



Fundamentals

# DIGITAL DEMOCRACY – DO SOCIAL MEDIA STEER OPINION FORMATION?

👤 PD Dr. Merja Mahrt

Weizenbaum Institute

Version: 📅 6/25/2024

Cite as Mahrt, M. (2024). Digital democracy: Do social media steer opinion formation? (fundamentals series). Berlin: Weizenbaum Institute.

**Are social media bad for opinion formation?**

**Why social media (have to) select opinions**

Facebook's newsfeed algorithms: Blowing the whistle on their effects

The YouTube debate: Radicalization through video recommendations?

**How social media curate**

What happens below the surface of social media? What are the covert effects of advertising? And how to bots affect communication?

**The effects of curation**

Who uses what for political information?

Experiments on opinion formation

**Two potentially dangerous outcomes of curation that should be investigated in the future**

**How should social media be designed to protect opinion formation as much as possible?**

Do social media deepen societal rifts? ⇔

Do social media only show users what they are already thinking? ⇔

Do social media users only see "more of the same?" ⇔

What role do social media play in the spread of false information? ⇔

Are users becoming more radical through YouTube's recommendations? ⇔

Are users becoming more radical through YouTube's recommendations? ⇔

What makes advertising in social media special? ⇔

Can opinion formation be manipulated through social media advertising? ⇔

What influence do bots have on communication in social media? ⇔

How does Russian propaganda affect opinion formation in other countries? ⇔

What role do social media play in the spread of false information? ⇔

Are users becoming more radical through YouTube's recommendations? ⇔

**References**



# ARE SOCIAL MEDIA BAD FOR OPINION FORMATION?

With the spread of social media in recent years there has been growing concern about their effects on opinion formation. Key suspects are fake news, radical content, and algorithms that mostly show users what they are already thinking, among others. Accordingly, researchers have worked on numerous

studies that investigate the effects of social media on democracy and its members.

This compact overview focuses on opinion formation on the level of individual users. Opinion formation is the fundamental basis of democracy as a person's opinion on political, social, or economic issues directly feeds into their voting behavior. The results of elections and referendums in turn determine which people or parties will make decisions during the next legislative period – decisions that are generally binding. If social media affect this process, it should be clear whether users, for example, no longer perceive different opinions or become ever more radical in their own worldviews.

What people see in their everyday social media use has potentially far-reaching consequences. This summary of current research offers an overview on the most relevant aspects of this topic: why social media need to select content and opinions in the first place and how this selection works; what consequences this selection has on opinion formation among the users; and what options to act exist for users, platform providers, and politics. Concrete fears that abound in public debate – for instance, on fake news, radicalization, or filter bubbles – are addressed in short profiles that can be read independently.

Profile: Digital divide

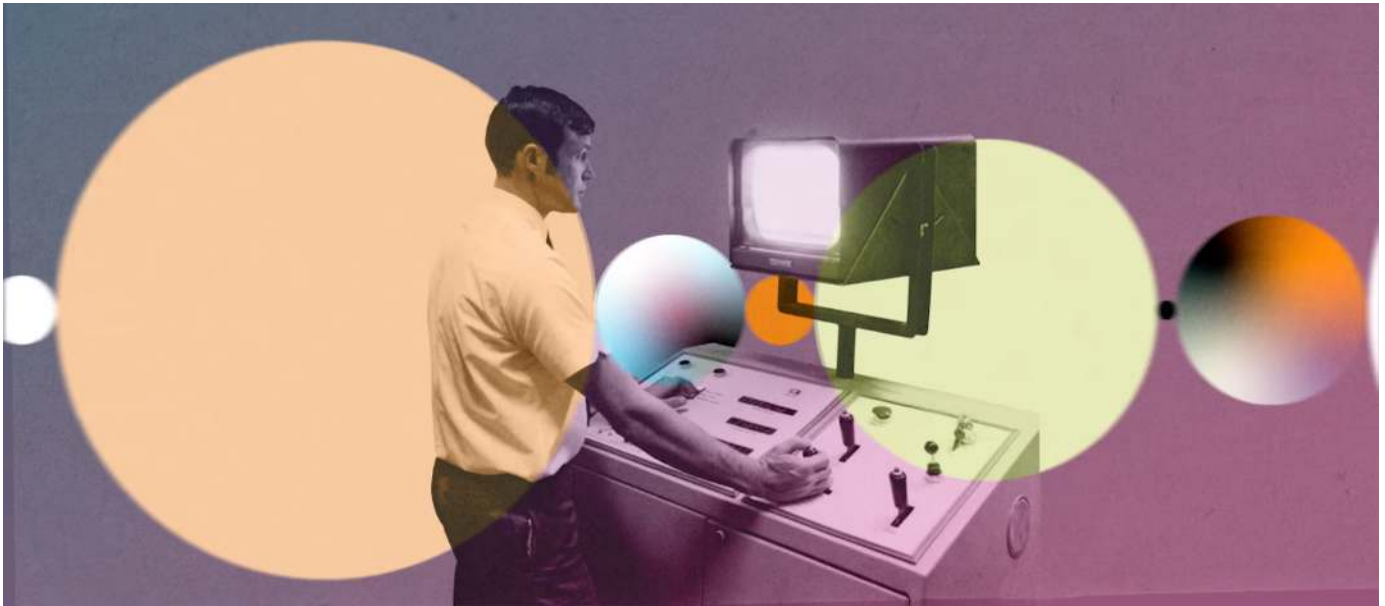
Do social media deepen societal rifts?

