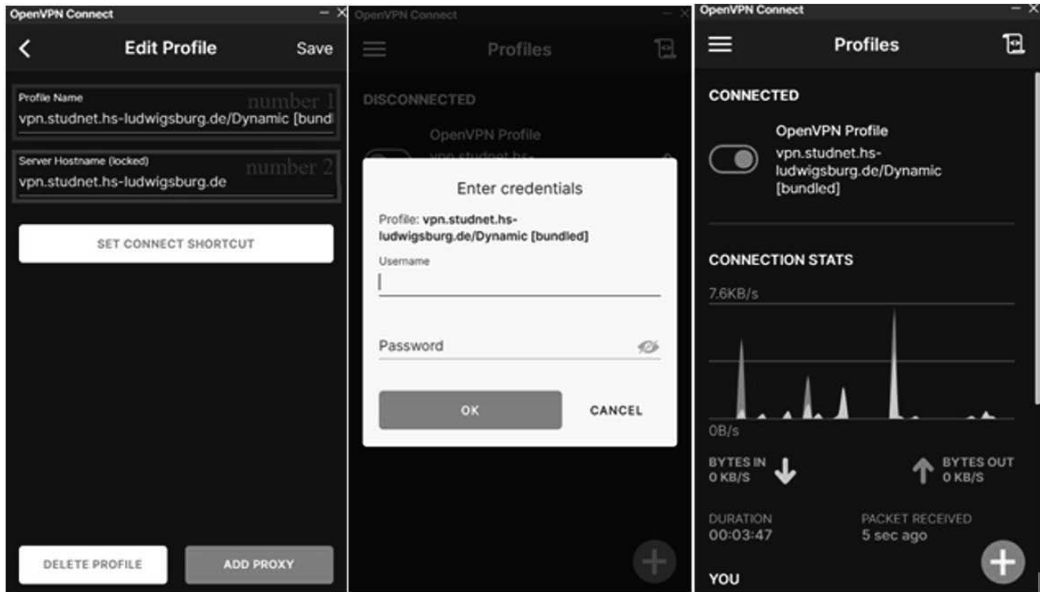
Figure 40: Own recordings from the OpenVPN Connect

Following this step, you can save the connection and proceed to connect, as presented above. When pressing "Connect", your client computer establishes a safe and encrypted TLS/SSL connection based on the public key of the server hs-ludwigsburg.de. The public key is a part of the asymmetric encryption used. It establishes an asymmetrically encrypted safe tunnel between the client (you) and the server (hs-ludwigsburg.de). All the work you must now do is to enter your username and your password to identify that you are a student or a teacher of the HVF and then you are connected to the university via an encrypted and safe VPN connection. Now you can use e.g., Beck Online with many features, that would otherwise be costly if not for the VPN connection.[194]

---

[194] https://www.hs-ludwigsburg.de/fileadmin/Seitendateien/einrichtungen/bibliothek/Dateien/OpenVPN_Stud.pdf (last accessed 20.12.2021)

Figure 41: Beck Online via VPN[195]

The advantages of using a VPN are diverse. To only name a few:

- Inaccessible data traffic

- Your VPN connection disguises and encrypts all data

- Secure connection

- The encryption of your VPN (depending on how reputable it is) can only be cracked with an encryption key. Without the key, it would take millions or billions of years.

- Disguise of your location

- Source of your data is after using a VPN the VPN-Servers so the location of the VPN is the only visible thing. All data that is transferred after this server is disguised so no one can trace it back to your location. And the IP address with which you access Facebook and post some hate speech is the one of the VPN – so authorities cannot catch you as long as the VPN provider does not deliver you to them.

- Access to regional content

- The source of your data is the location of your VPN, therefore you have access to the data in the region of your VPN server. Some VPN hosts offer the service of choosing the location of your VPN operations. This is called VPN-location-spoofing. A great example of the usage of this feature is Netflix accessibility in different regions. More series and films are available on Netflix in the USA than in Germany and the users circumvent it with a VPN.

- Secure data transfer

- The data that is transferred is safe from manipulation or spying. Most of the bigger companies use VPNs to transfer data between facilities.

---

[195] https://beck-online.beck.de/Search?pagenr=1&words=Hatespeech&st=&searchid= (last accessed 19.12.2021)

The only disadvantages are the minimal slower speed of your internet and the fact that you must trust the VPN host. If the VPN host is corrupted by the government or hijacked from a hacker attack this connection can be traced back to your position.

Besides that, the service (Chinese internet, Amazon, …) that you use can find out that you are using a VPN but not the exact data. Some governments that control the internet (e.g., China) can block the usage of a VPN via black lists. But they can't block every usage of VPNs because most of the companies also use VPNs for communication. Therefore it would have a significant negative economic impact on the country.[196]

3.5.3. Government and VPN

If a government is confronted with VPN users, it has the following options to deal with the situation:

First of all, the government has to simply accept the usage of the VPN. This is the simplest solution and overall, the preferred one. This is the opinion of the majority of the internet community and civil society. It also is an issue in the business world because the VPN is also used to communicate and deliver confidential business information between branches.

The second option is to prohibit the usage of VPN by its citizens (and all other users worldwide using VPN) either by blocking all the VPN traffic with a negative/black list of server-IP-addresses from the VPN servers or by enabling access to social media, blogs and governmental sites only when the user is listed on a positive/white list (like publishing companies open their journals and repositories to users from listed university IP addresses only).

Using positive/white list would dramatically reduce the traffic on the respective websites, turning the effective traffic to zero in the worst case. If e.g., a newspaper with a forum only accepts postings from specific white-listed servers, its dissemination will likely diminish.

Note that the basic and underlying technology of a VPN, namely a Public Key Infrastructure with private and public keys and establishing a symmetrically encrypted VPN-tunnel is exactly the same technology, which is used in online banking, webmail services and many other services on the internet. So, prohibiting or abolishing this technology or the demands of digitally illiterate politicians "all cryptographic keys must be accessible to the government" would de facto disable any safe application on the internet and for example, open your online banking account to the government.

Using a black list is very inefficient because there are many providers that can offer easy access to the VPN servers.[197] There are also numerous browsers that provide a VPN function, e.g., Mozilla Firefox or Opera, either as an addon or build-in in the main browser (as shown in the picture below).

---

[196] https://www.dw.com/de/zensur-mit-vpn-umgehen-ist-das-überhaupt-sicher/a-56816688 (last accessed 17.11.2021)
[197] https://www.heise.de/download/specials/Anonym-surfen-mit-VPN-Die-besten-VPN-Anbieter-im-Vergleich-3798036 (last accessed 07.12.2021)

Figure 42: Mozilla Firefox VPN[198]
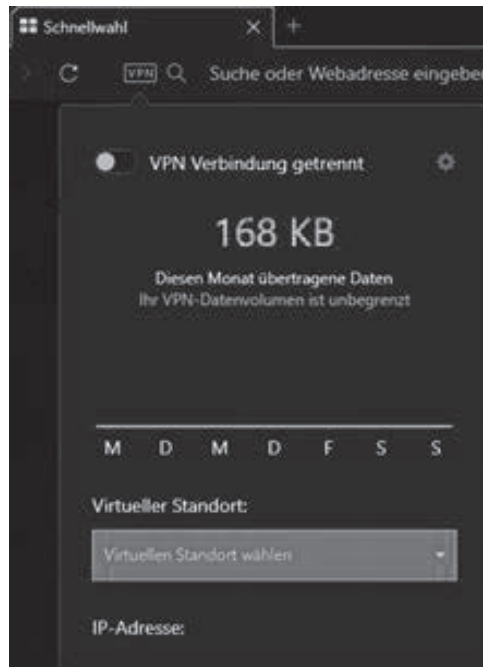
Figure 43: Own recordings from Opera GX

3.5.4. Data protection

This chapter will explain the role of data protection for VPNs. First of all, what is the meaning of "data protection"?

---

[198] https://www.mozilla.org/de/products/vpn/ (last accessed 05.12.2021)

"It is the protection of the individual against impairments of his rights to informal self-determination, under which every citizen can in principle determine himself about the disclosure and use of his personal data. (BVerfGE 65, 1)".[199]

This protection is on the legislative side granted in Europe because of the General Data Protection Regulation (GDPR). It regulates what personal data can be collected and processed. The scope of this regulation are organisations with the company headquarters in the EU and all organisations worldwide that process the personal data of EU citizens.[200]

The usage of a VPN is sometimes related to the lack of speech freedom. Often people in suppressed countries use VPNs or TOR (see next chapter) to communicate freely because otherwise, they face punishment. One example is a VPN provider in China, that was put in jail recently (December 2017) for five and a half years, just for providing users with a connection to the rest of the world.[201]

It is also a way to communicate freely without the risk of surveillance, which motivates people to use this communication channel. Such cases can be illustrated by whistleblowers such as Edward Snowden or Silver Meikar.[202]

3.5.4.1. Key escrow

Another approach to "controlling" the internet is a system called key escrow. It is also known as a "fair" cryptosystem. The mechanism behind this name is a simple agreement with a third party to store the keys that are used to decrypt the data. The keys stored by third parties can only be accessed by authorized persons or groups inside a business (e.g. head of the security) or in some cases the government itself. Note that the keys affected would also include keys for online banking and other very sensitive applications, therefore enabling a third party to trade your securities on the stock exchange, pretending to be you.

One of the downsides of key escrows is on the structural side. How is access granted only to authorised users? No system has been designed yet to overcome this challenge, mainly because the danger of abuse is very high. Many negative implications also arise from the use of this system on a national level. Many people don't trust the government or have concerns regarding keys' safety ensured by the government from a security perspective (e.g., hacker attacks).[203] Implementation of the system at the national level poses many struggles, one of the latest examples being France.[204]

3.5.4.2. NIS directive

The Network and Information Security Directive (NIS directive) is a part of the European cybersecurity strategy. The goal of this directive is to strengthen EU-wide cybersecurity. The main instrument is the enhancement of cooperation on more layers. This fits the timeline o the enforcement of EU cybersecurity: Regulation on establishing ENISA in 2004, EU Cybersecurity Strategy in 2013, the new

---

[199] https://wirtschaftslexikon.gabler.de/definition/datenschutz-28043 (last accessed 27.12.2021)
[200] https://www.atinternet.com/de/glossar/gdpr/ (last accessed 27.12.2021)
[201] https://www.heise.de/newsticker/meldung/Urteil-gegen-VPN-Dienst-Chinese-muss-fuenfeinhalb-Jahre-in-Haft-3926954.html (last accessed 03.01.2022)
[202] https://news.err.ee/104712/whistleblower-and-pm-put-scandal-in-perspective (last accessed 07.01.2022)
[203] https://jumpcloud.com/blog/key-escrow (last accessed 07.01.2022)
[204] https://www.icommercecentral.com/open-access/france-struggles-to-implement-worlds-first-trusted-third-party-infrastructure-with-key-escrow.php?aid=38879& (last accessed 07.01.2022)

regulation on ENISA in 2013, the NIS directive in 2016, and the Cybersecurity Act in 2019 [3-8, p. 84]. As a European Union directive, it has to be transferred into national law in all member states.

The NIS directive can be separated into three parts. These include the national capabilities, the cross-border collaboration and the national supervision of the critical sectors. National capabilities are cybersecurity capabilities of the individual countries, such as computer security incident response team (CSIRT), performing cyber exercises, etc. National supervision of critical sectors refers to the protection and supervision of the critical infrastructure cybersecurity, such as water, healthcare, energy, finances, and digital service providers, like online marketplaces or clouds.[205] It also refers to the Public Key Infrastructure and effectively hinders key escrow.[206]

3.5.5. A stronger alternative to VPNs: TOR

TOR is an abbreviation that stands for "The Onion Routing". This might seem a hilarious name, but if one understands the system and the procedures behind it, the name makes perfect sense.

Figure 44: TOR logo[207]

The origins of TOR are traced back to a project started by the US Navy. The network was developed for the US Navy and other military organisations to communicate online anonymously. It became popular because the project was public, to allow volunteers to work on it. Many users are choosing this browser nowadays because it is safe and has many functions that allow people to act without fear of being traced back or spied out.[208]

3.5.5.1. The technical structure of the TOR-browser

TOR provides safe operations because not even the browser knows your identity. The TOR browser builds tree "tunnels" to the destination. Instead of tunnels, these connections are like onion layers that are protecting and pile each other. That's why is called "The Onion Routing". These onion layers are not interconnected and no layer knows the destination and the identity of the user at the same time. Due to this structure, the TOR browser is secure - the internet cannot collect data if there is no data to collect. This type of procedure is called "privacy by design".[209]

---

[205] https://www.enisa.europa.eu/topics/nis-directive (last accessed 08.01.2022)
[206] https://digital-strategy.ec.europa.eu/en/policies/nis-directive (last accessed 25.01.2022)
[207] https://www.torproject.org/static/images/tor-project-logo-onions.png (last accessed 20.12.2021)
[208] https://vpnoverview.com/privacy/anonymous-browsing/tor/ (last accessed 25.11.2021)
[209] https://www.dw.com/de/zensur-mit-vpn-umgehen-ist-das-überhaupt-sicher/a-56816688 (last accessed 17.11.2021)
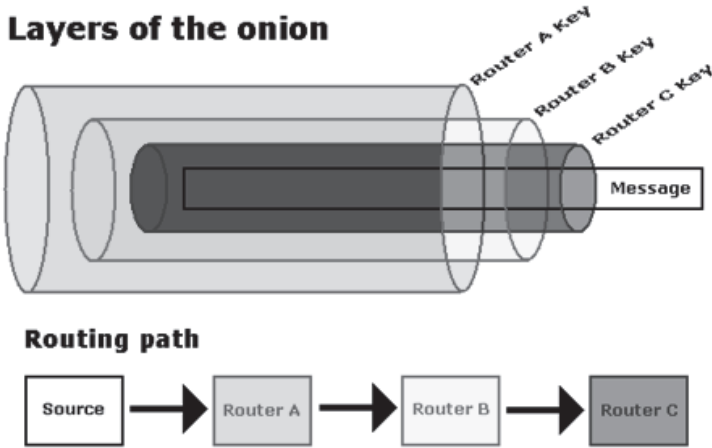
Figure 45: Onion routing[210]

As the picture below shows, the original data is encrypted and sent to a daisy chain of different servers. These encryptions are not edited but the encryption is added on top of the already existing encryption. Also, the servers that are used follow no strict pattern; they are randomly picked out. So, the users can surf and act anonymously. At the moment, this is the safest way to protect the identity of persons or to protect data.
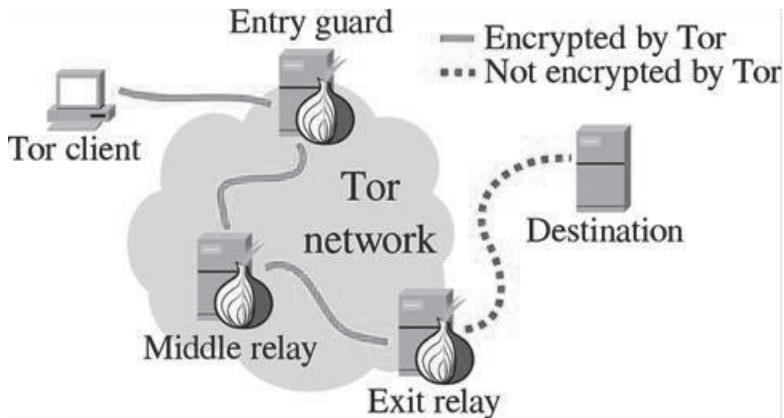
Figure 46. Onion routing (continued)[211]

The entry into TOR, namely the first TOR server or so-called entry guard can be any server whose administrator joins the TOR network. The government can of course blacklist such a server, but

---

[210] https://i3.moyens.net/de/images/2021/05/1621764758_981_Was-ist-Zwiebel-Routing-und-wie-koennen-Sie-Ihre-Privatsphaere-zurueckerhalten.png (last accessed 21.12.2021)
[211] https://br.atsit.in/de/wp-content/uploads/2021/06/download-tor-browser-fur-windows-mac-offline-installer.png (last accessed 21.12.2021)

millions of other servers could fill in. So, it is next to impossible to shut TOR down – and fully impossible for the government of a single state out of 194 states worldwide.

3.5.5.2. Advantages and disadvantages of TOR

The advantages of TOR are overwhelming and include:

- Inaccessible data traffic

  o TOR encrypts the data three times and covers its protection, so nobody can read the data except the addresses at the final destination.

- Secure connection

  o The way back to your final destination cannot be traced, because the existing encryption does not know your identity and the final destination simultaneously. Also, the brute-force attack to crack the encryptions used to secure the connection would take up billions of years.

- Disguise of your location

  o The source of your data when using TOR is the last used server of the TOR procedure, also known as the exit node, so the location of the TOR server is the only visible thing. All data that is transferred is disguised so no one can trace it back to your location. And the IP address you use to access Facebook and post hate speech is the address of the VPN – so authorities cannot catch you if you don't make any mistakes.

- Access to regional content

  o The source of your data is the location of the last used TOR server TOR. So you have access to the data from the region of the TOR server, as well as websites and services, that can't be found by regular browsers or search engines.

- Secure data transfer

  o The data that is transferred is safe from manipulation or spying because of the massive encryption.

- Absolute anonymity

  o Due to the encryption and the server structures, the identity of the person is protected, which benefits large groups of users.

TOR disadvantages are the slower speed of data connection, so streaming is possible at very low quality and the connection is more time-consuming, determined by the encryptions and the servers used. Another significant disadvantage is due to anonymity, which attracts many criminals. This is the consequence of making such a project public.[212]

---

[212] https://vpnoverview.com/privacy/anonymous-browsing/tor/ (last accessed 25.11.2021)

3.5.5.3. Users of TOR

The users of the TOR browser were originally the government agents and the military, employing it to communicate without fear of being intercepted or corrupted. But the user base has increased significantly. Today it is used by those who want to protect their confidentiality or profit from online anonymity, such as political activists, investigative journalists and whistle-blowers like Edward Snowden.

Another user group are the people in suppressed countries, where the authorities will punish you for certain opinions or online views. TOR allows them to communicate and enjoy more freedom because their statements cannot be traced back to a specific person. A third user group of TOR are individuals who bypass geo-restricted content or censorship and visit specific websites. TOR allows access to many web pages that are not visible for usual browsers because their addresses are not indexed by popular search engines Google or Bing.

Due to its anonymity, the browser also attracts criminals, that use it for communication, black markets trading illegal drugs and weapons and child porn. TOR browser is the only browser that lets you visit the dark web.[213]

3.6. Levels of the web

The *clear web* (also known as the *surface web*) is the section of the internet that can be publicly accessed from any browser. However, the other levels of the web, the so-called *deep web* and *dark web* are gaining more attention from the public. This chapter aims to deepen the understanding of these terms.

The differences between these levels of the web are determined by two attributes: indexing and encryption, which impact the transparency and visibility of the web and its content. The clear web is the visible part of the internet that gets indexed, meaning all search engines can scan these websites using crawlers and include the websites into a database of possible search results. The second characteristic, encryption is not usually found on the clear web, allowing users direct access. Some examples of the clear web are those accessible via search engines Google, Bing or online shops like Amazon. The ratio of the websites from the clear web to the whole internet is approximately 1 to 4%.[214]

Nowadays the borders of the clear internet and the deep web are fading. The deep web begins where websites are encrypted or not accessible via URL (free access hindered by means like mandatory logins, registration requirements, paywalls, etc.). So, when someone is buying from Amazon, to make the transfer via bank transactions on the online banking website, he transitions from the clear web (Amazon) to the deep web (bank websites). The deep web is mostly the source of the information about the clear web, accounting for cca. 90% of the websites, therefore without the deep web, the internet would be impossible. [215]

The deepest level of the web is the dark web or the darknet. The websites of the darknet are indexed and heavily encrypted, where URL is the first cryptographic key needed to decrypt the asymmetric encryption of a website. URLs don't include "http" or "https" or domain names like ".com"or ".org". Darknet accounts for cca. 6% of the total number of internet websites.[216]

---

[213] Ibid.
[214] https://techjury.net/blog/how-much-of-the-internet-is-the-dark-web/ (last accessed 30.01.2022)
[215] Ibid.
[216] https://www.anwalt.org/clear-web/ (last accessed 25.01.2022)
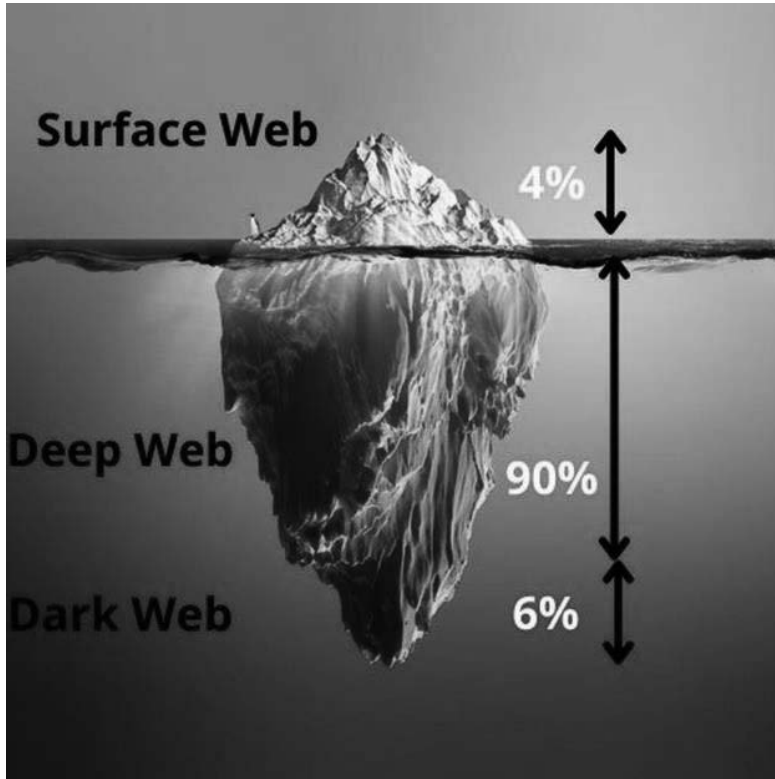
Figure 47: Web levels explained[217]

3.6.1. Domestic-only "Internet"

Most countries grant their subjects "free" use of the internet. "Free" in this case doesn't mean one can do anything on the internet, but rather the access to the internet is not restricted by the government or other groups of people. As usual, there are some exceptions, which will be exemplified in this chapter.