Profile: Echo chambers

Do social media only show users what they are already thinking?

Profile: Filter bubbles

Do social media users only see “more of the same?”



## WHY SOCIAL MEDIA (HAVE TO) SELECT OPINIONS

What content social media display to users is the result of several levels of selection that are necessary for platforms to work and that serve different purposes for them:

### 🔗 Variety of content

Social media make it very easy for users to create their own content or distribute content from third parties. Platforms themselves, on the other hand, only create small amounts of the available content. This means that the amount of content flowing through a platform at any given time is practically unlimited and needs to be made accessible in one way or another, for example, through clustering, filtering, or sorting.

### ⚖️ Legal framework

Social media need to respect legal boundaries for permissible content. Respective limits concern, among other things, individual rights, the protection of minors, copyright, and other boundaries of freedom of speech. Since these legislations differ between countries, globally active platforms get into conflict about the limits of permissible content in different national contexts.

### 🔄 Business models

Social media businesses depend, for the most part, on advertisers, and partly also on paying subscribers. Subsequently, social media companies need to offer their different stakeholders an accommodating platform. With advertising-based business, this platform needs to work for two groups simultaneously to create sufficient opportunities of contact between

advertisers and users. Platforms, therefore, often strive to keep users on their services for as long as possible or to make them engage with the available posts as much as possible.

### Own policies

Lastly, social media also follow an ideal of what their respective platforms should or should not be used for. Their terms of service thus typically contain rules for platform-appropriate content that go beyond legal requirements. In recent years, loose versus strict limits for nudity or the expression of political opinions have been a frequent topic in discussions about how social media select content.

Selecting content created by third parties for their users – so-called curation – is a central function of social media. Faced with large numbers of users, huge amounts of constantly posted content and the diverse international backgrounds of users and respective legislations, this selection is a complex task with potentially multiple consequences.

How this can result in conflicts of interest can be illustrated by the following two examples:

## Facebook's newsfeed algorithms: Blowing the whistle on their effects

In the fall of 2021, the former Facebook employee Frances Haugen [leaked internal company documents to the Wall Street Journal](#) and [testified in a US Senate](#) hearing on Facebook's business practices and goals. According to her statements and to the documents, the News Feed algorithms had been optimized to incite as much "user engagement" as possible, meaning interactions with posts through liking, sharing, or commenting. This favors posts that elicit strong emotions among users, including, for example, fake news stories and conspiracy theories designed to inflame the public and incite anger. The leaked material and Frances Haugen's testimony imply that the company had been aware of the potential negative effects of this type of content for years, but chose to prioritize other goals in the continuous reworking of their algorithms.

## The YouTube debate: Radicalization through video recommendations?

In several widely read articles in the *New York Times*, YouTube was criticized for allegedly radicalizing users via increasingly extreme video recommendations. [Then product chief and current CEO, Neal Mohan](#), said in an interview it would not make sense to assume YouTube would purposefully radicalize users through video recommendations to make them stay on the platform as long as possible. He pointed out that important advertisers would not want their ads shown alongside radical content, and usage time alone would not benefit YouTube financially.

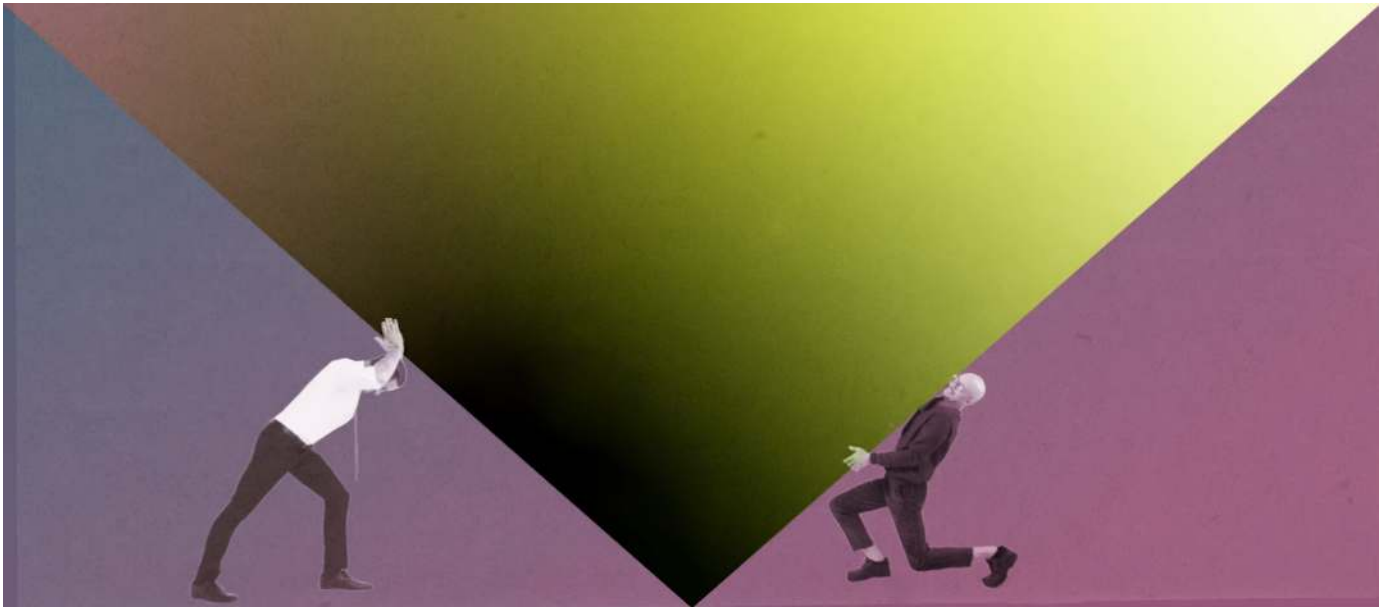
Both examples show that platform providers take into account different goals at the same time—and that it is difficult to satisfy them equally through the highly complex algorithms of social media.

Profile: Fake news

What role do social media play in the spread of false information?

Profile: Rabbit holes

Are users becoming more radical through YouTube's recommendations?



## HOW SOCIAL MEDIA CURATE

Compared to the curation of content at newspapers, radio, or television stations, social media differ in two major respects: First, they know in much deeper detail who their users are and what they select. Second, their identity is typically not centered on journalistic ideals of societally relevant reporting, background information, or editorial comment on current events. Quite to the contrary, Facebook has in fact stated for years that it is *not* a media company.

The detailed information that users themselves enter into their social media profiles and that manifest in their usage behavior over time allow platforms to curate content on the level of *the individual user*. Users directly steer the selection of posts they want to see by connecting with other accounts, subscribing to their posts, or searching for specific offerings. Additionally, platforms usually assume that people want to see more of the type of content they look at frequently or extensively, according to the readily available and detailed usage data. Subsequent curation of content for the users can mostly be subsumed under one of two types:

## 🕒 Recommendation systems

Recommendation systems work based on similarity between types of content and users: If many users select content type Y after having seen X, the former type will be recommended to future users who access X. It remains up to the users whether they actually select content type Y. Via users' IP addresses, recommendation systems can take into account the approximate location, time of day, or season when suggesting more content. This type of curation remains comparatively coarse and shares similarities with how traditional mass media suggest content; for instance, how broadcasters try to create "audience flow" through the scheduling of adjacent TV shows or how journalistic content and advertisements are tailored to the targeted audience of a magazine or program.