The first one is the Peoples' Republic of China, which runs a project called "The Great Firewall" or "Project Golden Shield". This project regulates the people's internet dramatically, with significant outcomes, like the limitation of information sources, blocking of internet tools like Google, Wikipedia, messengers, social networks (Meta (Facebook), Twitter, …) and mobile apps.[218] The censorship in China is implemented by modifying the search results and censoring the "wrong" opinions. Besides censorship, the PRC has influenced internet companies to provide the internal internet structure needed to diminish the effectiveness of other (foreign) companies.[219]

---

[217] https://postpear.com/wp-content/uploads/2021/06/Surface-Web-Deep-Web-Dark-Web-Internet-Explained.jpg (last accessed 24.01.2022)
[218] https://www.internetzensur.info/china/ (last accessed 30.01.2022)
[219] https://www.washingtonpost.com/world/asia_pacific/chinas-scary-lesson-to-the-world-censoring-the-internet-works/2016/05/23/413afe78-fff3-11e5-8bb1-f124a43f84dc_story.html (last accessed 30.01.2022)

The Chinese government has further plans to prohibit the usage of VPNs and TOR browser and make it punishable. In the recent past, a VPN provider received a five-year prison sentence.[220]

North Korea is another country with heavy network censorship established, resulting in a more restricted usage of the internet. Access to the internet, which is mostly used for governmental purposes, is only possible based on a special authorization.[221] These two countries are the biggest countries restricting access to the internet.[222]

3.6.2. Real name compulsory

Another approach to controlling the internet in some countries is the so-called real name compulsory system. Internet users are required to bind their virtual identity to their real one, using their legal name, including the government-issued e-ID.

The goal of this procedure is to prevent users from spreading false or denigrating information in internet forums, for example, because these people can then be identified and held accountable for their actions on the internet. But this procedure is only working when all internet providers require users to use their real name.[223]

There are various problems when it comes to real-name compulsory policy. First of all, at least in the CoE member states and the internet regime established there, it is easy to bypass this requirement via VPN. It means that even though a country may choose to enforce this policy, internet users can circumvent it by using a VPN server located in another country. Second of all, a vast majority is not using e-IDs provided by the government. Therefore, the Spiegel website, which only allows articles to be read by users with real names, would experience a huge decrease, by over 60% of their web traffic.[224] The consequences of such actions would be a significant decline in the company sales, eventually leading to its insolvency.

In conclusion, the real-name policy is not a viable approach to controlling the internet, at least not yet and not if enforced by only a small group of countries.

3.7. Social Media

Social media are mostly provided by big profit-oriented businesses, listed on the New York Stock Exchange, such as Meta Inc. or Twitter Inc. Social media provided by non-private entities are the rare exception, which raises the question of the business models behind them.

3.7.1. Business models and features

Social media business models vary greatly, all sharing the same end goal of gathering more traffic.

---

[220] https://vpntester.org/blog/china-statuiert-exempel-5-jahre-haft-geldstrafe-fuer-vpn-betreiber/ (last accessed 22.01.2022)
[221] https://www.bbc.com/news/world-asia-37426725 (last accessed 22.01.2022)
[222] https://www.wired.com/1997/06/china-3/ (last accessed 22.01.2022)
[223] https://www.golem.de/news/login-dienste-wer-von-der-klarnamenpflicht-profitieren-koennte-2002-146687.html (last accessed 29.01.2022)
[224] https://de.statista.com/statistik/daten/studie/777662/umfrage/nutzung-der-online-ausweisfunktion-des-npa-in-deutschland/ (last accessed 29.01.2022)

The "freemium model" offers several basic services for free and users need to upgarade to access other services. The providers have to figure out exactly how many services can be free so that the users are willing to upgrade, otherwise, the users may decide against it and only use the free services.

In the "affiliate model" a business makes money by guiding users to generate leads or sales on the websites of other affiliated companies. Nowadays many businesses rely on affiliated websites, to increase traffic to their websites and sell their products. Upon purchase or participation of users, the affiliated business receives a share of the transaction.

A "subscription model" requires users to pay a monthly or annual fee to access a product or service. It is common for monthly membership websites to have a high attrition rate because many users forget about the site after their first or second login and never visit it. Therefore, owners of such websites should make consistent efforts in keeping the site interesting and up-to-date.

"Virtual goods model" is another business model, where users pay for virtual goods like upgrades, points, or gifts on a website or a game. Three main categories of goods include functional, decorative and status items. The owners of such websites have to produce things that users want and need and that are relevant to the community, in order to be able to sell them.

The "advertising model" means the operator of a website sells advertisements, based on their internet traffic. The higher the traffic, the higher are the advertising charges. Therefore, users can still use the website for free while the operator can monetize it through advertising.[225]

There are also publishing and planning tools, which enable users to garner more attention for their content, especially if they post it at certain times when the engagement rate is higher. These tools are based on analytical features, that determine the peak engagement rates, most visited content items, competitors' activity levels etc. Based on such analyses, the users acquire more views, sell their products or have higher chances of becoming influencers.[226]

Besides the business model, a social media platform has to be appealing for the users. Hence it needs to have some essential features.

- *Simple and friendly user interface* – incorporating different elements, such as the content and media layout, input controls, navigation etc. user interface has to be simple and easy to navigate, regardless of the target audience.

- *Versatile and responsive* - user interface has to adapt and be responsive to different devices (smartphone, iPad or notebook) and screen sizes, without any loss of functionality or quality.

- *Visually appealing and accessible design* - the design elements have to be consistent and well-organized, so they don't create sensory overload and are accessible to everyone, the fonts and colour schemes are carefully chosen to promote a cohesive and pleasant user experience.

- *Secure login* - social media platform should deliver safe procedures for a unique user account with personal login settings and identification methods, such as backup email or code authentication, to prevent malware attacks or identity theft. Users must have a choice of what

---

[225] https://mashable.com/archive/social-media-business-models (last accessed 17.12.2021)
[226] https://sproutsocial.com/insights/best-times-to-post-on-social-media/ (last accessed 17.12.2021)

personal information they are willing to share and make publicly available (name, contact information, location, occupation, etc).

- *Networking elements* – one of the most sought-after features, allowing users to create personal or professional networks of their choice, which may consist of friends, family, colleagues or people with similar interests. The app should allow users to add other accounts into their networks and follow each other.

- *Content sharing* is one of the best social media app features because it enhances communication between people and strengthens the feeling of connection. Content sharing may include posting and sending photos or videos and the possibility to comment on what other users are sharing.

- *Public and private messaging* is a very valuable feature, enabling an easier and cheaper communication channel with other people (group chats, video calls), as compared to long-distance calls or expensive text messages. Assuming the user is connected to a WiFi network, messaging will not affect the data plans.

- *The open forum* offered by social media platforms allows users to voice their opinions, rally together for a cause, or even discuss their hobbies, with like-minded individuals.

- *Real-time notification and an activity feed* keep people up-to-date and informed, therefore being an essential feature of social media.

- *Privacy settings* should be an essential feature of any social media platform. Users should be able to determine who can see their profiles, what personal information is shared and have the ability to opt-out of certain marketing tactics, like tracking inline browsing or shopping experiences.

These are the most important features of any social media platform, enabling them to gain users.[227]

3.7.2. Algorithms

Algorithms are usually defined as sets of rules or instructions focused on solving a problem or fulfilling a task. Algorithms are not inherent to the digital world, a recipe is also an algorithm, as it has instructions on how to perform the task of making a meal. Digital devices like computers or smartphones need algorithms to execute the functions of various hardware or software-based routines.[228]

Algorithms are essential for social media. But how do they work? Most social media providers consider their algorithms a business secret and it's often not known how exactly they work, because the algorithm is subsequently not published. They influence our use of the Internet and, especially our use of social media. The content shown on the internet depends on algorithms, which are run when the respective website is accessed. For instance, the shown content may depend on the browser version or the language setting of the individual user accessing the website with his device. To a certain extent, algorithms work similarly to a strong filter. Only a small part of all available information will be presented to the individual user, therefore influencing users' perceptions or opinions. Moreover, they are not transparent (only the creator and distributor of the software know

---

[227] https://www.koombea.com/blog/10-top-features-of-social-media-apps/ (last accessed 20.01.2022)
[228] https://www.investopedia.com/terms/a/algorithm.asp (last accessed 30.01.2022)

exactly how they function) and most of them are hidden, i.e. the source code is not published; hence it is unknown where they are used and how they influence the usage of the Internet.

Algorithms have a significant impact on social media. For instance, the algorithm decides what posts are shown on the front page, as they are considered more interesting and relevant for the user. This decision is based on certain statistics: how often the user has hovered, read, liked, clicked, shared and commented. The more two users interact with each other, the more the algorithm interprets them as belonging to the same group and their respective content being of interest. It also means that the rest of the content is not visible to and from the user. Algorithms are also strongly influenced by the activity level: users have to be very active to be seen. Subsequently, the user needs to spend as much time as possible on social media platforms if he wants to be seen, shared, followed and liked.

When algorithms define content as relevant for each user, a single user runs the risk of being incased in a "bubble" of content a group of others wants to see. This is quite risky, if a user gathers information through social media because the algorithm does not care if the news is true or fake, the number of likes, shares or comments is relevant. As a result, an emotional or provoking comment can become much more popular than an objective article, even if is more interesting or important.[229] The users interact mainly with users which think alike and block users they do not like or disagree with. Like-minded users support each other and encourage each other's opinions about who they are or what they do, fortifying each other in their opinion or actions.[230] The algorithms have a great impact on these internet bubbles.

Censorship also contributes to enforcing these bubbles, because people get banned or their posts and comments get deleted, making them think it must be true and the government wants to hide it. This also drives the shift towards other platforms. When people who used Facebook and WhatsApp get banned they will shift to Signal or Telegram, where they can communicate directly with like-minded people.[231]

3.7.3. Censorship

To figure out how censorship works, we must understand how the internet works. The connection to the internet is always through an Internet Service Provider (ISP). This ISP allocates an IP address to every computer, which is similar to a postal address, identifying the person and transport information. Everyone who knows this IP address can figure the address authority country and even municipality. When the computer is used at an internet café or office, it is possible to event determine the building, office and the exact computer that is used. This information is often available to government agencies.

---

[229] https://webcare.plus/algorithmen-social-media/ (last accessed 06.12.2021)

[230] https://webcare.plus/zwischen-wohlfuehl-oase-und-meinungsvielfalt-in-den-sozialen-medien/ (last accessed 23.12.2021)

[231] https://www.dw.com/de/meinung-facebooks-querdenker-zensur-geht-zu-weit/a-59216883 (last accessed 10.01.2022)
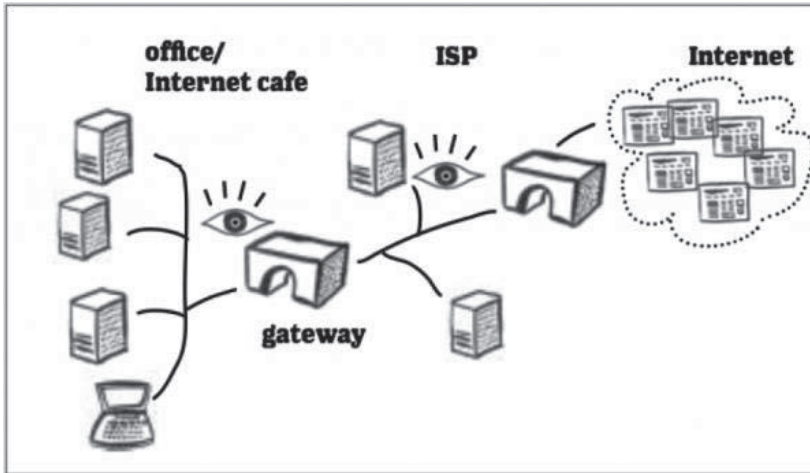
Figure 48: Internet connection[232]

However, not only computers have an IP address, websites also have them. To access a website, the IP address of the website can be entered in the address bar, not just the website address. Only, the IP addresses are convoluted and difficult to remember, so the Domain Name System (DNS) associates IP addresses with human-readable "domain names".
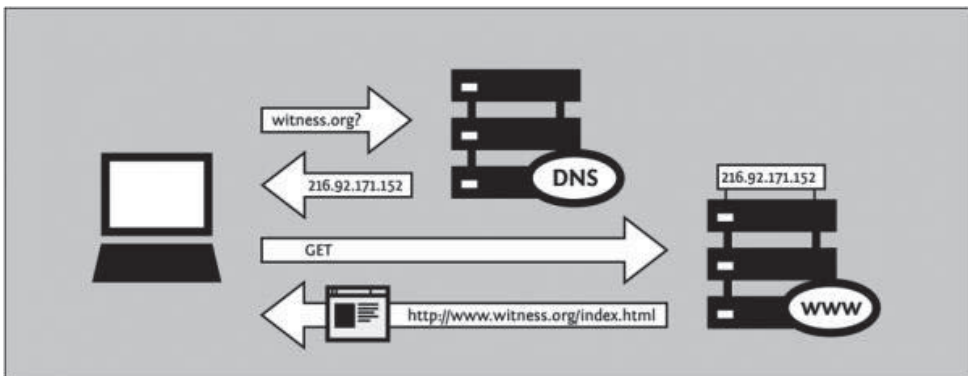
Figure 49: DNS[233]

To enable sending packets of information from server to server, the ISPs have to trust the established internet protocols on national and international infrastructure. This structure and conventions are normally referred to as the "backbone" of the internet. The backbone also consists of major network equipment installations that are interconnected via fiberoptic cables and satellites. Communication between internet users from different countries or even continents is enabled through these connections. Providers connect through routers, which are also known as gateways, which enable diverse networks to communicate with each other. But the gateways are also a point where the internet traffic can be monitored or even controlled.

---

[232] https://townsendcenter.berkeley.edu/blog/internet-censorship-part-1-technology-working-web (last accessed 04.01.2022)
[233] Ibid.

These complex processes are not seen by the average internet user. The understanding of these processes is necessary because they underline censorship on the internet. There are different types of censorship, that can be enforced at different levels of internet architecture.[234]

Censorship can be performed via different methods, such as DNS tampering or IP blocking. DNS tampering is one of the most common technologies, that can be used in countries where authorities control domain name servers. Officials can "deregister" the DNS to the censored content and these websites become invisible to the users because the DNS tampering will prevent the translation of domain names to website IP addresses.

IP blocking is enforced where governments have control over internet service providers, blacklisting specific IP addresses. When a user wants to access a certain website, the request is monitored by surveillance computers, that check the request with the blacklisted IP addresses. If the website is backlisted, the ISP will cancel the connection. This technology is frequently used in China, where international-gateway servers control the flow of internet information in and out of the country through mega-servers.

While IP blocking allows the government to block certain blacklisted websites, there are billions of websites with new ones created every second, making it impossible to keep updated blacklists. Keyword filtering could be a more powerful tool, it scans the Uniform Resource Locator (URL) string for keywords. If one of the forbidden words like "fascist" is encountered in the URL, the connection is cut. This also means that www.antifacist-initiative.org would subsequently be cut.

One of the newest and most sophisticated internet censorship techniques is packet filtering, meaning the actual contents of each page are scanned. Data sent via the internet is grouped in small units – packets - that are passed from one computer to another via routers. While IP address filtering only blocks websites based on where packets are going to or coming from, the packet filtering also inspects the content for banned keywords. If a forbidden keyword occurs in a packet, the connection is cut. The user may get an error message, without indicating he or she just got censored. It is important to note that packet filtering does not work when the content of the communication between the user and the website is encrypted – like in every online-banking session, which is encrypted via TLS/SSL.

Besides these wide-ranging internet censorship techniques, others like traffic shaping may be used. This is often used by governments or corporations, by delaying access to certain websites and simulating a slow-loading or unreliable website. A commonly used technique among companies is to blacklist individual port numbers, like Web or email, thus regulating certain employee behaviours, such as instant messaging.

Internet censorship is often disguised as a technical error or connection problem; therefore making it difficult to identify as censorship, which technology is used or who is blocking the website. This also makes it difficult to prevent censorship, but proxy servers or virtual private networks (VPN), filters can be bypassed, although not always rendering consistent results, as the following graph depicts.[235]

---

[234] https://townsendcenter.berkeley.edu/blog/internet-censorship-part-1-technology-working-web (last accessed 04.01.2022)
[235] https://townsendcenter.berkeley.edu/blog/internet-censorship-part-2-technology-information-control (last accessed 23.12.2021)
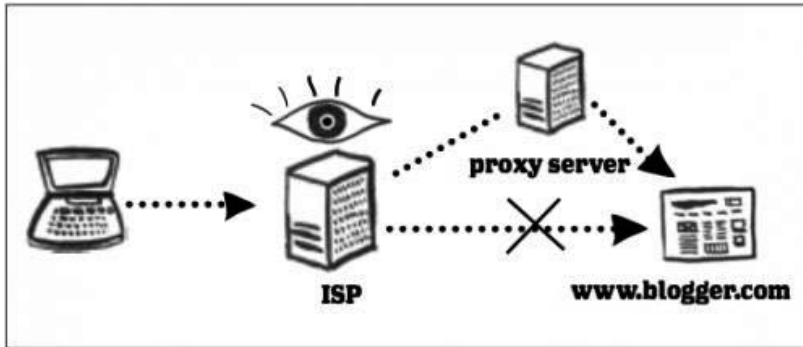
Figure 50: Avoiding internet censorship by using a proxy server[236]

Internet censorship has become an issue in countries with freedom of expression, mainly to the recent COVID-19-pandemic. In Germany many problems arose with the so-called "Querdenker".[237] This is an initiative hostile towards COVID-19 policies of the German government, which was blocked by Facebook. Following many demonstrations, these decisions were reversed, because the ban was not considered reasonable enough by administrative courts. This is a relevant case when because of a country's inability to tackle the issue, a private company became the real censor.

Facebook and other big social media providers are committed to taking action against those users on the internet who discriminate, insult, threaten other users or incite to violence. But it has to be reasonable. When 150 user profiles are banned because of "hate speech and inciting to violence" it's reasonable, but "publication of health-related misinformation" doesn't sound as reasonable. Many statements are covered by freedom of expression for good reason, no matter personal preferences.

These actions of the companies are not governed by law and arbitrated by courts, but by decisions based on property rights. When a company acts on its own, beyond the control of legislation and courts, just to satisfy the expectation of market demand, this is unacceptable. The current ban on the "Querdenker" was enforced on Facebook and Instagram but not on the WhatsApp messenger, although it also belongs to Facebook. The "Querdenker" already communicate and exchange via Telegram, as it's not forbidden to search for alternative channels to disseminate information, but everyone can file charges if they find discriminatory, violent or other criminal content there.[238]

The other social media giant Twitter got criticized when they banned the profile of the then-current president of the United States Donald Trump. Social media became a broadcast platform to reach out to the masses. And because the internet was based on the premise that, if you do not like it, you don't look, the government did not get involved or didn't impose regulations on the Internet. Most experts agree that this is not a censorship issue because the government is not the censor.[239]

---

[236] https://townsendcenter.berkeley.edu/blog/internet-censorship-part-1-technology-working-web (last accessed 04.01.2022)
[237] A German equivalent of QAnon, very similar.
[238] https://www.dw.com/de/meinung-facebooks-querdenker-zensur-geht-zu-weit/a-59216883 (last accessed 20.01.2022)
[239] https://www.forbes.com/sites/petersuciu/2021/01/11/do-social-media-companies-have-the-right-to-silence-the-masses--and-is-this-censoring-the-government/ (last accessed 20.01.2022)

References Chapter 3

[3-1] Leiner, Barry M., Cerf, Vinton G., Clark, David and Kahn, Robert E. (2009). A Brief History of the Internet. In: ACM SIGCOMM Computer Communication Review vol. 39 no. 5, pp. 22-31, https://dl.acm.org/doi/10.1145/1629607.1629613

[3-2] Treaty of Bern in the version Bukarest 2004 as published on the Austrian Federal platform as "Gesamte Rechtsvorschrift für Weltpostverein – Weltpostvertrag (Bukarest 2004), Fassung vom 5.10.2021", https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20006773 (last accessed 08.02.2022).

[3-3] H.R.3162 - Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001, https://www.congress.gov/107/plaws/publ56/PLAW-107publ56.pdf (last accessed 08.02.2022).

[3-4] Bendix, W., & Quirk, P. J. (2016). Deliberating Surveillance Policy: Congress, the FBI, and the Abuse of National Security Letters, Journal of Policy History, 28(03).

[3-5] Deutscher Bundestag (2017), Beschlusempfehlung und Bericht des 1. Untersuchungsausschusses gemäß Artikel 44 des Grundgesetzes, Drucksache 18/12850, https://dserver.bundestag.de/btd/18/128/1812850.pdf (last accessed 08.02.2022).

[3-6] Bederna, Zsolt et al. (2021) Modelling computer networks for further security research, Security and Defence Quarterly 9(36) 16 p. doi: 10.35467/sdq/141572.

[3-7] Szadeczky, Tamás (2018), Security of E-Government Website Encryption in Germany and Hungary, Academic and Applied Research in Military and Public Management Science 17(2) pp. 127-138 doi: 10.32565/aarms.2018.2.9.

[3-8] Szádeczky, Tamás (2020) Governmental Regulation of Cybersecurity in the EU and Hungary after 2000, Academic and Applied Research in Military and Public Management Science, 19(1), pp. 83–93. doi: 10.32565/aarms.2020.1.7.

4. Legal foundation – do legal remedies work?

Authors: Konstantinos Katevas, Timo Steidle and Max Winter
Academic supervisor: Sebastian Brüggemann

4.1. Introduction

With hate speech and fake news on the rise across the internet, politicians are faced with the responsibility to act decisively against their spread.

Several questions need to be answered to approach this issue: How are hate speech and fake news defined in a legal context? What legal remedies or potential policies can be put in place to stifle them? What hurdles and challenges will the government face when enacting these policies? What is the potential impact on universal human rights like the freedom of speech and information?

In the following section, the legal definitions and approaches of individual countries in regards to both hate speech and fake news will be analyzed and compared.

Further, several methods will be evaluated for their expected efficacy in the pursuit of dealing with hate speech and fake news online, while predicting potential short- and long-term effects.

4.2. Fake news

The difference between fake news and hate speech is that hate speech generally harms individuals or members of a specific group, whereas fake news is arguably damaging to society as a whole. This raises problems and questions for governments: whether they should try to regulate fake news with legal restrictions or not. Often those laws directly contradict other basic rights like freedom of speech.

The legal situation regarding the fake news is comparable to the section concerned with hate speech. It can be rather difficult to find common ground, especially on an international level and between different cultures.

The idea of combating fake news is not a subject exclusive to the 21st century. In 1936 the member states of the League of Nations agreed on the "International Convention Concerning the Use of Broadcasting in the Cause of Peace". One of the main aspects of this agreement was to prohibit the spread of fake news for propaganda purposes. Regarding the period of the agreement, the main concerns were about war propaganda and the related consequences. [4-1]

Articles three to five of the "International Convention Concerning the Use of Broadcasting in the Cause of Peace" mainly address the issues with fake news. Article 3 thereby states, that "any transmission likely to harm good international understanding by incorrect statements shall be rectified at the earliest possible moment". The following Article 4 elucidates the goal of the agreement by petitioning the participating states to "ensure [...] that stations within their respective territories shall broadcast information concerning international relations the accuracy of which shall be verified".

To reiterate, following the agreement meant that information needed to be checked and verified before being spread via radio broadcast. If a published statement was found to be incorrect or false, the responsible nation had to ensure that the spread of fake news is stopped and corrected as soon as possible.

While this agreement may be outdated in many aspects, the intentions are comparable with the current state of affairs regarding the spread of fake news.

These days different criminal codes acknowledge fake news as part of the definition of fraud. This means that in many countries, like for example the UK or US, fraud can be committed by spreading fake news if there is a related benefit for the perpetrator or harm for the victim. In German criminal law, the crime of fraud (§ 263 StGB) also requires the perpetrator to "distort or suppress true facts" ("Entstellung oder Unterdrückung wahrer Tatsachen"), which can be described as propagating fake news.[240]

Additionally, the crime of fraud according to the German criminal law also requires immediate disposal of property, which could cause some problems when trying to compare it to the spread of fake news. It might be quite difficult to prove a direct connection between an article that contains fake news and immediate disposal of property.

In addition to the offense of fraud, fake news can potentially fulfill the elements of offenses such as defamation (§ 187 StGB) or incitement of the people (§ 130 StGB), to name a few examples. A closer look on the offense of defamation (§ 187 StGB) shows, that if a perpetrator "asserts or disseminates an untrue fact about another person" ("in Beziehung auf einen anderen eine unwahre Tatsache behauptet oder verbreitet"), or in other words "if someone is spreading fake news about another person", he can be legally punished for that.

This rather small national sample size of legislation already shows the difficulty when trying to combat fake news effectively with the help of legal remedies. There are many different offenses in already existing criminal codes, which can be fulfilled under specific circumstances by spreading fake news.

In the following section, different examples of local legislation regarding fake news will be discussed and further examined. The analysis will focus on the criminal codes of member states of the CoE.

4.2.1. Local legislation regarding fake news (in Europe and other countries)

4.2.1.1. Germany – Netzwerkdurchsetzungsgesetz (NetzDG)

In 2018 the so-called "Netzwerkdurchsetzungsgesetz" (NetzDG) or "Network Enforcement Act" was passed in Germany to stifle different kinds of harmful actions on the internet such as illegal material, hate speech and fake news. The NetzDG applies to social media platforms with at least 2 million members in Germany [4-2].

According to section 3 of the act, social media networks have to implement an "effective and transparent procedure for handling complaints about unlawful content" (NetzDG, paragraph 3) which also includes the likes of fake news. In addition to that, any social media network that reaches this large of an audience has to "remove or block access to content that is manifestly unlawful within 24 hours of receiving the complaint" (NetzDG, paragraph 3).

---

[240] "Fighting Fake News or Fighting Inconvenient Truths?" - https://verfassungsblog.de/fighting-fake-news-or-fighting-inconvenient-truths/ (Last accessed 26.10.2021).

The responsibility of taking down potentially damaging content and thus supporting the competent authorities is delegated to social media networks. Some of them have already voiced their concerns about including Fake News in the scope of the NetzDG. In a study by Prof. Dr. Marc Liesching from July 2020 Facebook addresses the issue, that the NetzDG only applies if a statement or comment contains criminally relevant contents. Fake news can often not be considered in this context as they fall under freedom of speech [4-3]. Furthermore, this form of prosecution raises the question of whether a company should be eligible to decide if a post or comment is within the legal boundaries or not.

The second problem regarding the NetzDG is the national jurisdiction. On one hand, the scope of this legal regulation is clearly stated in Art. 1 to only social media networks with more than two million registered users in Germany. The scope of the NetzDG is pretty much narrowed down to only the providers of social media networks that meet these requirements.

On the other hand, the problem of limited national jurisdiction is addressed in Art. 4 NetzDG which outlines that "the regulatory offense may be sanctioned even if it is not committed in the Federal Republic of Germany". With this article, the German legislator states that their scope of national jurisdiction can be applied outside of the national territory. This raises concerns regarding the international law principle of territoriality [4-4].

This legal problem is not exclusive to the NetzDG. As already mentioned above, the German criminal code "StGB" also addresses the dissemination of fake news in section 263. The StGB regulates the jurisdiction of crimes committed outside of its territory as follows: "If the participant to an offense committed abroad acted within the territory of the Federal Republic of Germany, German criminal law applies to the participation even if the act is not a criminal offense according to the law of the place of its commission" (§9 StGB). Because almost all websites and contents are accessible in Germany, paragraph 9 StGB is very widely applicable. If executed strictly, this section would lead to an international jurisdiction for the German prosecution authorities when it comes to cybercrime.

4.2.1.2. France – Law against the Manipulation of Information

In comparison to Germany, the French legislation against fake news is focused on a different aspect of the topic. The so-called "Law against the Manipulation of Information", passed in December 2018, is centred around the fight against election misinformation. This legislation was caused by the attempt to interfere with the 2017 presidential election in France. Before the election took place, a coordinated attempt to undermine the presidential candidacy of Emmanuel Macron with the help of a systematic misinformation campaign was started. As a direct result of these actions the "Law against the Manipulation of Information" was enacted to prevent such campaigns in the future.[241]

To respect other rights such as freedom of expression and communication a fake news campaign has to meet certain requirements to be considered a punishable offense in the sense of the "Law against the Manipulation of Information". As such, the digital information has to be objectively false, misleading and threatening to the honesty of an upcoming election. This means that information or news must be considered manifestly false to be punishable. With this large barrier implemented freedom rights can be respected and protected [4-5, p.12].

The "Law against the Manipulation of Information" only addresses a small field regarding the topic of fake news. This further limitation of the material scope may raise questions and concerns about

---

[241] "Measures to tackle disinformation in selected places" - https://www.legco.gov.hk/research-publications /english/2021in14-measures-to-tackle-disinformation-in-selected-places-20210623-e.pdf (Last accessed 29.11.2021).

the overall usefulness of the legislation. The law only applies to fake news and misinformation that is demonstrably false. However, fake news is often not entirely false but exaggerated, sensationalized or taken severely out of context [4-5, p.13].

On the other hand, this limited scope of the French law guarantees the respect of the freedom rights mentioned above. These strict requirements ensure that only objectively and truly false information is considered illegal while dissenting opinions or statements are still protected from the danger of state-imposed censorship.

4.2.1.3. Global

On a global scale, several countries have addressed the issue of fake news in special sections within their respective criminal codes.

Canada used to have a legal section that addressed the spreading of false news in their criminal code. According to Criminal Code Section 181, if a perpetrator were to publish a statement or news that is known to be "false and that causes or is likely to cause injury or mischief to a public interest", they can be punished with imprisonment for up to two years.[242]

This section was repealed in 2019 because it interfered with the constitutional right of freedom of expression and was thus deemed unconstitutional.

This decision was made, after a neo-Nazi, who had published antisemitic literature, was prosecuted under section 181 of the Canadian criminal code. The Supreme Court of Canada ruled, that there is a fine line between the truth and falsehood, that cannot be defined by law. This then led to the repeal of section 181.[243]

Another prime example of legislation against fake news is the Anti-Fake News Act (AFNA), created by the Malaysian parliament in 2018. This ordinance punishes actions like the creation, publication, and distribution of fake news, "with intent to cause, or which is likely to cause fear or alarm to the public, or any section of the public" (AFNA, section 4). The AFNA was repealed in December of 2019 because of a new coalition after the national elections of that year.

This law is under heavy criticism because it was re-enacted in 2021 via a Proclamation of Emergency to fight the COVID-19 pandemic, whilst some believe the ordinance was promulgated to restrict media reports of the pandemic and thereby constrain aspects like free speech and press freedom.[244]

In comparison to the others, South Korea took quite a different approach. To combat fake news in their respective state, the ruling party wanted to change the "Press Arbitration Act" so that publishing fake news or false information could be punished. The bill specifically targets media outlets such as newspapers, magazines or TV and radio channels. If a media outlet publishes fake news, the

---

[242] Criminal Code (R.S.C., 1985, c. C-46)" https://laws-lois.justice.gc.ca/eng/acts/c-46/section-181-20030101.html (Last accessed 26.10.2021).
[243] "Is Canadian Law Better Equipped to Handle Disinformation?" - https://www.lawfareblog.com/canadian-law-better-equipped-handle-disinformation (Last accessed 27.10.2021).
[244] "The rebirth of Malaysia's fake news law – and what the NetzDG has to do with it" - https://verfassungsblog.de/malaysia-fake-news (Last accessed 27.10.2021).

retribution will be five times the estimated damage caused by false information. This damage will be calculated according to the social influence, as well as the total amount of sales or views.[245]

Instead of battling fake news on social media like Germany or in the context of presidential elections like France, the South Korean focus is more narrowed down towards "classical and established" media.

Within this approach, one big advantage can be determined: the prosecution of this bill. As outlined in chapter 3 – Technical Foundations, there are many options for internet users to conceal themselves and hinder or impede the law enforcement authorities. Another problem, especially regarding multi-national participants can be the jurisdiction in the context of state sovereignty. These difficulties are circumvented by the South Korean bill. Because the remedy focuses on media companies that are situated within their state, the law enforcement is similar to any other crime and does not have the issues associated with the prosecution of online crimes.

This bill still faces a lot of criticism by the government's opposition and international journalism organisations for encroaching on the "freedom of the press". As a result, the law was not passed as intended and the vote has been delayed until 2022.[246]

The differences between the main examples of national legislation against fake news, Germany, France and South Korea are listed in the following table to add clarity.

---

[245] "The Trouble With South Korea's 'Fake News' Law - " https://thediplomat.com/2021/08/the-trouble-with-south-koreas-fake-news-law (Last accessed 20.12.2021).
[246] "How South Korea Is Attempting to Tackle Fake News" - https://thediplomat.com/2021/11/how-south-korea-is-attempting-to-tackle-fake-news (Last accessed 20.12.2021).

	Germany	France	South Korea
Legislation	Network Enforcement Act	Law against the Manipulation of Information	Press Arbitration Act
Scope / Objectives	Preventing illegal content online (like fake news, hate speech, etc.)	Countering pre-election disinformation campaigns	Preventing media outlets from spreading false information
Parties regulated	Social network platforms (with over 2 million users in Germany)	Online platforms (exceeding a certain amount of distinct French users)	Media outlets (TV, radio, newspapers)
Regulatory tools	Social network platforms are required to remove illegal content within 24 hours	Online platforms have to provide a mechanism to report fake news, judges are responsible for deciding about the content before a general election	Courts can issue heavy fines on media outlets, if they are found guilty of spreading fake news
Criticism / Issues	Social network platforms have to decide in the first place whether a post is illegal or not; leading to the danger of overblocking	Fake news has to be determined objectively false within a short timeframe; in practice, this can be quite difficult to decide	Possible infringement of basic rights, such as the Freedom of Speech may occur

Table 1. National legislation against fake news

4.2.1.4. Conclusion

As shown with these brief examples, legislators in different countries all over the world have been trying to quash fake news with the help of regulations and laws. Their approaches might differ, but the development in legislation against the harm of fake news shows, that many countries are well aware of this problem and actively try to prevent the negative impact via legal remedies.

A common criticism against such laws is their encroaching on the freedoms of speech, expression and the press. The national legislators acknowledge these concerns and try to balance the fight against fake news with the preservation of individual and societal rights.

Further problems and legal issues, like the national jurisdiction, the applicability of laws online, as well as the principle of proportionality in the context of freedom-rights versus restrictive legal remedies, will be discussed and explained in section 2.2.3 of this chapter.

4.2.2. Legal Prerequisites

Local legislation concerning fake news can vary significantly, as showcased in section 1.1. As such, there are no general or international prerequisites, which would make something qualify as fake news and be punishable all over the world.

In the following sections, some examples of legal prerequisites will be showcased and discussed whilst referring to the previously outlined national laws from section 1.1.

4.2.2.1. A common definition is desirable to be able to take legal action against fake news

As explained in the beginning chapter of this book, an all-encompassing and universally correct definition of the term fake news simply does not exist. In different societies or legal and political systems, the conception of what exactly is fake news may vary in considerable ways. Therefore, legal prosecution especially across national borders is pretty much impossible.

A common definition of fake news may solve this issue or is at least desirable for governments to be able to suppress fake news effectively using legal remedies. Because most criminally relevant actions regarding fake news are often taken via the internet, a common definition is even more important due to the global reach of the medium.

Another reason that speaks to the importance of a clear-cut definition of fake news is the legal principle of "nulla poena sine lege" ("no penalty without law"). According to this principle, a clearly defined criminal offense is needed to enforce a prosecution. It is applied when there are no clear definitions or assessments for a criminal offense [4-6]. If "nulla poena sine lege" was neglected, judgments in fake news cases would be easy to challenge.

Even though a common definition would be ideal for legislators to be able to fight fake news with legal remedies, the controversies around the restrictions a law like that would pose to the freedom of the press, showcase that such a common definition is unlikely to find purchase in the near future.

Governments should therefore try to find other ways to counteract fake news on a smaller scale, whilst still working on a global solution.

4.2.2.2. Different treatments depending on the perpetrator

Is the legal treatment of fake news different, when it originates from an "official" news source, or even a government institution when compared to a private blogger who is spreading disinformation?

This question is not only considered as an ethical or political problem but can be examined in a legal context. First of all, it is important to determine whether a distinction between a private person/company or the public authorities is even necessary for a legal examination. From a societal point of view, this difference is a substantial part. If a person only speaks for themselves and shares fake news on his or her private blog, the amount of damage from this action is most likely much smaller than if a government organisation is spreading false information. The public authority would arguably have a higher responsibility to be truthful. The shown examples of national legislation further above do not differentiate between those two groups of perpetrators. This further emphasizes the thesis that the distinction is not relevant in a legal context and more of an ethical aspect.

A recent example of fake news published by a public authority is the case of the former Austrian chancellor Sebastian Kurz, who has allegedly spread disinformation as part of his political campaign. Before further explaining these actions, it is important to state, that by the time of writing this paper those actions are still under investigation and the presumption of innocence is still entitled.

The former Austrian chancellor and his allied colleagues are suspected of paying tabloids to publish fake opinion polls and overly positive media coverage before the national elections of 2016. The financing was allegedly funded by state money.[247]

Even though not all details of this incident are screened as of now, it already shows the difficulties and problems that the connection of false information and public authorities can cause. Whether a legal remedy against such behaviour would work is up for debate.

4.2.2.3. Are those regulations applicable to the internet?

At first glance, the internet might appear as a legal environment with a completely different set of rules or even as a place without any form of regulation. Whereas this might be a somewhat common point of view, it generally does not apply in a legal sense. Many laws that apply offline are equally applicable online. It is mandatory to state that this might vary regarding different national or local legislation.

Additionally, in this context, the question of fake news being a cybercrime or not can be raised. As with many other terminologies regarding this topic, there is also no clear-cut definition of what can be considered a cybercrime. One definition of cybercrime was published by Nir Kshetri in 2010, which states that "[...] a cybercrime is defined as a criminal activity in which computers or computer networks are the principal means of committing an offense or violation of laws, rules, or regulation" [4-7].

This pretty narrowed-down definition only focuses on crimes like fraud, forgery, data manipulation and other similar actions. In recent years other criminal offenses in the online spectrum have emerged and are no longer part of this definition. Actions like online stalking or bullying are noteworthy examples of such [4-8].

So, regarding Kshetri's definition, spreading fake news would not be considered a cybercrime. However, there are opposing opinions who are including fake news in the terminology of cybercrime. For example, Robert Smith and Mark Perry state in their article about "Fake News and the Convention on Cybercrime", that spreading fake news on social media can be considered a cybercrime [4-9].

Even though these statements support different opinions they have something in common, both of them only focus on the online aspect of fake news. In these articles, publishing or spreading of fake news in other forms of media, such as TV, radio or newspapers, are not taken into consideration. Other than hate speech, which is committed most of the time in an online environment, fake news is not necessarily a cybercrime. This marks a clear distinction in a legal context between the likes of hate speech in comparison to fake news.

Fake news can be published online, but are also oftentimes spread via other forms of media. Therefore, the term cybercrime is not perfectly suited for fake news, as the criminal acts are equally happening offline and online. Regarding fake news, the term cybercrime may be applicable for some aspects of it but certainly not the whole topic.

---

[247] "Austria: First arrest in Kurz corruption probe – reports" - https://www.dw.com/en/austria-first-arrest-in-kurz-corruption-probe-reports/a-59483916 (Last accessed 30.12.2021).

Besides the question regarding the applicability of laws online and the terminology of cybercrime, another urgent problem is the legal jurisdiction, particularly concerning legal cases across national borders. The global nature of the internet might lead to conflicts between different national laws and moral concepts. Crimes like the spread of fake news introduce a new possibility: a criminal action can be committed in one state whilst unfolding its effect in a different state. One action might be legal in the state where it is taken but be illegal in the state where its effects take place. This undermines national sovereignty with the principles of national legislation and criminal prosecution. As already stated in section 1.2.2 the legal jurisdiction can be a crucial aspect for example whilst debating between contradicting legal regulations.

In the EU there is a specific regulation for organisations regarding this exact problem. The so-called country of origin principle (with a few exceptions) was introduced by the e-commerce directive (2000/31/EC) to address this issue [4-10].

Article 3 of the directive states that "Each Member State shall ensure that the information society services provided by a service provider established on its territory comply with the national provisions applicable in the Member state in question which fall within the coordinated field" [4-11].

This means that in the EU the establishment of an organisation is the decisive reason on which national laws apply to them. So, if for example, an organisation located in member state A spreads fake news regarding a topic on their website, hosted by a server in member state B, the legal regulations of member state A are applicable and not the national laws of state B. The jurisdiction is thereby regulated clearly by article 3 of this direction. In a case of fake news spreading within the EU the jurisdiction resides with the member states in which the organisation that caused the incident is established.

As already stated, this regulation only applies to organisations or companies and not to private individuals. Furthermore, the directive is inapplicable to organisations established outside of the EU and its member states. Therefore, organisations that want to cause harm by spreading fake news can circumvent the regulation quite easily by simply setting up their head office in another country outside of the EU.

Another difficulty, which has been briefly mentioned before in this chapter, is the principle of proportionality regarding legal remedies introduced by constitutional states. This is one of the main reasons why many examples from section 1.1 "Local legislation regarding fake news (in Europe and other countries)" struggled with their respective legislation. Legal regulations against fake news almost always come with a more or less extensive restriction on other fundamental rights, such as the freedom of speech or the freedom of information. While the principle of proportionality applies to the internet as well as to offline circumstances, about the freedom of speech, the internet is oftentimes seen as an "unregulated place", where anybody can share their thoughts freely. Thereby making a regulation to combat fake news in an online environment is a difficult task for any legislator because arguments like censorship almost immediately arise.

Whereas these principles are present in all constitutional states, the reach or possible limitation of those freedom rights are different in their specific manifestations. When comparing the European country Germany to the United States of America, this difference becomes much more prevalent. In Germany, for example, the freedom of speech is restricted by section 130 of the criminal code, where expressions like approving or denying "an act committed under the rule of National Socialism" (§ 130 StGB), are legally punished.

On the other hand, in the United States of America, attempts to combat fake news with the help of legal restrictions contradict the First Amendment where "freedom of expression by prohibiting Congress from restricting the press or the rights of individuals to speak freely" is guaranteed by the constitution. This showcases, that different countries place different emphasis and boundaries on their citizen's fundamental rights.