## 🔍 Personalization algorithms

Personalization algorithms, on the other hand, "learn" over time in much more detail what is used on a given device (through cookies or tracking pixels) or by a specific person (logged into their account) and can suggest seemingly matching content for further use. On social networking sites, lots of information entered by the users themselves or by their contacts on the platform can be included in this type of curation, which contains much more detailed information than recommendation systems. For example, previous interactions with another user or their posts can be included in the decisions regarding if, how often, at what rank, or with what time delay content from this user will be included in a person's personalized social media feed.

Users can influence this curation to a certain degree, for instance, by heavily interacting with another user on the platform. For the most part, however, the rules that determine what one's own screen shows are opaque and cannot be tailored by the user themselves (while with traditional media like broadcasting or print, users can be certain that the selection and order of content is standardized for the entire audience). Even the list of topics that are displayed as popular on a digital platform or at least within a given country (often as "trending") may be personalized, unbeknownst to the users.



TED talk: How do filter bubbles work?

YouTube

For the necessary selection of posted content, social media use targets such as frequent, regular, or extensive usage, watching videos until the end, or engagement (liking, sharing, commenting...). It is, however, tricky to strike an attractive balance of known versus new or surprising content, to not keep recommending the same content to users over and over again. When Netflix was still mailing DVDs instead of streaming, the company even opened a competition based on this problem: Whoever could create a better recommendation system than their own “Cinematch” would win a prize of one million US dollars, which highlights the value of good content recommendation. Recommendation and personalization algorithms in social media face a similar dilemma to Netflix: If they only display content that fits a user’s previous selections the feed could become boring after a while and result in shorter or less frequent use of the platform. Conversely, YouTube has been criticized for including more radical content in its recommendations to users who watched political videos, including extremist content and conspiracy theories [1]. Equally problematic for the video platform was the fact that in 2017, ads from large advertisers had been shown next to videos with extreme political content and hate speech. The cancellation of billions of dollars’ worth of advertising deals put YouTube under a lot of pressure to adapt its curation not only to the users’ viewing behavior, but also to the interests of their advertisers. The platform subsequently changed its rules for monetizing videos, which in turn had repercussions for content creators, whose videos form the basis of bringing together viewers and ads.

In summary, there is a strong interrelation between what social media users want to see (in the sense of: what they click on, and how often); what appears useful for platform providers and their own, usually commercial, goals; what advertisers perceive as a friendly environment for their purposes, and what content gets produced by creators. Long-term trends towards more and more radical content recommendations or like-minded people sharing one-sided political content on social networking sites are seen as particular threats to opinion formation.

### **Why has online content become more and more negative?**

A prime example of how user behavior, platform curation, and content creation feed into one another is the phrasing of online news headlines. Two recent studies have shown that curation on social media and in journalism tends to make news headlines more negative in online settings: they spread more quickly on social media, meaning that they receive more clicks [56]. This is an incentive for journalists and editors to use negative expressions more frequently in headlines to increase their usage numbers. Such a trend towards more and more negative headlines has been documented for 47 US news outlets since 2013 [57]. These findings highlight how social-media curation through purposeful or incidental usage behavior as well as through the programming of algorithms is embedded in further contexts and can have lasting effects on what gets posted and receives attention.

What happens below the surface of social media? What are the covert effects of advertising? And how do bots affect communication?

Profile: Rabbit holes

Are users becoming more radical through YouTube’s recommendations?



Profile: Microtargeting

What makes advertising in social media special?

Profile: Dark ads

Can opinion formation be manipulated through social media advertising?

Profile: Bots

What influence do bots have on communication in social media?



## THE EFFECTS OF CURATION

Scientific investigations into the effects of social media curation typically follow one of two paths: A first option is to ask users. This is considered the most pertinent way to capture people's opinions and allows for surveying a wide range of potentially confounding factors (which, in an experimental setting, can also be controlled by the researchers). On the flipside, survey-based study designs can only relatively broadly gauge what types of content users came into contact with on social media, as their memory, available time, and patience pose natural limits. The second option consists of capturing social media usage automatically and in great detail. However, this often limits what information about political opinions or other individual differences of the users can be gathered.

## Who uses what for political information?

Annual surveys such as the Reuters Digital News Report show that social media are a staple of many German users' media diet [2]. They are especially popular among young people: during the 2021 Federal Elections in Germany, almost half of first-time voters said that social media were their primary source of political information [3]. Since social media are typically combined with many other media outlets, it is difficult to assess the specific effect of social media use on opinion formation or other outcomes.

Milieu-based research suggests that social media use is most likely to influence the opinions of two groups of users: people who almost exclusively come into contact with information about current events via social media; and people whose contacts in social media are very homogeneous with regard to political opinions [4]. All other milieus use many different sources of information so that the possible effects of social media use are mitigated by these other outlets.

A large-scale analysis of digital usage data additionally shows that social media serve as distributors of content that bring average users in contact with more news than non-users [5]. Due to the study design, it cannot analyze the opinions contained in the news items; however, US-based studies of large usage datasets have found a way to approximate this: The American two-party system allows researchers to estimate both the political leaning of news posts via the proportion of users who lean Republican, Democrat, or Independent, as well as to establish who accessed the news based on their partisanship. Such studies show that the vast majority of users selects news items with a neutral leaning that are used across party lines [6]. Only small groups at the ends of the political spectrum tend to use mainly news that are exclusive to their partisan camp.

These results are a good reminder of the fact that many political opinions exist prior to media use, including social media use, and that can influence what content is selected. It is plausible to assume that social media use can reinforce previously held political opinions when like-minded content is consumed or that it can balance opinions when the user is exposed to a variety of viewpoints. In fact, a German survey of Facebook users did not find an effect on users' opinions [7]. Unlike Facebook usage, the same study showed that both gender and reading newspapers are, however, predictors of political opinion: Women and newspaper-readers are clearly more moderate in their opinions than men or non-readers. This result again underlines the importance of usage of other types of media on opinion formation beyond social media.

## Experiments on opinion formation

Instead of conducting surveys on the effects of social media curation on users in everyday life, which can be affected by numerous confounding factors, other researchers design experiments to control the latter. They typically show participants' opinion posts from social media and ask about potential effects of the users' political opinions after exposure. This allows for specifying the effect of different types of posts in a systematic manner. A review of seven such experimental studies revealed that preexisting differences in opinion can be deepened by exposure to social media content [8]. In an Austrian study, for instance, left-leaning users perceived a populist right-wing politician more negatively after seeing two of his anti-migration tweets [9].

Such an increase in difference of opinion through social media use is called polarization. Experiments like this one, however, cannot glean whether an opinion actually *changed* through exposure or if the users simply become more

For about half of first-time voters, social media are the main source of political information

2021 Federal Elections in Germany

aware of their already established opinion of the politician in question. It is equally unclear for how long after the end of the experiment the effect persists. Lastly, given the many instances in which social media users select among options in their everyday usage, a situation of chicken or egg arises: If users typically connect to like-minded people on social media and like posts that align with their opinions, thereby making the curating algorithms show more similar content in the future, to what degree has an opinion already been polarized before exposure and how big is the additional effect of social media usage?

In summary, scholars have investigated various types of effects of platform curation on news use in social media as well as on opinion formation. Experimental research does confirm polarization effects; however, typical usage patterns of social media platforms bring people in contact with a variety of posts that tend to be moderate with regard to opinions. In addition, social media are only one source of news for many age groups.



## TWO POTENTIALLY DANGEROUS OUTCOMES OF CURATION THAT SHOULD BE INVESTIGATED IN THE FUTURE

The state of research on the effects of social media curation is complex, and many results are far from alarming. However, two aspects deserve continued attention, both in future research and in public debate:

First, a small effect on opinions can have broad ramifications. A prominent example is the close result of the Brexit referendum, which was preceded by a large amount of fake news (also in social media) as well as attempts from different online actors to influence the vote. In a similar vein, the last two US presidential elections were won by a small margin, which illustrates that even a small change in (or confirmation of) opinions can potentially impact a larger decision.

Second, social media can play a relatively big role for users with a very narrow or already radical news repertoire compared to the majority of users. People who spend a lot of their time online in comparatively closed groups and are exposed to a homogeneous news diet may become more radical in their opinions over time. It is in this respect that YouTube, for example, has been criticized for its supposedly rabbit-hole-like recommendation algorithm. Users with low political knowledge who only use news sporadically might also be at risk of greater effects on their opinions through social media.

Profile: Russian propaganda

How does Russian propaganda affect opinion formation in other countries?

Profile: Fake news

What role do social media play in the spread of false information?

Profile: Rabbit holes

Are users becoming more radical through YouTube's recommendations?