Described by Dr. Ron Paul as a "war on Free Speech" the proposed legal regulation against fake news is seen from a different perspective, with a much bigger focus on freedom rights and a distinct refusal of any sort of government restriction [4-12].

So as a conclusion, it is important to state, that applying legal regulations to the internet is very much possible and laws are often either applicable online as well as offline, or even specifically meant to combat crimes on the internet. The difficulties hereby are usually less of a legal problem like a missing criminal offense in the applicable laws and more of practical execution.

In many cases, the prosecution of internet crimes is too complex and disproportionate to the seriousness of the offense. Also, the legal jurisdiction is an important aspect when trying to combat fake news online, as the perpetrator, potential victims, the server, and the internet platform or social media network can all be located in different countries with a variety of potentially applicable laws.

4.2.3. Hate speech

4.2.3.1. The legal definition of hate speech

4.2.3.1.1. Definition of hate speech as determined above

"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, 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."

That is the definition of hate speech from the Council of Europe as it was shown in the first chapter. But just as there is no clear and universal default definition, there is also no internationally legally recognized definition, what is considered hateful is disputed [4-13].

Therefore, each country has its own definition and its own way of dealing with hate speech, as the following examples will show:

4.2.3.1.2. Different legislations

4.2.3.1.2.1. Germany

In Germany, criminal offenses are punished under the German Criminal Code (Strafgesetzbuch). For the offense of hate speech, several paragraphs can be used. For example, the sections "§ 130 Volksverhetzung" or "§ 185 Beleidigung" come into question. Hate speech is also defined in § 130

section 1 No. 2 of the German Criminal Code. This is understood to mean a disruption of public peace by inciting hatred, violence or arbitrariness towards groups, population groups or individuals based on nationality, religion or ethnic origin. This also includes an attack on human dignity in which an individual is maliciously insulted, defamed or despised, based on his or her membership in a particular group or part of the population [4-14].

As already mentioned in chapter 1.1 Fake news - local legislation, the NetzDG came into force on September 1st, 2017, to combat not only fake news but also other cybercrimes such as hate speech [4-15].

It is intended to ensure that criminal content is deleted in social networks within a reasonable time frame. Similar to fake news, in the case of hate speech, difficulties arise when assessing the content, or the network provider does not adhere to the given deadlines for deletion. The Federal Office of Justice acts as a supervisory authority and can impose fines in the event of non-action.[248]

4.2.3.1.2.2. France

In France, criminal offenses are punished according to the French criminal code "code penal" Hate speech is punishable there as "discours de haine" [4-16].

The proposal for a law to combat hateful content on the Internet is intended to increase the obligation to remove hateful comments by making criminal reactions more efficient and preventing dissemination. Furthermore, the responsibilities should be clarified and the platform operators should be asked to cooperate more. Proposition de loi visant à lutter contre les contenus haineux sur internet, "Avia Law" is based on the German NetzDG and also has a definition of hate speech [4-17].

4.2.3.1.2.3. Austria

In Austria, hate speech, if it is made public and accessible to many people, is defined as incitement to hatred under § 283 of the Austrian Criminal Code. This paragraph also contains a definition that is very similar to the German definition. [4-18]

4.2.3.1.2.4. Interim conclusion (member states of the EU)

As can be seen in the examples shown above, the European countries all have a similar point of view regarding hate speech. Nevertheless, there is still a need for action, as an example from 2015 shows. In a report published in 2015 by the European Commission against Racism and Intolerance (ECRI), incitement to racism was only punishable in Estonia when the victim's health, life or property were threatened [4-19].

However, in the European Union, the definition, the elements of the crime and the opinion that hate speech should be legally prosecuted are largely the same. Other cultures may have completely different moral values and views. An example of this is the USA, which has a very different view on hate speech.

4.2.3.1.2.5. United States of America

In principle, it must be mentioned that constitutional protection is very pronounced in the United States of America. Among other things, this is due to the fact that the American Constitution is very

---

[248] "Hasskriminalität in sozialen Netzwerken bekämpfen" - https://www.bundesjustizamt.de/DE/Themen/Buergerdienste/NetzDG/NetzDG_node.html (Last accessed 12.01.2022).

old compared to most other countries. In addition, the American courts grant almost absolute protection to freedom of expression, which also distinguishes them from the courts of other countries. This is based on the fact that there is a strong distrust of government. People believe in a competition of opinions and are not supposed to distinguish between good or bad opinions [4-20].

A famous United States Supreme Court ruling has defined the limits of freedom of speech. In the 1969 case of Brandenburg v. Ohio, a Ku Klux Klan leader (Clarence Brandenburg) called for the possibility of revenge against "Jews" and "Negroes" and announced a march on the United States Congress. Brandenburg was initially convicted for this speech, but the Supreme Court overturned the conviction, calling Brandenburg's actions "imminent lawless action", which cannot be punished under the Constitution [4-21].

4.2.3.1.3. Conclusion / Problem

Aiming to avoid serious differences in the handling of hate speech, as shown by the example of the USA and European countries above, international rules must be established. This would facilitate prosecuting and punishing hate speech crimes internationally in a uniform way. These should best be established by major international associations or organizations to include as many different countries as possible. In the first instance, this should be incorporated into international law by the United Nations.

4.2.3.1.3.1. International law / United Nations

Under international law, no provision would prohibit hate speech at the moment. This is even though hate speech has a big impact on many areas of activity from the United Nations, such as the protection of the population and the fight against violence, racism and discrimination. Taking into account the international Human rights norms and standards and the right to freedom of expression, the United Nations has adopted a UN Strategy and a Plan of Action on Hate Speech. These grant the United Nations the necessary resources to take action against hate speech. The United Nations tactic is to undermine the causes of hate speech and to find effective responses to the impact of hate speech on society.

Also, incitement to violence, discrimination or hostility is nevertheless prohibited. This is a special form of hate speech, which is particularly harmful because of the risk of misdeeds or even terrorism [4-22].

4.2.3.1.3.2. European Union

Another supranational organization is the European Union, which, with its many bodies and member states, also has the possibility of drawing up supranational rules. Because of its smaller size compared to the United Nations, it also has a greater chance of ensuring that the rules it draws up are accepted by all and that compliance can be better monitored. So far, there is no legal regulation on hate speech at the European level. However, there are various approaches from European bodies such as the Council of the European Union or the European Commission to take action against hate speech. One possibility is "soft-laws", which prescribe certain things just like real laws, but these do not have to be adhered to and non-compliance is not punished [4-23].

European Code of conduct

The European Commission, Facebook, YouTube, Microsoft and Twitter signed the Code of conduct (CoC) on countering illegal hate speech online on May 31, 2016. Instagram, Dailymotion, Snapchat

and Goggle+ were added later. Jeuxvideo.com joined in January 2019, also TikTok in September 2020 and LinkedIn in June 2021 [4-24].

The Code of Conduct obliges the companies to check reports of hate speech within 24 hours, delete illegal content or block users. As reported by the Commission, IT companies investigate 89% of reported cases within 24 hours, of which 72% are deleted due to illegal content [4-25].

Thus, the Code of Conduct is the significant instrument for self-regulation of illegal hate speech on the Internet [4-26, p.53].

Guideline-Audiovisual Media Services

The European Audiovisual Media Services Directive (AVMS) was also amended to combat hate speech. The regulations, which previously applied only to broadcasters, now also cover video-sharing and video-on-demand platforms such as Netflix, YouTube and Facebook. As a result, video platform operators are required to create easy-to-understand and use mechanisms through which videos containing hate speech or glorifying violence can be reported and, after subsequent review, deleted by the operators.[249]

European Commission

Combating hate speech and hate crime is also a priority for the President of the European Commission, Ursula von der Leyen. On February 23, 2021, the European Commission published a proposal to declare hate speech an EU crime. At the moment, the European Commission is working on an initiative that should lead to a Council decision against hate speech. If this Council decision is taken, the European Commission will then have the power to propose substantive legislation. This would make it possible to standardize definitions and penalties for hate speech.[250]

4.2.3.1.3.3. Council of Europe

Minimum rules for the definition of criminal offenses and sanctions are necessary and decided to align laws and regulations for the implementation of a common policy of member states. According to Art. 83 section 1 TFEU, the European Council may, through a legislative procedure, lay down guidelines for minimum rules on criminal offenses. However, a special need and a cross-border dimension are required for this. The directives apply to many areas of crime, including cybercrime. If there are concerns about a directive, under section 1 or 2 regarding its compatibility with the respective criminal laws in the countries, a member of the Council of the EC can refer this to the ER for consideration and request a suspension and deliberation on it. A decision will be made within 4 months.

There are often differing opinions when it comes to establishing guidelines, but if at least nine Member States reach an agreement about cross-border cooperation, they manifest their decision to the European Parliament, the European Council and the European Commission within the defined 4 months. The conclusion is deemed to have been granted [4-27].

---

[249] "EU-Richtlinie für audiovisuelle Mediendienste" - https://www.medienkorrespondenz.de/politik/artikel/eu-richtlinie-fuer-audiovisuelle-mediendienste-umsetzung-bis-septembernbsp2020.html (Last accessed 04.01.2022).
[250] "Commission: Hate Crime Should Become an EU Crime" - https://eucrim.eu/news/commission-hate-crime-should-become-an-eu-crime/ (Last accessed 04.01.2022).

European Commission against Racism and Intolerance

Besides these legal attempts, the European Union has established some bodies that are only there to protect human rights such as the European Commission against Racism and Intolerance (ECRI), which specializes in combating discrimination and racism. This is closely networked with the Equal Treatment Bodies of the Länder and thus monitors them. The ECRI monitors Member States by analysing the circumstances and the actual state of affairs. When problems arise, the ECRI put forward proposals and makes recommendations. The equality bodies are independent authorities that combat racism and discrimination at the national level. In addition, relations are maintained with international organizations, such as the United Nations [4-28].

European Court of Human Rights

Another body of the European Union is the European Court of Human Rights (ECHR), which is a supranational judicial body that ensures that member states respect the human rights set out in the Convention on Human Rights. All 47 member states are also members of the Council of Europe. If the ECHR would develop a common practice/definition per jurisdiction, where the important parameters are defined, the 47 member states with 47 different criminal codes would not have 47 ways the courts apply them. For hate speech to be punishable uniformly, it is important to define the parameters. Does the post need to be public and have a specific reach or is an insult through a private message enough to be considered hate speech? Does it have to be specifically directed and received by the targeted person, or is it enough if the hate speech is posted in a private forum, which excludes the targeted group?[251]

Summing up, there are several enactments and rules concerning the handling of hate speech, but a common European guiding principle is missing. The present definitions of hate speech, no matter nation or international, differ about the modalities, the impacts as well as the consequences.

A guidance note issued by the "High-Level Group", which was launched by the European Commission in 2016, refers merely to the observance of case law in the responsible member states. Combating illegal online content and disinformation cannot be solved nationally. It must be tackled by the Member States together. Basic approaches already exist, for example, "Tackling online disinformation: a European approach," but need to be expanded. The focus should be on promoting tolerance and plurality in public institutions and in positions with political, media and financial responsibilities. The policy must represent values and set standards as well as communicate improved, wide-ranging and coherent solutions [4-29, pp.21,53].

4.2.4. Are those regulations applicable to the internet?

Hypothetically, if there was a common legal definition for hate speech, could regulations against it be applied to the internet?

As already stated in section 1.2.3 of this chapter, the question of whether legal remedies are applicable online or not can often be quite difficult to answer. In the following segments, these problems are going to be further addressed and discussed to emphasize the explanations given in the section about fake news prior.

---

[251] "What is the European Convention on Human Rights?" - https://www.amnesty.org.uk/what-is-the-european-convention-on-human-rights (Last accessed 20.01.2022).

4.2.4.1. Determining the perpetrator

As someone can be using the profile of someone else, either by hacking into their account or simply using their computer to post something through their profile it is nearly impossible to be able to determine a poster's identity with 100% certainty. This legally makes determining the perpetrator without reasonable doubt complicated. If hate speech was posted through someone's personal computer on their account and they live alone, it's difficult to argue that they weren't the perpetrator.

However, if it was a PC at work and they happened to not lock their screen when for example getting a cup of coffee, the perpetrator could have been several people who had access to that PC. In the German legal system, the company owns the account and is thereby liable to third parties under private law for any actions taken via this account.[252]

4.2.4.2. Legal jurisdiction

In addition to the many different definitions, the preconditions for committing a crime, the way the courts prosecute these crimes, and the problem that these rules cannot be perfectly applied to the Internet, there is another key issue. In the case of crimes committed over the Internet, it is also not clear at the first moment which court has jurisdiction at all. The territorial principle that assigned jurisdiction in the past no longer works in today's world with modern technology. When clarifying jurisdiction, several factors must be taken into account, such as the location of the perpetrator, the nationality of the perpetrator, the location of the victim or the nationality of the victim. Also to be included is the portal or platform through which the crime was committed. Thus, the laws of the country in which the company headquarters or servers are located may also become relevant.[253]

As the internet is accessible all over the world, determining which country's laws apply to a situation can be complicated. This situation can even get more complicated depending on which of the locations are within the EU or outside the EU. There are several factors to consider, for example, if person A posted hate speech against person B on a popular platform, that is accessible from the EU Member States, which of the following is relevant to pinpoint which countries' jurisdiction applies to the case.

- Person A's nationality/country of origin/ citizenship?

- Person A's location when making the post?

- Person B's nationality/country of origin/ citizenship?

- The location where the information is accessible?

---

[252] "Use of company internet connections and e-mail accounts as well as mobile phones and notebooks" - https://www.anwalt.de/rechtstipps/nutzung-betrieblicher-internetanschluesse-und-e-mail-accounts-sowie-von-mobiltelefonen-und-notebooks_062654.html [Last accessed: 01.21.2022].

[253] "Territorial principle of Germany" - https://www.juraforum.de/lexikon/territorialprinzip (Last accessed: 01-21-2022).

Victim/Perpetrator	Same Country	Different Country within CoE	Country outside CoE
Country A (CoE)	Laws of country A fully applicable	Application of Country A's laws limited	Likely no application of Country A's laws. In rare cases bilateral treaties
Country B (non-CoE)	Laws of country B fully applicable	Likely no application without bilateral treaties	Likely no application without bilateral treaties

Table 2. Sorting out legal jurisdictions

If this would not make the clarification of jurisdiction difficult enough, it must also be taken into account that each country has its own special rules for law enforcement.

Example: In Germany, the legal jurisdiction is determined in §§ 5 and following the German Penal Code (StGB). The so-called "Handlungs- / Erfolgsort-Prinzip" contains rules on jurisdiction. For websites, this means that the availability of the content in Germany is sufficient for the German Penal Code to apply, thus practically everywhere. Also, some crimes can be generally persecuted by law even if they were committed outside Germany, for example, §130 section 2 No. 1 StGB "Volksverhetzung", which can also be applied to hate speech [4-30].

Server location

In addition to the nationality of the perpetrator/victim, the server or business location may also be relevant. The Canadian Court of Justice has faced this issue before. The case considered whether a Romanian website that had no servers or business locations in Canada was subject to Canadian laws. The content was passed from the website, which was taken from the Canadian database of court decisions CanLII.org. The original website ensured that personal information provided by litigants could not be found by search engines. This was not done by the Romanian website. The Canadian court then found jurisdiction. The reason given was that although the business location and the server location were in Romania, there was nevertheless a sufficient connection to Canada, as Canadians were affected.[254]

A similar point of view was shown at the conference "Law, Borders and Speech", which was held at Stanford. At this conference, the importance of the server location in clarifying jurisdiction was discussed in particular. From Silicon Valley, the opinion was expressed that the server location should be the decisive factor in determining jurisdiction. The U.S. Court of Appeals for the Second Circuit takes a similar view, mentioning this in the comments to the decision in the lawsuit "14-2985 In the Matter of a Warrant to Search a Certain E-Mail Account Controlled and Maintained by Microsoft Corporation" with Microsoft.[255]

---

[254] "Server location not definitive in determining jurisdiction over foreign defendant" - https://www.lexology.com/library/detail.aspx?g=9ad16ad9-c363-4e00-a293-c61f19f9fcf6 (Last accessed: 01-20-2022).
[255] "Server Location, Jurisdiction, and Server Location Requirements" - https://blog.ericgoldman.org/archives /2016/12/server-location-jurisdiction-and-server-location-requirements-guest-blog-post.htm (Last accessed 15.01.2022).

Bilateral treaties

Enforcement of court judgments in many countries, including the US, depends on the principles of comity, reciprocity and res judicata and ultimately on the internal laws of each country. Bilateral or multilateral treaties and agreements can be made between countries to regulate legal issues and the recognition of judgments and their enforcement. There is no agreement in this regard with the United States. Reasons for this seem to be, among other things, the high sums imposed by US courts in connection with liability claims. Many countries consider the fines to be too high. There are also different opinions regarding extraterritorial jurisdiction, which prevent joint agreements with the US. However, in most countries, foreign judgments not providing for damages can generally be enforced if the following conditions have been met:

- the court, that made the judgement was authorized to judge and had jurisdiction in the designated case;

- the defendant was informed of all relevant facts about the case;

- the process was not influenced by fraud;

- the judgment was compatible with the public order of the country.

If a judicial decree does not require compensation for damages, after approval by the domestic local court, the judgement can in many cases be enforced, despite differences in the procedures of the countries.[256]

4.2.5. Hate speech vs. freedom of expression and freedom of religion

In the second chapter, it was explained that the right to freedom of expression is a fundamental human right, which was included in the Universal Declaration of Human Rights. In Germany, it is enshrined in the Basic Law of Germany (Grundgesetz). Art. 5 section 1 Grundgesetz grants everyone the right to freely express and disseminate their opinions in speech, writing and images. In the well-known Lüth decision, the Federal Constitutional Court describes this fundamental right as a fundamental element of democratic state order and a direct expression of human personality [4-31].

However, unlimited freedom of expression is not granted, so in certain cases, restrictions are applied. For example, statements that incite, encourage or justify hatred based on intolerance. Certain statements thus fall into the category of hate speech and are therefore no longer protected by the fundamental right to freedom of expression [4-29, p.16].

Moreover, the Federal Republic of Germany guarantees according to Art. 4 sections 1 and 2 Grundgesetz, everyone the right to confess a religion, to join it or to change religious affiliation, as well as the right not to confess any religion or to leave a religious community.[257]

Freedom of religion, also guaranteed by Art. 18 ICCPR (International Covenant on Civil and Political Rights) often leads to discussions and is considered controversial because it affects other fundamental

---

[256] "Enforcement of Judgments" - https://travel.state.gov/content/travel/en/legal/travel-legal-considerations/internl-judicial-asst/Enforcement-of-Judges.html (Last accessed 10.01.2022).
[257] "Religious Constitutional Law" - https://www.bmi.bund.de/DE/themen/heimat-integration/staat-und-religion/religionsverfassungsrecht/religionsverfassungsrecht-node.html (Last accessed 23.01.2022).

rights. Freedom of religion can be restricted too, to protect fundamental rights and freedoms, health and morals, and in the event of violations of public order and security.[258]

Concerning freedom of expression and freedom of religion, each case must be carefully considered. In regards to different speech restrictions, there are distinctive case-laws applied by the European Court of Human Rights, but there is still no precise specification about hate speech. So, there is a need for clarification and guidelines [4-26, p.34].

The answer to the question of whether freedom of expression also applies to use on the Internet is yes. However, it is not permitted, and therefore no longer covered by freedom of expression, if the personal rights of persons or groups of persons are thereby violated. This includes insults or falsehoods about them. It is also not allowed to incite violence on the Internet and it is forbidden to post pictures with certain symbols, for example, the swastika. All statements classified as hate speech are punishable, whether on the Internet or in real life [4-32].

Thus, everyone should be aware that statements made in the online world can have serious consequences. Some comments are posted in a small or closed group within social media, but others are posted on a public, mass-accessible platform that reaches a global audience. It seems that social networks are communicating with an ever-increasing number of people. All statements and content can be called on the Internet at any time and can therefore also spread uncontrolled. All statements once posted are available in the social networks until they are deleted, in contrast to verbal statements. Quick deletions are difficult but necessary to minimize the potential damage of some utterances.

4.3. Possible legal approaches

This section will explore possible approaches to combat both hate speech and fake news. By analysing each approach and determining its short- and long-term effects, the goal is to find a recommendation for the procedure.

4.3.1. Platform Liability

This approach makes any platform that is accessible from within the European Union liable for the content that is on their platform. It aims to shift the responsibility of keeping a website clear of hate speech and fake news to the platform owner themselves by relying on their economic interest to have access to the European userbase. A similar strategy was utilized for the German Netzwerkdurchsetzungsgesetz as showcased in 1.1 of this chapter.

The perceived advantage of such a method is that large platforms are easier to address and sue when compared to individual users. One of the prerequisites to apply such a measure is clearly defining what constitutes hate speech and fake news respectively, as the platform owners themselves would have to determine what content they would have to take down.

However, several problems may arise when using this strategy:

Large platforms host huge amounts of content, with additional massive amounts being uploaded daily. For example, on the social media platform Twitter the daily upload is on average 500 million posts

---

[258] "Freedom of religion and freedom of speech" - https://menschenrechte-durchsetzen.dgvn.de/menschenrechte /politische-buergerliche-rechte/religionsfreiheit-und-meinungsfreiheit/ (Last accessed 23.01.2022).

which translate to 200 billion posts per year.[259] As such, the expectation of content being checked manually is not realistic. A combination of automatic filters, as well as a reporting system, would be required, that allows visitors of the platform to report each other's content in case of violation.

Automatic filters using algorithms or artificial intelligence have come a long way in the past decade and are actively in use to take down posts that infringe the copyright or that depict child sexual abuse. However, they haven't been sufficiently developed to be able to discern the intent of a post [4-33]. Whether something qualifies as hate speech or remains within the boundaries of fundamental rights such as the freedom of expression, can be controversial [4-34].

A well-known example of courts being in disagreement over a matter of hate speech was the case of the German politician Renate Künast, who was the target of a large mass of insults online because of a remark she made regarding sexual abuse towards children. The first instance court ruled that the insults directed at her were not hate speech and that, as a politician speaking about a delicate topic, she was supposed to withstand harsher forms of criticism. Künast appealed against this ruling and the case went to higher courts that ended up revising the decision several times, each court disagreeing with the previous ruling [4-35, 4-36].

As such, writing a general predetermination into a program aiming to distinguish between something being hate speech or protected speech is unlikely to produce sufficiently accurate results – especially when dealing with 47 different legal systems in the 47 member states of the Council of Europe.

Depending on the severity of the sanctions toward the platform owner, they may choose to use filters that follow the philosophy of "rather safe than sorry". Doing so will result in a lot more content being flagged and removed than intended by the policy writers, resulting in so-called "overblocking" [4-37, 4-38].

Smaller companies and platforms, that do not have access to advanced filtering technology, would struggle to keep up with the Silicon Valley tech giants, forcing them to rely on their technical solutions and thus creating additional market entry barriers [4-39].

Platform owners may also choose to simply not make their content accessible for Europe-based users, as a way to avoid liability issues – similar to what happened when the GDPR was introduced.[260] If enough platforms react in this fashion, citizens of member states may end up in digital isolation from the rest of the global userbase. That could, in turn, result in a sizeable public backlash.[261]

In short – making the platform liable for the content on it would arguably be an effective way for governments to restrict undesirable content. However, the de-facto delegation of complex judiciary tasks to private companies without any financial compensation is likely to have external effects. It could lead to considerable downsides for internet users such as the restrictions being heavier than intended by the legislators. It would make the owners of social media platforms the arbiters of what falls under freedom of speech and what needs to be taken down ebased on either fake news or hate speech. It would also shift the liability towards the platform owners and generate additional costs for them.

---

[259] "Internet Live Stats", Available at https://www.internetlivestats.com/twitter-statistics/ (Last accessed: 25 January 2022).
[260] SENTANCE, REBECCA, GDPR: Which websites are blocking visitors from the EU?, 2018, Available at https://econsultancy.com/gdpr-which-websites-are-blocking-visitors-from-the-eu-2/ (Last accessed 28 January 2022).
[261] SOUTH, JEFF, More than 1,000 U.S. news sites are still unavailable in Europe, two months   after GDPR took effect, 2018, Available at https://www.niemanlab.org/2018/08/more-than-1000-u-s-news-sites-are-still-unavailable-in-europe-two-months-after-gdpr-took-effect/ (Last accessed 28 January 2022).

4.3.2. Blocking Access

A straightforward way to deal with platforms not abiding by European laws, regulations, or standards would be blocking access to them for users within the member states.

To successfully set this up, a list of standards would need to be created that are to be followed by websites wanting access to the European market. Those standards would have to be attainable and maintainable. Within the context of hate speech and fake news, however, this might prove more difficult. Hate speech in particular is most often produced in social media by the platform's users - rarely by the platform owners themselves. As such, it may be difficult to curate the posted content, especially for larger platforms for the reasons outlined under 3.1 of this chapter.

A blanket-blocking of (often foreign) websites and platforms might lead to geopolitical consequences. Similar requirements of following foreign standards may be imposed on European-based platforms by other countries in the long term, thus potentially disassembling the internet along national lines due to cultural differences.

Therefore, a possible public backlash needs to be taken into account in this case, as it constitutes a heavy restriction on the freedoms of European citizens. The preventative attempts may be viewed as exclusion and censorship and might even increase the interest in the banned sites, thus causing a so-called Streisand effect [4-40].

Finally, blocking access to certain websites based on a user's physical location or country of origin might very well be futile. As outlined in chapter 3, such a restriction can be easily circumvented by utilizing VPN clients and proxies. Successfully stifling access to the targeted websites seems unlikely as a result.

4.3.3. Liability of the Individual

An opposing approach to making platforms liable for content that was posted on them is to attempt to prosecute the individual who uploaded the illegal content in the first place.

This avoids having to use the platform as a scapegoat or legal arbiter as well as establishing direct consequences for transgressions in matters of hate speech and fake news. Additionally, it doesn't move the responsibility of prosecution away from the judicial branch.

This method may also face several issues, that can be divided into three categories: Detection, Identification and Prosecution.

4.3.3.1. Detection

Detection involves recognizing illegal content as such and commencing the legal process against the responsible individual.

As illustrated under 3.1 the sheer mass of data that gets posted onto social media daily makes a manual checking of everything nearly impossible. A system like that would require either automated checking of content before it is uploaded or an integrated report system for other users to utilize. Both of these would have to be implemented into all social media platforms. Furthermore, the intricacies of distinguishing hate speech and fake news from the legal and protected speech are difficult to program into an algorithm. If the platform is not the one responsible for checking whether a post is illegal,

then government or non-government agencies would have to be established to do so. Even with assuming that common definitions and outlines are accepted internationally, the number of media to check would likely be overwhelming. To reduce the workload the responsible agency could focus on content flagged by users, though even that is unlikely to make the number manageable.

4.3.3.2. Identification

Assuming a piece of uploaded content was found to be illegal and the poster needs to be held accountable, it would be required for the responsible individual to be identified without a doubt. Such an identification can be problematic through the internet for several reasons: On the one hand, users on social media platforms tend to make use of pseudonyms, rather than using their real names – and on the other hand, a perpetrator could be using someone else's profile, account or computer to mask their identity. While one could attempt to trace a user's IP address back to them, there are several ways for them to avoid being successfully tracked, as discussed in Chapter 3.

Several politicians and influential political figures have instead called for an enforced deanonymization of social media platforms with the hopes of increasing accountability on the internet.[262] [4-41] This would mean that a person could only register into social media (or other websites that allow users to interact with each other) using their real first and last name, possibly in addition to other personal data that can be used to identify them. The proponents of this method aim to improve internet culture by taking away a user's ability to conceal themselves.

Experts in the field are concerned that such a mandate could bring about a lot of undesirable side effects [4-42]. A common example is that an employee could no longer criticize their company or superiors without fearing retaliation.[263] Marginalized groups may face difficulties when wanting to express their opinions and voices – leading to a chilling effect that stifles freedoms of speech and free expression [4-43]. Additionally, users having to display their real names and possibly additional personal data could make it easier for them to get targeted by doxing – meaning that their private address and other sensitive information could be leaked into the internet, making them easy targets for retaliation, stalking or other crimes.

Studies on online culture have shown, that changes in tone are negligible between users using pseudonyms or their real names [4-44, 4-45, 4-46]. A user who wants to spread hateful speech will do so regardless of whether their real name is shown or not.

In the case of South Korea, a country that enacted a real-name policy on social media back in 2007, malicious comments on internet forums decreased only by 0.9%. At the same time, hackers were able to take advantage of the citizen's personal information being stored in the website databases for identification purposes. A single cyber-attack leaked the personal information of over 35 million Koreans – which amounted to more than half of South Korea's national population at the time [4-47].

---

[262] WITTENHORST, TILMAN, Gegen Hetze im Netz: Schäuble fordert Klarnamen-Pflicht, 2019, Available at https://www.heise.de/newsticker/meldung/Gegen-Hetze-im-Netz-Schaeuble-fordert-Klarnamen-Pflicht-4425451.html (Last accessed 28 January 2022).

[263] KELBERER, ULRICH, "Klarnamenpflicht im Netz vertreibt nicht den Hass, sondern unsere Freiheit", 2020, Available at https://www.focus.de/digital/internet/gastbeitrag-von-ulrich-kelber-eine-klarnamenpflicht-im-netz-vertreibt-nicht-den-hass-sondern-unsere-freiheit_id_11614881.html (Last accessed 29 January 2022).

Finally, the South Korean Constitutional Court declared the real-name policy unconstitutional in 2012 and it was abolished as a result.[264]

A recent ruling by Germany's federal court lines up with this analysis, as they support their citizen's rights to make use of pseudonyms – preventing Facebook from demanding real names from users that have been on the platform for longer than four years.[265]

Identification of Users on the internet has become more reliable over time with the help of algorithms, meta-data and machine learning,[266] but it remains difficult and unreliable when an individual knows how to conceal themselves.

4.3.3.3. Criminal Prosecution

Even when a post has been flagged and determined as either hate speech or fake news and the user who is responsible for that post has been positively identified, that is still no guarantee for prosecution or any consequences for that individual.

The internet is globally accessible, but perpetrators are not. As showcased in 2.2.3. of this chapter, a state's jurisdiction will rarely allow them to prosecute criminals beyond their borders. Those cases are rare for murderers, war criminals and other forms of wanted individuals.[267] [4-48] Expecting to be able to extradite someone who may or may not have posted hate speech on the internet seems optimistic at best.

Another thing to be considered is that the damage of hate speech campaigns or the spread of fake news is inflicted quickly. It has been shown that falsehoods are several times more likely to be shared on social media when compared to facts [4-49]. Thus, even if a smear campaign is shown to be false, the damage has already been done.

A prosecution across national borders, even if successful, would potentially be a years-long process. With the ubiquity of hate speech across social media, processing the cases would drain sizeable resources both in time and money.

In conclusion, attempting to prosecute for hate speech and fake news on an individual level would use vast amounts of resources and face numerous difficulties in the process – all for likely disappointing results.

4.4. Conclusion

There are numerous definitions and approaches to fake news and hate speech across different countries. This difference will potentially make it difficult to agree on an internationally uniform approach to resolving these problems.

---

[264] KYUNGHYANG, SHINMUN, Internet "Real Name" Law Violates the Constitution, Of Course, 2012, Available at http://english.khan.co.kr/khan_art_view.html?artid=201208241354087&code=790101 (Last Accessed 29 January 2022).
[265] BUDRAS CORINNA, Facebook muss Pseudonyme auf seiner Plattform dulden, 2022, Available at https://www.faz.net/aktuell/wirtschaft/digitec/facebook-muss-pseudonyme-auf-seiner-plattform-dulden-17756980.html (Last Accessed 29 January 2022).
[266] STOKEL-WALKER, Chris, Twitter's vast metadata haul is a privacy nightmare for users, 2018, Available at https://www.wired.co.uk/article/twitter-metadata-user-privacy (Last Accessed 29 January 2022).
[267] Ibid.

Legal remedies are hindered by the internet's global nature and the respective nation's lack of jurisdiction outside of its borders. A perpetrator outside of a legal authority's sphere of influence faces next to no effective consequences.

Resorting to perceived easy solutions, such as delegating the prosecution to the platforms themselves or blanket-blocking the access to them, leads to numerous unintended side effects. These include but are not limited to digital isolation, a heavy restriction on the freedom of expression and the citizens side-stepping the measures altogether.

Therefore, existing and intended legal remedies on a national legislative basis do not provide a feasible solution to combat fake news and hate speech on a global scale. It might be preferable to spend the resources on extensive education with digital media in conjunction with more steps towards political transparency. Open government initiatives might prove to be a better path towards mitigating the spread of fake news.

The next chapter will offer some insight into these initiatives and analyze their potential benefits.

References Chapter 4

[4-1]   UN GENERAL ASSEMBLY, International Convention concerning the Use of
         Broadcasting in the Cause of Peace (Geneva, 1936) Available at
         https://www.refworld.org/docid/3b00f0838.html (Last accessed 25 January 2022).

[4-2]   KALSNES, BENTE, Fake News, Kristiania University College, 2018, Available at
         https://doi.org/10.1093/acrefore/9780190228613.013.809 (Last accessed 05 November 2021).

[4-3]   LIESCHING, MARC, Das NetzDG in der praktischen Anwendung, 2021, P. 75, Available
         at https://library.oapen.org/handle/20.500.12657/48794 (Last accessed 20 January 2022).

[4-4]   CLAUSSEN, VICTOR, Fighting Hate Speech and Fake News. The Network
         Enforcement Act (NetzDG) in Germany in the context of European legislation, in: media
         laws 03/2018, Available at https://www.medialaws.eu/wp-content/uploads/2019/05/6.-
         Claussen.pdf (Last accessed 24 November 2021).

[4-5]   COUZIGOU, IRENE, The French Legislation Against Digital Information Manipulation
         in   Electoral Campaigns: A Scope Limited by Freedom of Expression, in Election
         Law Journal   Vol. 20, No. 1, 2021. p. 112, Available at
         https://doi.org/10.1089/elj.2021.0001 (Last accessed 29 November 2021).

[4-6]   SANZ-CABALLERO, SUSANA, The Principle of Nulla Poena Sine Lege Revisited:
         The Retrospective Application of Criminal Law in the Eyes of the European Court of
         Human Rights, in European Journal of International Law, Vol. 28, 2017, P. 788,
         Available at https://doi.org/10.1093/ejil/chx049 (Last accessed 19 January 2022).

[4-7]   KSHETRI, NIR, The Global Cybercrime Industry, Springer, 2010, P. 143.

[4-8]   BUSSMANN, KAI-D., Organisationen als Opfer, in BKA Reihe Polizei und Forschung,
         Viktimologie Deutschland, 2015, P. 395, Available at https://www.bka.de/SharedDocs
         Downloads/DE/Publikationen/Publikationsreihen/PolizeiUndForschung1_47_1_Viktimi
         sierungsbefragungenInDeutschland.html (Last accessed 30 December 2021).

[4-9]   SMITH, R., PERRY, M., Fake News and the Convention on Cybercrime, in Athens Journal of
         Law-Volume 7, 2021, P. 336, Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id
         =3878059 (Last accessed 10 January 2022).

[4-10]   KOTOWSKI, MATEUSZ, The Country of Origin Principle and the Applicable Law for
         Obligations Related to the Benefit of Information Society Services, in Adam Mickiewicz
         University Law Review, 11, 161-183, Available at
         https://doi.org/10.14746/ppuam.2020.11.09 (Last accessed 30 October 2021).

[4-11]   EUROPEAN PARLAMENT; Directive 2000/31/EC, 2000, Art. 3, Available at
         http://data.europa.eu/eli/dir/2000/31/oj (Last accessed 28 December 2021).

[4-12]   PAUL, RON, War on "Fake News" Part of War on Free Speech, Available at
         https://mises.org/wire/war-fake-news-part-war-free-speech (Last accessed 20 January 2022).

[4-13]   UNITED NATIONS, Strategy and plan of action on hate speech, 2019, p.1f., Available at https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml (Last accessed 30 December 2021).

[4-14]   FEDERAL OFFICE OF JUSTICE, German Criminal Code, 2019, § 130, Available at https://www.gesetze-im-internet.de/englisch_stgb/ (Last accessed 25 January 2022).

[4-15]   FEDERAL OFFICE OF JUSTICE, Law to improve law enforcement in social networks - Netzwerkdurchsetzungsgesetz – NetzDG, 2017, Available at https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html (Last accessed 13 January 2022).

[4-16]   EUROPEAN COURT OF HUMAN RIGHTS, Press Unit, January 2022, p. 1., Available at https://www.echr.coe.int/documents/fs_hate_speech_fra.pdf (Last accessed 13 January 2022).

[4-17]   FRENCH SENATE, Fight against hate on the internet, September 2021, Available at https://www.senat.fr/dossier-legislatif/ppl18-645.html (Last accessed 06 January 2022).

[4-18]   GERMAN BUNDESTAG, Scientific Services, Regulating hate speech and fake news on social media networks through selected countries, p. 9., 2019, Available at https://www.bundestag.de/resource/blob/662048/190949149266f3df2e27a0f098a53026/ WD-10-059-19-pdf-data.pdf (Last accessed 30 December 2021).

[4-19]   COUNCIL OF EUROPE, Reports of the Anti-Racism Commission on Estonia, Austria and the Czech Republic, 2015, Available at https://go.coe.int/K94ds (Last accessed 30 December 2021).

[4-20]   BRUGGER, WINFRIED, Ban on or protection of hate speech - some observations based on German and American law, 2019, Available at https://journals.tulane.edu/teclf/article/view/1662 (Last accessed 20 January 2022).

[4-21]   SUPREME COURT OF THE UNITED STATES, U.S. Reports: Brandenburg v. Ohio, 395 US 444 (1969), Available at https://www.loc.gov/item/usrep395444/ (Last accessed 19 January 2022).

[4-22]   UNITED NATIONS, Strategy and plan of action on hate speech, 2019, p.2., Available at https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml (Last accessed 30 December 2021).

[4-23]   EUROPEAN PARLIAMENT, Policy Department - Citizens rights and constitutional affairs, p. 126, 2015, Available at https://www.europarl.europa.eu/RegData/etudes/ STUD/2015/536460/IPOL_STU(2015)536460_EN.pdf (Last accessed 17 January 2022).

[4-24]   EUROPEAN COMMISSION, The EU Code of conduct on countering illegal hate speech online, 2019, Available at, https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en (Last accessed 14 January 2022).

[4-25]   GERMAN BUNDESTAG, Regulating hate speech and fake news on social media networks through selected countries, p. 11, 2019, Available at

https://www.bundestag.de/resource/blob/662048/190949149266f3df2e27a0f098a53026/WD-10-059-19-pdf-data.pdf (Last accessed 30 December 2021).

[4-26]   EUROPEAN PARLIAMENT, Hate speech and hate crime in the EU and the evaluation of online content regulation approaches, p. 53, 2020, Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/655135/IPOL_STU%282020%29655135_EN.pdf (Last accessed 10 January 2022).

[4-27]   EUROPEAN UNION, Consolidated version of the Treaty on the Functioning of the European Union – Part Three, Article 83 (ex Article 31 TEU), 2008, Available at https://eur-lex.europa.eu/eli/treaty/tfeu_2008/art_83/oj (Last accessed 07 January 2022).

[4-28]   COUNCIL OF EUROPE, European Commission against Racism and Intolerance ECRI, Available at https://rm.coe.int/leaflet-ecri-2019/168094b101 (Last accessed 08 January 2022).

[4-29]   EUROPEAN UNIVERSITY INSTITUTE, Handbook on Techniques of Judicial Interaction in the Application of the EU Charter: Freedom of expression and countering hate speech, pp. 21, 53, Available at https://cjc.eui.eu/wp-content/uploads/2020/05/eNACT_Handbook_Freedom-of-expression-compresso.pdf (Last accessed 08 January 2022).

[4-30]   FEDERAL OFFICE OF JUSTICE, German Criminal Code, 2019, § 5, Available at https://www.gesetze-im-internet.de/englisch_stgb/ (Last accessed 08 January 2022).

[4-31]   FEDERAL CONSTITUTIONAL COURT GERMANY, Decision of January 15, 1958 –1 BvR 400/51, Available at https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/1958/01/rs19580115_1bvr040051.html (Last accessed 09 January 2022).

[4-32]   BERLIN STATE CENTER FOR POLITICAL EDUCATION, Hate speech and fake news - questions and answers, p. 9, 2018, Available at https://www.amadeu-antonio-stiftung.de/w/files/pdfs/hate_speech_fake_news.pdf (Last accessed 20 January 2022).

[4-33]   EUROPEAN PARLIAMENT, The impact of algorithms for online content filtering or moderation, p.44, 2020, Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU(2020)657101_EN.pdf (Last accessed 26 January 2022).

[4-34]   HOUSE OF LORDS LIBRARY, Freedom of speech: Challenges and the role of public, private and civil society sectors in upholding rights, 2021, Available at https://lordslibrary.parliament.uk/freedom-of-speech-challenges-and-the-role-of-public-private-and-civil-society-sectors-in-upholding-rights/ (Last accessed 26 January 2022).

[4-35]   KORNMEIER, CLAUDIA, Gericht mit Kehrtwende im Fall Künast, 2020, Available at https://www.tagesschau.de/inland/kuenast-beleidigung-103.html (Last accessed 27 January 2022).

[4-36]   KÖVER, CHRIS, Interview zum Fall Künast - Dieses Urteil ist ein gutes Zeichen, 2020, Available at https://netzpolitik.org/2020/dieses-urteil-ist-ein-gutes-zeichen/ (Last accessed 27 January 2022).

[4-37]   HELDT, A. P., Intelligente Upload-Filter: Bedrohung für die Meinungsfreiheit, 2018, Available at https://www.ssoar.info/ssoar/bitstream/handle/document/57609/ssoar-2018-heldt-Intelligente_Upload-Filter_Bedrohung_fur_die.pdf (Last accessed 28 January 2022)

[4-38]   DREXL JOSEF, Bedrohung der Meinungsvielfalt durch Algorithmen, ZUM 2017, 529

[4-39]   SPOERRI, THOMAS, On Upload-Filters and other Competitive Advantages for Big Tech Companies under Article 17 of the Directive on Copyright in the Digital Single Market, 10 (2019) JIPITEC 173 para 1., Available at https://www.jipitec.eu/issues/jipitec-10-2-2019/4914 (Last accessed 27 January 2022).

[4-40]   CACCIOTTOLO, MARIO, The Streisand Effect: When censorship backfires, 2012, Available at https://www.bbc.com/news/uk-18458567 (Last accessed 28 January 2022).

[4-41]   SCOTT, JENNIFER, Can Online Safety Bill tackle social media abuse of MPs?, 2021, Available at https://www.bbc.com/news/uk-politics-58958244 (Last accessed 28 January 2022).

[4-42]   NEUERER, DIETMAR, Digitalverbände und Datenschützer warnen vor Klarnamenpflicht, 2019, Available at https://www.handelsblatt.com/politik/deutschland/cdu-vorstoss-digitalverbaende-und-datenschuetzer-warnen-vor-klarnamenpflicht/24446344.html?ticket=ST-3333381-QQCgAfzLxmTFvGCebieC-ap3 (Last accessed 29 January 2022).

[4-43]   SCHWANDER, TIMO, Das digitale Vermummungsverbot – eine irreführende Analogie, in ZRP 52/7 (2019), p. 207-208.

[4-44]   HAARKÖTTER, HEKTOR, Anonymität im partizipativen Journalismus, in: Zöllner, Anonymität und Transparenz in der digitalen Gesellschaft, 2015, p.133-149.

[4-45]   CARAGLIANO, DAVID, Real Names and Responsible Speech: The Cases of South Korea, China, and Facebook, 2013, Available at https://www.yalejournal.org/publications/real-names-and-responsible-speech-the-cases-of-south-korea-china-and-facebook (Last accessed 29 January 2022).

[4-46]   THE CHOSUNILBO, Real-Name Online Registration to Be Scrapped, 2011, available at http://english.chosun.com/site/data/html_dir/2011/12/30/2011123001526.html (Last accessed 29 January 2022).

[4-47]   KIM, KATE JEE-HYUNG, Lessons Learned from South Korea's Real-Name Policy, 2012, available at http://www.koreaittimes.com/news/articleView.html?idxno=19361 (Last accessed 29 January 2022).

[4-48]   SALUZZO, STEFANO, EU Law and Extradition Agreements of Member States: The Petruhhin Case, 2017, Available at https://www.europeanpapers.eu/en/europeanforum/eu-law-and-extradition-agreement-of-member-states-the-petruhhin-case (Last accessed 29 January 2022).

[4-49]   VOSOUGHI, SOROUSH, The spread of true and false news online, 2018, Available at https://www.science.org/doi/10.1126/science.aap9559 (Last accessed 29 January 2022).

5. Open Government & Open Data as a feasible solution?

Authors: Timo Vogt and Erik Wurzbach
Academic supervisor: Silvia Ručinská

5.1. Introduction

As shown in the preceding chapters, neither technical nor legal remedies seem capable of "saving us" of fake news and hate speech, at least not in a liberal and Human Rights-oriented regime like in the CoE Member States. The question arises, whether other remedies are (better) applicable. If it seems impossible to remove both hate speech and fake news like an unwanted weed from a flower bed, could it be more feasible to cover them with the more desired plants? In our situation, could a culture of more Open Data and Open Government, via increased transparency and subsequently accountability lead towards a reduction of people listening to hate speech and fake news?

The terms Governance, Open Governance, Open Government and Open Data and the term transparency are often confused. The authors attempt to clarify these terms and bring them into a comprehensive and understandable order, as you can hopefully agree on when having read this chapter, starting with principles of (Good) Governance.

5.2. Principles of Good Governance

In 2008, the Council of Europe published a list of 12 arguments that can be used as a guide for Good Governance. The 12 Principles are enshrined in the Strategy on Innovation and Good Governance at the local level, endorsed by a decision of the Committee of Ministers of the Council of Europe. They cover issues such as ethical conduct, rule of law, efficiency and effectiveness, transparency, sound financial management and accountability [5-1].

Good Governance is transparent, efficient and gives account to the entire population including all minorities. The participation of all subgroups of the population is paramount. All citizens are provided with all the services and public goods they need [5-2].

"12 Principles of Good Governance [COE]

1. Fair Conduct of Elections, Representation and Participation

- Local elections are conducted freely and fairly, according to international standards and national legislation, and without any fraud.

- Citizens are at the centre of public activity and they are involved in clearly defined ways in public life at the local level.

- All men and women can have a voice in decision-making, either directly or through legitimate intermediate bodies that represent their interests. Such broad participation is built on the freedoms of expression, assembly and association.

- All voices, including those of the less privileged and most vulnerable, are heard and taken into account in decision-making, including over the allocation of resources.

- There is always an honest attempt to mediate between various legitimate interests and to reach a broad consensus on what is in the best interest of the whole community and on how this can be achieved

- Decisions are taken according to the will of the many, while the rights and legitimate interests of the few are respected.

2. Responsiveness

- Objectives, rules, structures, and procedures are adapted to the legitimate expectations and needs of citizens.

- Public services are delivered, and requests and complaints are responded to within a reasonable timeframe.

3. Efficiency and Effectiveness

- Results meet the agreed objectives.

- The best possible use is made of the resources available.

- Performance management systems make it possible to evaluate and enhance the efficiency and effectiveness of services.

- Audits are carried out at regular intervals to assess and improve performance.

4. Openness and Transparency

- Decisions are taken and enforced in accordance with rules and regulations.

- There is public access to all information that is not classified for well-specified reasons as provided for by law (such as the protection of privacy or ensuring the fairness of procurement procedures).

- Information on decisions, implementation of policies and results is made available to the public in such a way as to enable it to effectively follow and contribute to the work of the local authority.

5. Rule of Law

- The local authorities abide by the law and judicial decisions.

- Rules and regulations are adopted in accordance with procedures provided for by law and are enforced impartially.

6. Openness and Transparency

- The public good is placed before individual interests.

- There are effective measures to prevent and combat all forms of corruption.

- Conflicts of interest are declared in a timely manner and persons involved must abstain from taking part in relevant decisions.

7. Competence and Capacity

- The professional skills of those who deliver governance are continuously maintained and strengthened in order to improve their output and impact.

- Public officials are motivated to continuously improve their performance.

- Practical methods and procedures are created and used in order to transform skills into capacity and to produce better results.

8. Innovation and Openness to Change

- New and efficient solutions to problems are sought and advantage is taken of modern methods of service provision.

- There is readiness to pilot and experiment new programmes and to learn from the experience of others.

- A climate favorable to change is created in the interest of achieving better results.

9. Sustainability and Long-term Orientation

- The needs of future generations are taken into account in current policies.

- The sustainability of the community is constantly taken into account.

- Decisions strive to internalise all costs and not to transfer problems and tensions, be they environmental, structural, financial, economic or social, to future generations.

- There is a broad and long-term perspective on the future of the local community along with a sense of what is needed for such development.

- There is an understanding of the historical, cultural and social complexities in which this perspective is grounded.

10. Sound Financial Management

- Charges do not exceed the cost of services provided and do not reduce demand excessively, particularly in the case of important public services.

- Prudence is observed in financial management, including in the contracting and use of loans, in the estimation of resources, revenues and reserves, and in the use of exceptional revenue.

- Annual budget plans are prepared, with consultation of the public.

- Risks are properly estimated and managed, including by the publication of consolidated accounts and, in the case of public-private partnerships, by sharing the risks realistically.

- The local authority takes part in arrangements for inter-municipal solidarity, fair sharing of burdens and benefits and reduction of risks (equalisation systems, inter-municipal co-operation, mutualisation of risks…).

11. Human rights, Cultural Diversity and Social Cohesion

- Within the local authority's sphere of influence, human rights are respected, protected and implemented, and discrimination on any grounds is combated.

- Cultural diversity is treated as an asset, and continuous efforts are made to ensure that all have a stake in the local community, identify with it and do not feel excluded.

- Social cohesion and the integration of disadvantaged areas are promoted.

- Access to essential services is preserved, in particular for the most disadvantaged sections of the population.

12. Accountability

- All decision-makers, collective and individual, take responsibility for their decisions.

- Decisions are reported on, explained and can be sanctioned.

- There are effective remedies against maladministration and against actions of local authorities which infringe civil rights." [5-1]

5.2.1. Why is the 4th Principle "Openness and Transparency" so important?

One of the basic principles of good governance is transparency. This means that the public should have a deep insight into the work of the public administration. Citizens should be able to scrutinise the work of the public administration and monitor it by providing tools to monitor the decision-making process. Furthermore, citizens should be familiarised with the rules that are applied in the exercise of their rights.

Transparency is important for the reform of public administrations. The goal is to fight corruption as well as to strengthen citizen participation. This is not possible without a sufficient level of information, which can only be obtained through transparent work.

Transparency and accessibility can become relevant in public administration in two ways. One is proactive transparency, which aims to make information public before the public calls for it actively. It takes the approach that all information of an administrative body that could be of importance to the public should be accessible. This theory holds the belief that there is a general right to publish relevant information [5-3].

The principle of openness and transparency makes government decisions easier to understand. As a result, conformity with the law can be maintained and proven before the citizens. Untransparent decisions are a thing of the past when this principle is observed. This deprives critics as well as opponents of the government of the basis for fake news and hate speech. With public access to all government information, every citizen has the same access rights and no one needs to feel excluded or disadvantaged. Of course, some data still needs special protection, for example when it comes to personal data. In our opinion, data protection should not be neglected even in the approach of openness and transparency. Concrete strategies should be developed on how data protection and transparency can go hand in hand.

It is particularly important to make the information easily accessible to everyone, bearing in mind that language barriers may exist for various reasons. As a result, the information must be provided in such a way that it can basically be read by everyone. Likewise, the data on the website should be easy to find to avoid long complicated searches, including falls also the announcement of the portal among the citizenry to increase the popularity and discoverability. In some cases, a notice on the homepage referring to the corresponding access or a notice in the town hall advertising the information available online is sufficient. Likewise, the data must also be made available to citizens without Internet access, but in this case, it is sufficient to provide the possibility of viewing the information in the town hall. All in all, it is important to make citizens feel that decisions have been made following all applicable regulations, and that data are provided in a nature that they can be easily understood.

5.3. Transparency

Congress recognized the importance of transparency by issuing Resolution 435 (2018) and Recommendation 424 (2018) together with an Explanatory Memorandum on 7 November 2018 [5-15].

In the political sense, transparency means making decisions known to the population and informing the population about political activities. The basic goal is to make important information public for everyone, to make the flow of money from public authorities and politicians verifiable (prevention of corruption). It also regulates activities that active politicians are prohibited from engaging in.[268]

5.3.1. Definition

At present, there is no generally valid definition of transparency. What is certain, however, is that transparency is a multi-layered concept that must be mentioned in the same breath as accountability, corruption, impartiality and the rule of law. In the narrower definition, transparency can be defined as the release of information relevant to the evaluation of various pieces of information.

Vishwanath and Kaufmann in 1999 define transparency as the increased flow of timely and reliable, economic, political and social information that is accessible to all relevant stakeholders. Thus, they emphasize not only the availability of information but also its reliability and accessibility to potential stakeholders.

---

[268] https://www.politik-lexikon.at/transparenz-transparenzgesetz/ (last accessed 25.01.2022).

Most literary definitions and statements on transparency today deal with the fight against corruption. However, accountability should also be addressed and improved governance should not be neglected. A key element is that the focus is not only on the provision of information but also on the ability of external actors to have access to it.

Thus, transparency can be defined as the availability and ability for internal as well as external actors to access and disseminate information. Stakeholders must be able to access information relevant to the evaluation of the institutions. This must happen concerning rules, procedures and results. The most widely used measures of transparency are the World Bank Governance Indicator or the TI Corruption Indicator. However, there is no universally valid measure (cf. [5-4], p. 5.).

5.3.2. The Six Faces of Transparency

Type A: Will formation

Type A transparency is intended to ensure in a society that citizens actively participate in the formation of public opinion. The concept of citizen participation, which has become increasingly important in modern times, falls under this type of transparency. To strengthen citizen participation, everyone must be able to have access to relevant information. Due to the broad scope, all relevant information held by public authorities falls under the obligation.

This lends itself to passive access for implementation, as it is the applicant who decides which information is to be released. Exceptions are determined by the possible violation of personal rights or if the refusal of an application has no or only limited effects on the formation of public opinion. Nevertheless, the formation of public opinion must not be impeded by the authority, as otherwise, the purpose of Type A transparency would have failed.

However, the exceptions must be accompanied by a reasonable justification even if they are rejected, otherwise, a false image could be achieved with the applicant.

Type B: Public participation and accountability

The purpose of Type B transparency is to ensure that the government and public authorities represent the public interest and implement the decisions taken through the public will. This can be achieved by allowing citizens to see and understand what the administration is doing. Here, too, the addressee of transparency must be the citizens directly. In contrast to Type A transparency, the scope here is limited and refers only to the actions of public authorities.

Under this point, only the information that directly concerns the actions of the authority is published; all other information does not fall under this type. The exceptions include information that, if published, could impede or restrict the actions of the public authority. In EU law, however, there is generally no presumption of a restriction on administrative action unless the public authority can put forward reasonable counter-arguments. For information to be used effectively, it should be disseminated to the public as early as possible. This kind of participation can make a positive contribution to democracy, but in certain cases, it can also have a paralysing effect on the decision-making process and the public interest. In principle, citizens should always understand what the authority has done. Even arguments against citizen participation are not in contradiction with the publication of information afterward and should never exclude it.

The accountability of public authorities can be exercised proactively, as this information is necessary for the performance assessment of the authority. The information should be presented in a simple and understandable manner for citizens to make good use of it.

Type C: Efficient decision-making

Type C transparency is intended to ensure that the quality of decisions and the associated improvement in the overall efficiency of the EU internal market increases. With this type of transparency, information is not made available to the general public, but only to economic actors who can use it to optimise their decisions.

By making it difficult to identify which information might be of interest to which economic operator, the opportunity should be taken to make the information publicly available to all operators. The scope of Type C transparency applies only to information that could potentially influence economic actors' decisions and affect the functioning of the market. The information can be either market regulator or market participant.

The time of provision must be such that it is still possible to act on it, i.e. it must take place before the measure is taken. An interpretation of the information must be easy for market participants to interpret and must be unambiguous. For the reason that the market participants do not know when information is generated, the authorities must approach the participants proactively. After notification, however, participants may be free to request further information or not.

Exceptions are difficult to justify because of the principle of equal treatment and the right to free movement. In EU law, these have a high priority and should not be lightly circumvented. It is questionable, however, whether transparency helps to improve efficiency. The EU institutions are often in a poor position to judge and are therefore often reluctant to impose strict obligations.

Type D: Compliance with economic law

Type D transparency is necessary for ensuring that public authorities comply with EU single market rules. It is possible to review the actions of public authorities and hold them accountable.

However, not all economic operators have been granted the right to transparency for valid reasons. The main interest of these is not compliant with EU law, but only their interests. While this is understandable, EU law provides the rule only for public authorities to comply with the applicable rules.

In public procurement law, however, transparency is not beneficial to the authorities, as economic actors can manipulate decisions here to their advantage, thus preventing fair competition.

In case of refusal of transparency by public authorities, the effects are limited. It is possible to give a commission the task of monitoring the behavior of member states.

Type E: Respecting the intrinsic worth of homo dignus

Type E transparency aims to facilitate autonomous decision-making. Transparency here is only required towards people whose rights it affects in terms of human dignity.

The information concerned is that which people need to make autonomous decisions in their private and family lives or to secure their human rights. Included is all information that governments possess, regardless of whether the authority needs the information for its work or only stores it. It is assumed that each person knows best when they need this information in their life, so it should be given out upon request.

An exception to the duty of providing information on request is difficult to justify because denying transparency means violating the rights of the individual.

Type F: Ensuring respect for homo dignus

Type F transparency ensures that public authorities respect people's dignity rights. One has the right to see what authorities do and hold them accountable for their actions. The right is to help people whose rights are affected by decisions made by authorities.

It is to be made transparent why a decision is made or why a procedure is initiated. Everyone who is affected has the right to participate in the procedure or to challenge it.

The quality of the information received should be such that it can be seen whether the authority respects the rights. In addition, the information should be given out as early as possible, so that the person concerned can influence the outcome of the procedure. With this form of transparency, the authority must act proactively, since this is the only way the person will learn about it.

A denial of this transparency endangers the substantive law and EU law on the transparency obligation the authorities have towards the data subjects, likewise it is a violation of the right to human dignity. It denies the data subject to fight for his due right. Delayed transparency is usually less harmful than a complete lack of transparency.

An early release of the information is better for the data subject than a delayed one since a delayed release could severely violate his or her rights. [5-5]

5.3.3. Benefits of Transparency[269]

• Build trust within your community

Citizens' trust in public authorities can be strengthened by making more information publicly available. A Gallup survey shows that local governments score significantly higher on trust than federal or state governments.

• Gain new ideas

Through an online forum, you can get citizens to participate with their ideas in the community. The advantages of an online forum are that all citizens can access data at any time and no one is excluded due to physical limitations.

• Increase community engagement

---

[269] https://icma.org/articles/article/top-10-benefits-transparency (last accessed 02.02.2022).

By using internal and external communication, you can build an engaged community. Internal communication is the exchange of information within an organization, external communication occurs when the organization communicates with external parties e.g. citizens.

- Understand your community's needs better

Measure effectiveness and your performance to better understand the needs of your citizens and your community.

- Empower citizens

Transparency in administrations increases the trust of the citizens and if the trust is high enough, the citizens feel responsible and it also leads to citizens identifying more with their administrations.

- Showcase reform

By highlighting growth and change, they can show citizens what has been done and where there is room for improvement. Through analysis and research, these can be clearly shown to citizens.

- Attract citizens to your government

The use of social media and geo-information systems can tremendously increase operational efficiency and also improve clarity for citizens. Again, transparency of map data and accessibility of information through social media is a good way to improve circumstances.

- Boost your economy

Transparency can be a good way to give a boost to the economy, which can be very important, especially in difficult times. Commissioning designs and advertising campaigns can create jobs for these industries. Information release portals also need to be created and designed by companies if this cannot be done in-house. A well-designed information system can create a trustworthy impression with the citizen.

- Foster a local government with professionalism

By promoting municipalities and cities, transparency can be expanded throughout the country. Through the resulting process of better information gathering for companies and potential new citizens, the economic power of municipalities can be strengthened. This is an effective and, above all, inexpensive means of bringing them into the municipality or city.

- Educate your citizens

Social media can be an important factor for a municipality's internet presence nowadays. Especially for the younger generation, who spend a lot of time on Facebook, Twitter and other portals, these platforms are a low-threshold way to reach them. In addition to the homepage, this kind of internet presence represents the future for simple information dissemination.

Reactive transparency, in contrast to proactive transparency, takes the approach of publishing knowledge only at the request of the public. Reactive transparency in this context means Freedom of Information like enshrined in many FoI legislative acts. The likely most renowned is the Freedom of

Information Act, 5 USC §552 et seq. (1966).[270] Note that the US Department of Justice operates the Office of Information Policy and publishes annual reports on FoI [5-60].

For us, transparency means that the processes in public administration and politics can be monitored by the public, at least on an aggregated level. Many people in today's world are surprised when procurement contracts are awarded, large-scale projects are prepared or even when a simple construction site is established in the neighbourhood. One recent example is the procurement of protective masks in the Corona pandemic in Germany, where the whole process was very non-transparent and far too many masks were ordered at very high prices and with corruption charges included. Such incidents diminish the belief in transparent and trustworthy administrative processes; after all, the entire expenditure is taxpayers' money. Particularly important for transparency is the trust of citizens in the date made transparent; without this, transparency cannot exist, since trust and transparency are directly linked. Once this trust is lost, it is difficult to regain.

In our view, there are six types of transparency, all of which have their justification, even if they are defined very differently in some cases. In the end, they provide an overall picture of transparency that covers all facets.

In today's world, transparency is one of the most important prerequisites for the functioning of a regulated democracy in which citizens have sufficient trust in the government. Every society should make more efforts to give transparency a high priority.

5.4. Open Government

5.4.1. A first short definition of Open Government

Different actors and policymakers can mean different things by open government and administrative action, which is influenced by political, social and cultural factors. Even though open government may be defined differently in different countries, the evidence suggests that government is open when it follows the principles of open government is open when it complies with the principles of transparency, accountability and participation. For the successful implementation of open government initiatives, it is important to have a consistent definition that is fully recognised and supported by the entire public sector and that is convincingly communicated to and supported by all stakeholders. Communicated convincingly to and accepted by all stakeholders.

Open government describes participatory, accountable and transparent government. The concept can be applied to any government size, locality, regionality and nationality are not important. Many regional authorities have already implemented reforms to open the government, not only to increase transparency to citizens but also to improve efficiency. The work of administrations should be traceable, which means that citizens should be able to follow what their government is discussing and producing at any time and from anywhere. Public authorities should also facilitate access to information and make it available through open data systems, as well as introduce procedures for the management of records. Open government also requires citizen participation that encompasses both government work and work in the civic space. To encourage and enable participation, care must be taken to prevent undue restrictions or potential repercussions of such activities. However, the information itself is also in need of protection. Accountability is the third pillar of open government, along with transparency and participation. Citizens should be able to hold their government

---

[270] Available at https://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/amended-foia-redlined.pdf (last accessed 10.11. 2021).

accountable for actions and results achieved. Accountability, in general, can be promoted, for example, through audits and scrutiny by civil society and the media.

The three pillars of Open Government (transparency, participation and accountability) can and should be applied to the 5 main functions of local government: Budgeting, contracting, lawmaking, policy-making and service delivery [5-7].

These three pillars are also referenced in the Explanatory Memorandum on Transparency and Open Government [5-15, p. 25].

5.4.2. Benefits of Open Government

• Increases transparency and accountability

The trend towards open data means that members of the public can stay connected, informed, and up to date with the day-to-day operations of their local government. The public nature of this information holds governments accountable to the results they produce. Residents can see exactly what their government has achieved, and how much more needs to be done. Failure to attain certain results or meet a particular milestone or goal will be publicized and up for public scrutiny. Conversely, achieving or exceeding goals will help to establish a greater and more trusting relationship with residents.

• Develops trust, credibility and reputation

The transparent nature of publicly accessible data exposes a side of an organization that is quite often kept under wraps. This sort of openness and vulnerability is comparable to sharing aspects of your personal life with another person. There is a considerable amount of trust and respect that comes with an open and honest conversation, and the result is quite often a closer and more dependent relationship between the two parties. In the same way, open government data helps to establish trust and credibility with citizens. Open data can give residents peace of mind that their local government is continually working to deliver on promises and making decisions in the community's best interests.

• Promotes progress and innovation

The value of key performance data has few bounds when set loose in the public sphere. Open data provides new opportunities for commercial applications, improves time-to-market for businesses, and can form the foundation for new technological innovation and economic growth. Third parties without the resources to gather this data for themselves will be able to re-purpose it and utilize the information to develop new applications and services. Information provided in this way is also significant for academic, public-sector, and industry-based research communities. Open data vastly increases the value of information and allows it to travel and be utilized to its full potential.

• Encourages public education and community engagement

What better way to educate the community on the progress and performance of the city than to have all the information displayed in a clear and user-friendly display? Open government data enables you to proactively answer those frequently asked questions by making the information freely accessible. Information can be made available as quickly as it is gathered, which means that the public can become involved and offer valuable feedback from throughout the entire process. Access to meaningful data aids in unifying a community and empowering them to help shape the direction for the future.

• Stores and preserves information over time

Finally, the availability of consolidated information in a single and easily accessible location is advantageous for the use of both current information and for historical data that has been gathered over time. This method of data storage ensures that all information will appear where and how it is supposed to, and that it will remain in that location for future reference. This also allows for the potential to observe trends and changes in the data over time.

Openness and transparency are two characteristics we greatly value at Envisio. Consequently, we have developed the Envisio Public Dashboard, enabling our customers a convenient way to share their strategic plan progress and performance measurement results with the public, harnessing all the benefits of open data in a single, easy-to-use platform.[271]

Today, Open Government is at least formally included in many governments' policies, but governments cannot rest on their status quo achieved but must continue to strive to improve in this regard. Especially in countries, where Open Government has not yet been widely adopted, a start should be made to give it a higher priority. Open Government having different meanings in different countries, complicates the uniform implementation of the principle. Nevertheless, the basic pillars of the Open Government movement should be adhered to in all countries. The variable applicability in terms of size, locality, regionality as well as nationality means that reform can be initiated without major adjustments.

Open Government is not a project of the central government, but can also find its way into local and regional authorities. In this way, citizen-oriented implementation can be guaranteed and credibility gained, even from the lowest level.

The three-pillar model of transparency, participation and accountability should be applied at all levels to the five main functions of government. In this way, citizens' trust in the administration can be improved.

The numerous benefits of Open Government for administrations are unmistakable. The most important benefits are the increase in transparency and accountability as well as the increase in trust and credibility, closely followed by progress and innovation. An administration that aspires to be among the best cannot do without these benefits of Open Government if it wants to offer its citizens the best possible service.

5.5. Open Data

5.5.1. Definition

Open data is the publication of data and information in a format that may be freely used, modified and shared. The OECD states that open data is *"a set of policies that promote transparency, accountability and value creation by making government data available to all"*. By making data generated through the activities of public bodies available, government becomes more transparent and accountable to citizens. It also supports business growth and the development of services centred on citizens, and provides important data for research and innovation by public bodies, the private sector, and civic stakeholders [5-8].

---

[271] https://envisio.com/blog/5-benefits-of-open-government-data/ (last accessed 02.02.2022).

Congress recognized the importance of Open Data by issuing Resolution 417 (2017) and Recommendation 398 (2017) together with an Explanatory Memorandum on 28 February 2017 [5-16].

5.5.2. Benefits of Open Data

Performance can be improved through Open Data and contributes to improving the efficiency of public services: Thanks to the provision of data across sectors, better efficiency can be achieved in procedures and the delivery of public services, for example by providing an overview of unnecessary spending.

Social security can be improved, as the society can benefit from the information that is more transparent and accessible. Open data improves collaboration, participation and social innovation.

Transparency can be increased: Open data increases transparency on the part of the Public Administration towards citizens, but also towards other administrations. Transparent government behaviour is a foundation for trust and collaboration.

Stores and preserves information over time: Finally, the availability of consolidated information in a single and easily accessible location is advantageous for the use of both current information and for historical data that has been gathered over time. This method of data storage ensures that all information will appear where and how it is supposed to, and that it will remain in that location for future reference. This also allows for the potential to observe trends and changes in the data over time.[272]

Open data is publicly available data from authorities and other institutions that can be freely used by everyone, access to it should neither be prohibited nor prevented, even if it could lead to opposing opinions. Everyone must have the right to form his or her own opinion and to represent it. The right to use this data to form opinions is unrestricted and transcends national boundaries. Especially in the context of freedom of the press, open data plays an important role, as it is made accessible by the press to the less interested citizens. But also each individual must have access to the data without any intermediary institution.

Open data is equally important for both the private and the public sectors, as both can benefit from this data. Be it in the area of research or the further development of processes, in all areas this data can be used purposefully.

The public sector can further use the data to improve the performance of its administration; specifically, the data can be used to increase the efficiency of processes and services. In addition, things like social security, increasing transparency and pinpointing redundant spending can be improved. However, the information is also needed for historical archiving purposes. This data must be prepared in a way that is understandable to everyone and must also be easily accessible. In this way, the broad mass can be reached through Open Data.

5.6. Issues with Open Data/Open Government/Transparency

Data protection problems are among the most common risks associated with increasing transparency. To enable more transparency, legislators should adapt the laws accordingly. The aim should be to do equal justice to both sides. Under no circumstances should personal data become the victim of increased transparency.

---

[272] https://envisio.com/blog/5-benefits-of-open-government-data/ (last accessed 02.02.2022).

However, data from public authorities and agencies should be disclosed as far as possible, and concrete guidelines should be defined for this purpose. These should regulate which data must be published and which should be kept under lock and key. These must be understandable and comprehensible to citizens. Citizens must also have the opportunity to check compliance with the guidelines. If this is not ensured, a feeling of false transparency and misuse or misappropriation can arise in the citizen.

This should not occur under any circumstances, as lost trust is difficult to regain.

When publishing data, good clarity and comprehensibility must be ensured, otherwise citizens will not be able to make use of this data. [5-14]

5.7. The distinction between Open Government and Open Data

For some years now, the terms Open Government and Open Data have gained considerable importance. However, many do not know in what the two terms differ. In the following, we will discuss these differences. Open Government is intended to promote the revitalization of democracy through the disclosure of government and administrative data. This can lead to new forms of cooperation by government, politics and administration with citizens and civil societies. Open Government should lead to citizen-oriented administrative action, better legitimized political decisions and improved cooperation between the state and society.

In contrast to this is Open Data, which is a part of Open Government. In practice, Open Data is aimed at transparency, free access and the dissemination and use of official data. This includes, among other things, budget data, maps or other documents of the authorities.

Open Data follows the approach of making data of all kinds publicly accessible without restrictions, structured and machine-readable.

The big difference is that Open Data does not only refer to the public administration but also includes other data, which can come from private companies or non-governmental organizations [5-10].

A term often mentioned in the context is Open Government Data, which is used for administrative and governmental institutions.

To get a better understanding of Open Government Data, the three components should be considered separately. The basics are the terms "open", "government" and "data". As shown in the diagram, this results in three overlapping terms: government data, open data and open government.
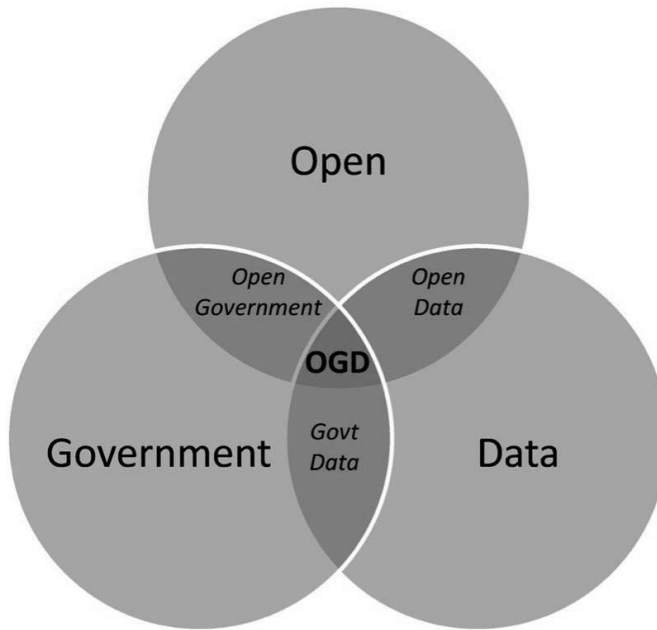
Figure 51: Open, Government, Data[273]

Government data refers to data sets held by the state as the largest collector of information. This includes, for example, information about citizens, organizations and public services. One concern of the state is to manage the information for the public sector in the best possible way.

Open data has its origins in information and communications technology and is intended to ensure that data is more easily accessible to citizens. The main aim here is to expand accessibility and accessibility to information.

Open government is intended to make government decisions and their measures more transparent and easier for citizens to understand. These measures are intended to find ways of empowering citizens and civil society organizations by opening up the administration.

The term Open Government Data is composed of these three approaches. In the general literature, Open Government Data is divided into four different perspectives.

First is the bureaucratic perspective, which is closely related to government data. In this one, OGD is understood as a government policy to support public services for improved handling of government data. Improved handling refers to internal government changes made by officials and staff. This refers to regulations, policies, and processes aimed at making government data handling more efficient and effective. Further, the opening of government data is used to help reduce costs and improve process quality.

---

[273] https://www.researchgate.net/publication/337398045/figure/fig1/AS:963465296502789@1606719435768/Open-government-data-foundations-Yu-and-Robinson-2011.png (last accessed 14.01.2022).

The second is the technological perspective, which is strongly associated with the idea of open data. This perspective considers OGD as a technological innovation made by changes in technical staff of information technology. Here, the design of formats, processes, and standards used to process public data is made. These changes are aimed at improving the data infrastructure within the administration. In this, the data must have a basic quality, which includes distributive aspects such as free availability, reusability or operability. This should enable easy availability for all actors.

The political perspective is linked to the ideas of open administration, where OGD is understood as a fundamental right for all citizens. This is intended to guarantee access to public sector data. Through the more open use of data, better governance should be possible. The data should help to increase transparency and enable greater participation by citizens in public sector decisions. This refers, for example, to policymaking and reducing imbalances between government and citizens.

The economic perspective has itself evolved from OGD; it is used to promote economic growth. By using freely available public sector data, businesses can act and react better. Likewise, it is intended to promote the creation of new products and services and to create new jobs. In this way, profits can be increased and investments improved. [5-11]

5.8. Transparency as a suitable way to avoid or reduce Fake News

The goal of full transparency requires the general belief that democracy is the rule of the people, and that representatives elected by the people exercise decision-making only temporarily, but are accountable to the citizens. In this sense, transparency requires public authorities not to put citizens on an equal footing with decision-makers, but to make information available at the same time as the rest of the administration [5-3].

This means, in our context, that a policy of Open Government proactively must provide data before it is explicitly requested and, if requested, deal with these requests in an adequate manner.

Our underlying assumption is, that hate speech and fake news can be more often perceived in an environment where there is no or little trust in government and authorities. This trust should be built and strengthened by the authorities through an Open Government Data strategy. This is how we believe the frequency can be reduced. This trust should be built and strengthened by the authorities through an Open Government Data strategy. This is how we believe the frequency can be reduced.

This assumption seems not to be explicitly proven in literature, but the following studies/publications provide hints in this direction:

Trust in governments is described by many studies as the cornerstone of democratic stability. Distrust in government action slows down progress and the functioning of a state. Statistically, trust in governments has declined in recent years. Understandably, there is now a desperate search for a way to reverse this trend and restore faith in governments. It is widely believed that transparent governance can greatly improve the credibility of government action. The disclosure of internal work processes and information should help to make government performance measurable and more comprehensible [5-18].

In the following part of the text, we have listed some existing studies on the topic of transparency and its effects, which we think provide helpful input. We are fully aware that this is no full proof of our assumptions – rather single observations, which point into this direction.

5.8.1. Questionnaire Buenos Aires

Buenos Aires is the capital of Argentina with almost 3 million inhabitants. Although the city has existed since the 16th century, citizens have only been able to elect a head of government since 1996, as he or she was previously appointed directly by the president. The administrative division of the city into 48 districts and 15 municipalities has greatly increased the decentralization of administration and citizen participation. Within this process, attempts have also been made to improve the transparency of government action. The mayor made a number of promises to the people when he took office to improve transparency. These are measurable targets based on the United Nations Development Goals. The city has received several awards for this commitment. Today, all of the more than 50 goals can be viewed and tracked transparently on the municipality's website.

Realization:

The Buenos Aires experiment was conducted by providing participants with equal probability information on a series of commitments promised by the mayor of the city of Buenos Aires when he took office. These were then followed up by making the results on these publicly available on the city's website.

The next step was to randomly assign a project that either highlighted a government promise of efficiency and good governance or contained a profound message with a government promise to improve the lives of residents. In this context, participants were provided with information that demonstrated either the fulfilment or non-fulfilment of the commitments.

A multidimensional approach was used to assess trust perceptions, encompassing all components of trust. These include competence, benevolence and honesty.

The interesting result for our project was that providing the information increased the perception of government transparency by about eight percentage points.

Further, it became clear that differences in performance can play a major role in trust in government. The group that received information about the government exceeding its targets subsequently showed significantly higher trust than the group that received information about the government not meeting its targets.

These findings underline the importance of providing transparent information to citizens. Furthermore, the results show that the way the message is expressed does not play an overriding role. Rather, the content of the message is relevant, especially whether the government delivers on its promises and goals.

Every government should draw the lesson from this study to do even more to achieve its goals, because the citizens honor this with higher trust. The example of a study illustrates this very well. If the government only publishes the results that are positive for it, the credibility of the good performance decreases. If the deception is exposed, the credibility in the government diminishes extremely and the credibility in other studies also diminishes. [5-19]

5.8.2. Threats of violence and harassment against politicians

For several years, there has been an increase in attacks and assaults on members of parliament and politicians at local, national and international levels. Most recently, a British Conservative MP was stabbed to death in his constituency. What makes this case even more frightening is that this did not happen without predictability, as there have been several other cases of this or similar nature in recent years.

However, such incidents are becoming more frequent and threats of murder and violence are unfortunately becoming almost daily affairs for many politicians.

Due to the Brexit and the COVID-19 pandemic, the existential fears of the British population are increasing and thus also anger and blame. An audit by the Hansard Society concluded that "opinions about the system of government have reached their lowest point in the 15-year audit series - worse than after the MPs' expenses scandal".

In the 2017 UK general election, 56% of parliamentary candidates surveyed said they were concerned about the level of intimidation they had experienced, with 31% saying they felt anxious during the campaign. Especially on the Internet, with the help of anonymous social media accounts, there are regular threats of all kinds of violence.

A study currently underway on trust and governance in five democracies around the world dramatically illustrates the result of the 2017 British general election. Nearly 40% of respondents could name at least one instance of abuse or threat of violence [5-13].

5.8.3. Corona vaccination in Portugal

The rate of fully vaccinated population in Portugal is 87.78% (as of 19 November 2021), making it the absolute leader in Europe, while globally only the United Arab Emirates and Singapore have a slightly higher vaccination rate. Far behind Portugal at the European level are Spain (about 79%) and Denmark (about 76%). France and Germany are even further behind at around 69% and 67% respectively.

It is questionable why Portugal performs so much better than the rest of the European countries because it is doubtful that the much higher vaccination rate is just a coincidence.

In the early summer of 2021, Portugal and Germany were still tied in vaccination rates, but vaccination fatigue has not set in in Portugal after initial successes. On the one hand, one could say that the Portuguese will to be vaccinated is due to the high infection and death wave at the beginning of the year. The Corona situation was completely out of control, the German army came to the rescue, and in one week at the end of January, 2 000 people succumbed to the virus.

The Portuguese vaccination coordinator Henrique de Gouveia e Melo explains that all Portuguese are pulling together after this traumatic experience. The vaccination process was also discussed very actively and openly. Transparency was at the forefront here, so that scandals, such as those that occurred with the procurement of masks in Germany, do not occur. Without such serious confidence-reducing actions by politicians, the Portuguese people's trust in the vaccination campaign was not endangered. So, there was no reason to doubt the vaccinations. The government made citizens aware from the beginning that although the vaccination had side effects, these were far milder compared to a severe course.

No small vaccination centers were set up, but large sports facilities were used, and every citizen was personally asked to be vaccinated at least three times, and those who did not respond were repeatedly contacted and reminded. E. Melo also claims to have made it clear to citizens that they were in a war

against the virus and that children needed to be protected from it. The Portuguese also rely heavily on their health system, introduced in 1970, and vaccination rates for measles, rubella and mumps are higher in Portugal than in almost all other EU countries. [5-14]

5.9. Conclusion

The possible remedies outlined here, in brief, are not validated yet, hence rather input for research projects and verification. The authors consider it likely that more Openness, both in terms of Government, Data and overall transparency could decrease the level of hate speech and fake news or at least significantly lower the portion of the electorate falling for that.

The survey we undertook and which is analyzed in detail in Chapter 6, supported our views and assumptions on the effectiveness of Open Data, Open Government and Transparency. Question 7.2 "Which political measures of your institution do you consider a viable option against fake news and hate speech?" showed that nearly 81 percent considered the approach of Open Data and more transparency in political decisions to be the most sensible means as a measure against fake news and hate speech. This was closely followed by a better explanation of decisions (approx. 75 % of votes), which is inextricably linked to Open Government Data, as it calls for data to be published in such a way that it can be understood by everyone.

That citizens should be involved in decision-making was supported by approx. 63 % of the votes) whilst more offline contacts should be maintained with citizens were supported by approx. 60 % of the votes.

These answers indicate, that Open Government and Open Data together with more transparency could be a feasible remedy – which of course must be analyzed and, if possible, verified by further research.

The authors are quite convinced that, as it was shown above, legal and technical remedies will be rather a placebo than a real remedy, because they can neither become universally enforced nor implemented without significant harm to the whole internet.

We are well aware that this situation is not satisfying for those people, especially local and regional politicians, who are confronted with hate speech and fake news daily. We call for the governments and civil societies of Europe to enable further research and political discussions on these topics.

If we succeed in increasing trust in government and authorities, this could, in our thinking, lead to a reduction of hate speech and fake news produced from within the society – let aside hostile attacks from third parties like other countries.

References Chapter 5

[5-1]   Council of Europe, 12 Principles of Good Governance. Available at
          https://www.coe.int/en/web/good-governance/12-principles (last accessed 25.01.2022)

[5-2]   Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung:
          Rahmenbedingung für Entwicklung – Good Governance. Available at
          https://www.bmz.de/de/entwicklungspolitik/good-governance (last accessed 25.01.2022)

[5-3]   Belgrad Open School: Transparency as a Principle of Good Governance. Available at
          https://bos.rs/du-eng/transparency/931/2017/06/29/transparency-as-a-principle-of-good-
          governance.html (last accessed 25.01.2022)

[5-4]   Monika Bauhr, Marcia Grimes: What is Government Transparency? In: Working Paper
          Series 2012:16, QOG, Department of Political Science, University of Gothenburg.
          Available at https://www.gu.se/sites/default/files/2020-05/2012_16_Bauhr_Grimes.pdf
          (last accessed 25.01.2022)

[5-5]   Buijze, A., 2013. The Six Faces of Transparency. Utrecht Law Review, 9(3), pp.3–25.
          DOI: http://doi.org/10.18352/ulr.233 (last accessed 27.01.2022)

[5-6]   U.S. Department of Justice, Annual FOIA Report 2020, Available at https://www.
          justice.gov/oip/department-justice-annual-foia-report-fy20 (last accessed 02.01.2022)

[5-7]   Council of Europe, 35th Session; Transparency and open government. Available at
          https://rm.coe.int/transparency-and-open-government-governance-committee-rapporteur-
          andre/16808eca29 (last accessed 27.01.2022)

[5-8]   DATA.EUROPA.EU, The benefits and value of open data. Available at https://data.
          europa.eu/en/highlights/benefits-and-value-open-data (last accessed 02 February 2022)

[5-9]   Johann Höchtl, Martin Kaltenböck, Peter Parycek, Judith Schossböck, Thomas Thurner:
          Open Government Data: Potentiale, Risiken und Hürden. In: Proceedings of 41.
          Jahrestagung der Gesellschaft für Informatik, 4.-7.10.2011, Berlin. Available at
          https://www.user.tu-berlin.de/komm/CD/paper/061121.pdf (last accessed 27.01.2022)

[5-10]   Emmie Tran, Ginny Scholtes: Open Data Literature Review. Available at:
          https://www.law.berkeley.edu/wp-content/uploads/2015/04/Final_OpenDataLitReview
          _2015-04-14_1.11.pdf (last accessed 27.01.2022)

[5-11]   Mila Gascó-Hernández, Erika G. Martin, Luigi Reggi, Sunyoung Pyo, Luis F. Luna-
          Reyes: Promoting the use of open government data: Cases of training and engagement.
          In: Government Information Quarterly Volume 35, Issue 2, pp.16-19. April 2018,
          Elsevier. Available at https://www.sciencedirect.com/science/article/abs/pii/
          S0740624X17302824?via%3Dihub (last accessed 27.012022)

[5-12]   Martin Alessandro, Bruno Cardinale Lagomarsino, Carlos Scartascini, Jorge Streb, Jerónimo
          Torrealday: Transparency and Trust in Government. Evidence from a Survey Experiment.
          In: World Development Vol. 138 February 2021, Elsevier. Available at

https://www.sciencedirect.com/science/article/pii/S0305750X20303508 (last accessed 27.01.2022)

[5-13]   James Weinberg: David Amess killing: threats of violence and harassment have become commonplace for politicians. In: THE CONVERSATION. Available at https://theconversation.com/david-amess-killing-threats-of-violence-and-harassment-have-become-commonplace-for-politicians-170078 (last accessed 29.01.2022)

[5-14]   Lea Schulze: Darum hat Portugal ohne Zwang alle im Impftempo überholt. In: DER TAGESSPIEGEL. Available at https://www.tagesspiegel.de/politik/an-die-weltspitze-geimpft-darum-hat-portugal-ohne-zwang-alle-im-impftempo-ueberholt/27695424.html (last accessed 29.01.2022).

[5-15]   Congress of Local and Regional Authorities, 35th Session, Report CG35(2018)14final, 7 November 2018, Available at https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016808d341c (last accessed 22.02.2022)

[5-16]   Congress of Local and Regional Authorities, 32nd Session, Report CG32(2017)15final, 28 February November 2017, Available at https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806fbdbc (last accessed 22.02.2022)

6. Empirical Analysis

Author: Alexander Prosser

6.1. Questionnaire

6.1.1. Background

The theoretical considerations called for an empirical validation among the members of congress (MC), which focused on three topics:

-   How do MCs see fake news and hate speech, how do they define it? This was done in Questions 1 and 2 building on Recommendation No. R(97)20 of the Council of Europe Committee of Ministers to the Member States on "Hate speech".

-   What is their experience with both phenomena? Questions 3 to 6 deal with that.

-   What countermeasures are recommended (Questions 7 and 8)?

Question 9 requested some general information. The questionnaire was established in close coordination with the General Secretariat of the Congress and implemented in an online tool at the University of Ludwigsburg, evasys[274]. The links were sent to MCs, Congress partner organisation delegates and youth delegates. The online questionnaire was open from December 13, 2021 till February 1, 2022.

187 questionnaires were returned of whom (types of respondents)

-   137 came from MCs;

-   17 from partner organisation delegates and

-   32 came from youth delegates.

One questionnaire was not attributed. The empirical analysis was done in IBM SPSS 28.[275] For verification and analysis the dataset and the spool files used in this study are made available from December 11, 2021 to January 31, 2022.

The following Section 1.2 presents the findings question by question and Section 1.3 attempts to relate data to one another. Section 1.4 presents a summary of the empirical findings and makes recommendations based upon them.

---

[274] https://evasys.de/evasys/
[275] https://www.ibm.com/analytics/spss-statistics-software The analysis was done in a German version of the software, hence on some occasions the German-language descriptors appear in the results copied into the text.

6.1.2. Descriptive Results

General caveat: Due to the number of respondents (less than 200), the confidence intervals of the descriptive analysis are relatively high and may only serve as an indicator.

6.1.2.1. Definition of Fake News

Question 2 was designed to elicit, what delegates understood as "fake news". The results are presented in Table 4 (empty entries omitted, no distinction between types of respondents):

No.	Text (English)	Yes, in %	No, in %
2.1	Verifiably false information that is disseminated with malign intent.	92.5	2.1
2.2	Verifiably false information that is disseminated bona fide.	69.0	15.5
2.3	Verifiably true information that is presented out of context or disproportionately.	56.7	27.3
2.4	Verifiably true information that is disseminated with malign intent.	41.7	43.9
2.5	Dissemination of information that can neither be verified nor falsified at the time of dissemination.	46.0	25.1

Table 3. Respondents specify what they see as "fake news", n=187.

Some of these results are worth noting:

-   The share of respondents classifying verifiably false information that is disseminated bona fide is more than one-fifth lower than the share of respondents classifying the dissemination of false information with malign intent as fake news. This shows that the *intention* with which false information is disseminated plays a role in the perception of fake news.

-   The dissemination of true information either in a distorted way or with malign intent is also classified as fake news by more than half of the respondents and over 40%, respectively. By inverse logic, 60% of the respondents say that it is fair and non-fake news to disseminate true information even with malign intent.

-   Dissemination of information that cannot be verified as true or false is considered fake news as well, irrespective of the intention by almost half of the respondents.

This shows that despite the somewhat ubiquitous usage of the term "fake news", there are substantial differences in the perception of what is fake information. The intent of dissemination seems to play an important part in this perception.

6.1.2.2. Personal Experience with Hate Speech

Questions 3 and 4 asked about whether and to what extent hate speech was experienced by the respondents (missing values omitted).

Question 3: Have you experienced hate speech in the above definition?

No.	Text (English)	Hardly ever, in %	At times, in %	Frequently, in %
3.1	Personally	42.2	44.9	11.8
3.2	Members of your city council / regional assembly	31.0	50.8	16.6
3.3	Our institution	46.0	42.2	9.6

Table 4. Personal experience with hate speech, n=187 (non-respondents not shown).

More than one out of 10 delegates frequently experiences hate speech on a personal level, well over a half either at times or frequently. This is quite a depressing result and may be one explanation why it is getting increasingly difficult to recruit political representatives on local and regional level.

Institutions as such, however, appear to be less prone to become the target of hate speech. The central takeaway here is that hate speech is something eminently personal – and it is not a fringe phenomenon.

6.1.2.3. Extent and Manifestation of Hate Speech

Questions 4 investigated the extent and manifestation of hate speech experienced by the respondents; questions are sorted in what the authors considered an increasing level of severity.

Question 4: Extent of hate speech (either 3.1 or 3.2, not the institution)

No.	Text (English)	Hardly ever, in %	At times, in %	Frequently, in %
4.1	Personal insults in media	34.2	46.0	14.4
4.2	Libel in media	40.6	39.6	13.9
4.3	Material damage in media (eg. cyberattacks against homepage)	67.9	20.3	4.3
4.4	Physical threats in media against the person addressed	56.1	31.0	5.3
4.5	Physical threats in media against the family of that person	66.3	22.5	3.2
4.6	Personal insults in the real world	40.6	45.5	7.5
4.7	Libel in the real world	44.4	39.6	9.1
4.8	Material damage in the real world	67.9	22.5	2.7
4.9	Physical violence in the real world against the person addressed	73.8	17.1	2.1
4.10	Physical violence in the real world against the family of the person addressed	75.9	13.9	2.7

Table 5. Extent and manifestation of hate speech, n=187 (non-respondents not shown).

Considering only the "frequently" answers, some patterns can be recognised:

- Personal insults and libel are the top scorers both in the digital and the real world.

- However, one fifth has encountered physical violence against themselves and one out of six against their families, in the real world either frequently or at times.

- On average, about half as many respondents were subjected to hate speech/acts in the digital media than in the real world. However, filtering for those respondents who replied with "frequently" in questions 4.1 to 4.5 (digital world) shows the following picture:[276]

Filtering for respondents indicating 4.1 = frequently				
4.6	Personal insults in the real world	33.3	33.3	33.3

Filtering for respondents indicating 4.2 = frequently				
4.7	Libel in the real world	15.4	38.5	46.2

Filtering for respondents indicating 4.3 = frequently				
4.8	Material damage in the real world	37.5	37.5	25.0

Filtering for respondents indicating 4.4 = frequently				
4.9	Physical violence in the real world against the person addressed	50.0	10.0	40.0

Filtering for respondents indicating 4.5 = frequently				
4.10	Physical violence in the real world against the family of the person addressed	33.3	0.0	66.7

Between 50% and 80% of those who "frequently" received threats in the virtual sphere were also attacked "frequently" or "at times" in the real world. It has to be re-emphasised that Question 4 only asks for the personal experience of the respondent or his/her colleagues from the representative body, not against the organisation. One may hence confirm the oft-used dictum that verbal abuse regularly leads to physical violence.

6.1.2.4. Personal Experience with Fake News

In a very similar way to hate speech, Questions 5 and 6 explored the extent to which Members of Congress were subjected to fake news (Question 5) and what form of fake news they experienced.

---

[276] With a caveat as to the small number of cases.

Question 5: Have you experienced fake news in the above definition?

No.	Text (English)	Hardly ever, in %	At times, in %	Frequently, in %
5.1	Personally	29.4	47.6	21.9
5.2	Members of your city council / regional assembly	28.9	54.5	15.0
5.3	Our institution	43.9	44.9	9.6

Table 6. Personal experience with fake news, n=187 (non-respondents not shown).

The percentage of respondents who have "frequently" experienced fake news personally is about twice as high as the percentage with a "frequent" experience with hate speech. However, the "frequently" responses concerning other members of the representative body and the institution itself are about the same as with hate speech. This looks odd and warrants closer investigation. One explanation may be that experience with hate speech is more readily shared among representatives than experience with fake news.

Naturally, the question arises, whether experience with hate speech and fake news correlates, particularly on a personal level. Since both variables are ordinally scaled, an X^2 test is the method of choice, the result is shown below (n=184 valid cases).

Q3.1 * Q5.1 Cross tabulation

			Q5.1			
			1	2	3	Gesamt
Q3.1	1	Anzahl	41	32	6	79
		% von Q3.1	51,9%	40,5%	7,6%	100,0%
	2	Anzahl	12	54	17	83
		% von Q3.1	14,5%	65,1%	20,5%	100,0%
	3	Anzahl	1	3	18	22
		% von Q3.1	4,5%	13,6%	81,8%	100,0%
Gesamt		Anzahl	54	89	41	184
		% von Q3.1	29,3%	48,4%	22,3%	100,0%

Chi-Square-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)
Pearson-Chi-Quadrat	78,613[a]	4	<,001
Likelihood-Quotient	70,133	4	<,001
Anzahl der gültigen Fälle	184		

Table 7. X^2 test personal experience fake news x hate speech

The result is unequivocal and on a significance level beyond 99.9%[277]:

Being subjected to hate speech and fake news strongly correlate. One entails the other. Therefore, one may also reject the consideration that fake news is the more "harmless" phenomenon here as compared to hate speech. The result suggests both are two sides of the same coin.

6.1.2.5. Extent and Manifestation of Fake News

Question 6 dealt with the form of fake news the respondents experienced.

No.	Text (English)	Hardly ever, in %	At times, in %	Frequently, in %
6.1	As part of hate speech	41.2	38.5	17.1
6.2	To influence decision making in our municipality / region	30.5	52.4	13.9
6.3	To influence elections for our city council / regional government	34.2	40.6	22.5

Table 8. Extent and form of fake news experience, n=187 (non-respondents not shown).

There is a clear tendency towards the use of fake news to influence elections. More than one-fifth of respondents indicate that they frequently experience fake news as part of an electoral campaign. This indicates that such interference is not an exception to the rule, but rather a commonplace occurrence. More research in this area is indicated as failure to ensure the integrity of the elections is a major issue in a democracy [1].

6.2. Countermeasures

Question 7 of the questionnaire inquired about suggested countermeasures, some of which are technically not feasible or at least not feasible in a non-police state. The following table lists the results and the technical feasibility.

6.2.1. Technological and Legal Remedies

Question 7.1 inquired about the proposed technological methods:

Which measures would you consider a technically and legally viable option against fake news and hate speech?

---

[277] In statistical tests, typically the hypothesis of independence is tested. The significance level indicates the probability with which the hypothesis of independence can be rejected – or inversely how likely it is that the two variables are indeed independent.

Text (English)	Proposed by %	Feasible?	Comment
Blocking of a web site in my own country	51.9	Yes	DNS entries of the site are replaced with a link to a page informing the user that the page is blocked
Blocking of a web site in another country	33.2	No	Only possible, when all DNS (see [2] and the standards bundle cited therein) and VPN [3] traffic outside the country is monitored/blocked; an example may be the Great Chinese Firewall.[278]
Identifying and blocking IP addresses of offensive posts in my own country	66.3	No	Pointless, most IP addresses are assigned by the provider dynamically.
Identifying and blocking IP addresses of offensive posts in another country	48.1	No	See above
Identifying posters of offensive content in my own country	64.7	Depends	If there is an obligation to use clear names at least known to the platform provider and the provider has to disclose them, yes. Otherwise, no.
Identifying posters of offensive content in another country	46.5	Depends	See above, but even more unlikely.
Blocking email addresses	43.9	No	Using a fake sender email, such as biden@whitehouse.gov is simple, for an example see https://emkei.cz/
Upload filters to social media platforms	55.6	Yes	Actually implemented, but with severe issues. Difficult for AI to recognize irony, figurative speech, memes etc. AI is here still in its infancy.[279]
Obligation to use clear name in social media	67.4	Yes	This is a highly effective way of tracing posters of offensive content, however it requires a legal basis to oblige operators of social media and discussion platforms to enforce clear names for users (at least known to the platform operator, not necessarily shown in the posts).[280] However, this measure is not undisputed.[281,282]

Table 9. Countermeasures proposed by the respondents (technical and legal) (n=187)

---

[278] Washington Post, China's scary lesson to the world: Censoring the Internet works https://www.washingtonpost.com/world/asia_pacific/chinas-scary-lesson-to-the-world-censoring-the-internet-works/2016/05/23/413afe78-fff3-11e5-8bb1-f124a43f84dc_story.html

[279] The author uses the following classroom example in the sentiment analysis library *sentimentr* fort he R studio development workbench:

The *sentimentr* package assigns a sentiment value to every string of English-language words between -1 (totally negative) and +1 (totally positive) with 0 being neutral value. The following values apply:

This is bad (-0.43)

This is pretty (+0,43)

This is pretty bad (0,00)

The figurative speech is lost on AI, "pretty" and "bad" cancel out each other. Basing upload filters on such a technology is highly problematic. Viennese museums show their works of art by Egon Schiele and others on OnlyFans as it is the only platform allowing the upload of "adult content", https://www.wien.info/de/sightseeing/museen-ausstellungen/of-411214

[280] A less obvious but still highly effective variation is to require a mobile phone verification where no anonymous pre-paid phones are possible.

https://www.parlament.gv.at/PAKT/VHG/XXVI/ME/ME_00134/index.shtml#tab-Stellungnahmen

[282] https://netzpolitik.org/2019/digitales-vermummungsverbot-oesterreich-will-klarnamen-und-wohnsitz-von-forennutzern/

The results indicate that many legislators/policymakers are not aware of the technological feasibilities and restrictions under which the Internet operates. One-third and one half, respectively, for example, believe it is possible to block a website or an IP address in another country. Four out of ten respondents believe it is possible to block email addresses. This may be possible locally in one's mailer – and even then the success is doubtful if the perpetrator uses different email addresses, which is simple using a fake mail site – on a general level it is simply not feasible.

More than half of the respondents believe that upload filters are a useful tool for stopping hate speech and fake news, which to some extent is, of course, possible, but with the side-effects shown in some examples in the above table.

Only in two instances, there are clear matches between the inclination of the respondents and the technical feasibility: (i) blocking websites in one's jurisdiction (51.9%) and (ii) the obligation to use clear names (67.4%).

It would suggest itself that similar answers may be obtained from members of other legislative bodies all over Europe. One may see a clear need for educational resources here for legislators regarding the internet and its functioning. The internet is not only a key economic factor and a critical infrastructure, it has become a cultural technique, where legislation should be based on informed decision making on the technology at hand.

6.2.2. Political Remedies

Question 7.2 enquired about political remedies against fake news and hate speech, here are the answers:

Text (English)	Proposed by %
Open data, transparency of the grounds of political decision making	80.7
Citizen participation in decision making	62.6
Better explanation of decisions to the citizenry	75.9
Increased own social media activity	42.8
Increased off-line contact with citizenry	59.9

Table 10. Countermeasures proposed by the respondents (political) (n=187)

The two top scorers here both refer to openness in decision making and transparent communication why certain decisions were made. Citizen participation, sometimes seen as a panacea in overcoming the tendency to people distancing themselves from politics comes in only a third. Increased own social media activity (which can maybe be dubbed as "counter-strike" strategy) is decidedly in the minority.

Also increased offline contact with citizens is a popular answer, here – and also generally – it remains to be seen, whether this depends on the size of the political entity. Hence, X^2 tests were run between these answers and the size category of the entity as shown below. First the descriptive analysis of the size question:

Nr	Text (English)	Number	Percent
1	A local authority of less than 50,000 people	60	39.2
2	A local authority of 50,000 to 500,000 people	41	26.8
3	A local authority of more than 500,000 people	10	6.5
4	A regional authority of less than 100,000 people	3	2.0
5	A regional authority of 100,000 to 1,000,000 people	21	13.7
6	A regional authority of more than 1,000,000 people	18	11.8

Table 11. Entity represented (n=153), percentage from valid answers

Joining categories 1 and 4 (smallest entities, encoded as 1), 2 and 5 (medium, encoded as 2) and 3 and 6 (large, encoded as 3) into transformed variable T9.2 yields interesting X^2 test results: none, literally none, of the answers to Questions 7.2 depends on the size category of the political entity on a significance level of 90%.

One may have surmised that, for instance, increased off-line contact to citizens may decrease the issue but no significant connection between entity size and the answers in Question 7.2 was observed (shown as an example below).

Recommendation as to these measures hence do not depend on entity size.

T9.2 * Q7.2.5 Cross tabulation

			Q7.2.5		
			0	1	Gesamt
T9.2	1	Anzahl	18	45	63
		% von T9.2	28,6%	71,4%	100,0%
	2	Anzahl	27	35	62
		% von T9.2	43,5%	56,5%	100,0%
	3	Anzahl	11	17	28
		% von T9.2	39,3%	60,7%	100,0%
Gesamt		Anzahl	56	97	153
		% von T9.2	36,6%	63,4%	100,0%

Chi-Square-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)
Pearson-Chi-Quadrat	3,127[a]	2	,209
Likelihood-Quotient	3,163	2	,206
Zusammenhang linear-mit-linear	1,742	1	,187
Anzahl der gültigen Fälle	153		

Table 12. Proposed remedies by entity size

6.2.3. Support Infrastructure

Question 7.3 inquired about the support infrastructure desired by the Members of Congress concerning hate speech: "What kind of resources or support would help you cope with hate speech".

Text (English)	Percent
Training and education of myself and my institution on this topic	71.1
Counseling and supervision by psychologists, coaches etc.	34.8
Specialized staff within the police force which I can directly approach	60.4
Taskforce within my political party to effectively deal with online hate speech at my request	42.2
Taskforce within my institution to effectively deal with online hate speech at my request	52.9

Table 13. Support infrastructure against hate speech attacks (n=187)

There is a clear and clearly articulated demand for training on how to cope with hate speech attacks. Respondents do not see themselves as an issue when it comes to hate speech as indicated by the relatively low demand for psychological counseling – hate speech is clearly (and rightfully) not seen as the psychological problem of the person attacked.

A strong(er) involvement of law enforcement is indicated by respondents as well as, to a lesser extent, a specialized task force within their political group and/or institution.

6.2.4. Motives

Question 8 inquired about possible, perceived motives for fake news and hate speech. Here are the results:

Text (English)	Percent "yes"
Hate speech and fake news are more likely when people lack trust in the government.	80.3
If the government keeps its action secret and hidden, fake news and hate speech occur more likely.	88.9
More Open Government could reduce both hate speech and fake news by increasing transparency and accountability.	89.5
Organized creation of hate speech and fake news cannot be countered by government actions.	29.0

Table 14. Motives (n=157, 153, 153 and 138), valid percentage only

The answers show a clear pattern:

- Hate speech and fake news susceptibility are seen as a failure in government performance alienating people. The first two questions point in that direction. This is a remarkably honest approach by the respondents to assign these issues fundamentally to something being wrong with politics or their communications to the citizenry.

- Openness and transparency are seen as effective countermeasure corroborating the results for Question 7.2.

- And finally, with all the issues being discussed, there is a clear message that something can be done against it.

6.3. Summary

Even considering the small sample size, some interesting results can be drawn from the survey:

- Fake news and hate speech are not distant, theoretical issues, but they are a real part of a representative's political and also private life. They have the tendency to go together and hate speech has a tendency to spill over from the cyber to the real-world domain.

- Both phenomena are seen as a failure of the political system and an indication of a lack of trust in government. Transparent decision making and open government measures are seen as a key element to counter these phenomena. These findings are independent of entity size.

- In many instances, the respondents' perception of technical countermeasures is technically not feasible. Here, sometimes an unrealistic expectation towards technology can be seen.

- Respondents see better education of themselves and to a somewhat lesser extent of law enforcement as a viable remedy to counter fake news and hate speech.

One may hence draw the conclusion that a specialised training and education package specifically designed for political representatives would have a clear value added. It would on the one hand help them to counter aggression via social media and on the other hand to make better informed decisions on the digital media in their political capacity.

References Chapter 6

[6-1]   Colomina, C., Sanchez Margalef, H., Youngs, R., The impact of disinformation
        on democratic processes and human rights in the world, Study requested by the European
        Parliament, 2021, download at https://www.europarl.europa.eu/RegData/etudes/STUD
        /2021/653635/EXPO_STU(2021)653635_EN.pdf

[6-2]   IETF, Network Working Group, RFC 1034: Domain Names – Concepts and Facilities,
        download at https://datatracker.ietf.org/doc/html/rfc1034

[6-3]   Ezra P.J., Misra S., Agrawal A., Oluranti J., Maskeliunas R., Damasevicius R., Secured
        Communication Using Virtual Private Network (VPN). In: Khanna K., Estrela V.V.,
        Rodrigues J.J.P.C. (eds) Cyber Security and Digital Forensics. Lecture Notes on Data
        Engineering and Communications Technologies, vol. 73. Springer, Singapore, 2022.
        https://doi.org/10.1007/978-981-16-3961-6_27

List of figures

List of Tables

Fake news and hate speech have become a significant factor in political discussions aimed at suppressing other opinions and/or unduly influencing democratic decision making. Both phenomena corrupt and distort civic dialogue. This study analyses them from a policy, social and technical angle and develops a foundation for policies fighting fake news and hate speech.

The study was conducted with the support of the Congress of Local and Regional Authorities of the Council of Europe. Part of the research was also an empirical survey among the Members of Congress.