## HOW SHOULD SOCIAL MEDIA BE DESIGNED TO PROTECT OPINION FORMATION AS MUCH AS POSSIBLE?

Social media do not only display what users are already thinking, and their effects on opinion formation are not as direct as people in everyday life or public debate might assume. Curation through usage behavior and platform algorithms should be considered separately for their potential effects on different types of content, and existing studies give little cause for alarm. However, problematic consequences may occur for some groups of users, for instance, people who rarely use other news sources or users in very homogeneous networks of contacts. Consequently, the following suggestions have been made to improve social media with regard to opinion formation:

## Platform providers

Companies operating social media platforms should acknowledge their importance for opinion formation. Today's dominant platforms were not created for news and current events, while journalism and its professional norms have evolved over centuries. In contrast, social media leadership typically try to eschew responsibility for their products' effects on opinion formation and democracy. Companies have successfully argued that they are not media companies and have avoided attendant regulation. The large US companies additionally follow a libertarian interpretation of freedom of speech that is not shared in all cultures where their platforms are being used around the globe.

## Politics

Since the late 2010s, media regulation in, among others, Germany, the European Union and the US state of California has introduced changes that try to better take into account the central role of large social media platforms in the distribution of information and opinion formation. These initiatives are intended to protect freedom of information and opinion, but they have been criticized as inadequate so far [10][11]. The long-term repercussions of laws and regulations will have to be monitored in the future.

## Usage

Lastly, users should be aware of the possibilities for access to information and opinion formation that social media provide and try to benefit from opportunities. A starting point is one's daily media repertoire: The positive effects of newspaper reading suggest that accessing reliable journalistic sources is beneficial. Via license fees for public-service broadcasting, all people in Germany have access to a wide range of information sources. This is complemented by a diverse newspaper offering that can be supported through subscriptions or individual purchases (for digital or printed editions). This type of usage should not only benefit a person's political knowledge and opinion formation, it also contributes to a healthy media system in which several newsrooms provide in-depth reporting and investigation—and also monitor their respective competitors for potential mistakes or manipulations.

Opinion formation as a central democratic process is thus worth protecting through a variety of measures, which nevertheless should take into account the positive effects that social media can have.

Profile: Bots


## What influence do bots have on communication in social media?

### What are they?

Bots (short for “robots”) are software whose automated communication imitates human users or carries out their tasks. With regard to opinion formation and public communication, specifically social bots are the most relevant, that are designed to appear to be regular social media accounts. They



The debate about whether or not Facebook is a media company has been going on for years

 Wired



publish posts, share others, or like and comment on them. Not all social bots are fully automated, but can also be run by people or combine human and automated posts. Subsequently, definitions of bots can be broad or narrow. This text will focus on inauthentic, seemingly human digital communication that does not have to be fully automated. Bot-like inauthentic patterns of communication include unrealistic networks of contacts that look suspicious themselves, very regular posting behavior, and spam that is posted with current trending hashtags even though there is no apparent link between the content of the post and the hashtag.

## Since when do they exist and how widespread are they?

The first bots were created in the 1960s, but this dating depends on which definition of bots we use. Joseph Weizenbaum's chatbot ELIZA, which in 1966 was able to have fully automated conversations with users, is frequently named as an early milestone of bot development ([a recreation of ELIZA can be experienced here](#)). But even before that and up until today, most bots are active behind the scenes, performing routine tasks of computer-based communication and thus stabilizing infrastructures and relieving human administrators.

Apart from chatbots, early social bots were also used in computer games and bulletin board services. Already in these instances, bot-based communication was seen by users as sometimes benign, helpful, or funny, while bots created with malicious intent caused unwelcome disruptions, for example through spam, vandalism, or provocative communication. Today, a large part of research as well as public discussion surrounding bots centers on inauthentic accounts created with improper intentions on social media. Especially X (formerly Twitter) receives a lot of attention in this regard.

## Which effects have been confirmed?

A large part of bot research is concerned simply with identifying bots: this is the necessary foundation for acting against undesirable bots, ideally in real time. However, bots have to be reliably detected to assess their effects. Many methods to detect bots in large datasets exist, each with their own costs, strengths, and weaknesses [52]. Using different methods of detection on the same dataset can lead to very different results. In one study, the proportion of Twitter accounts that were labeled as bots varied between 3 and 25%, while only 0.25% of accounts were flagged as bots by all three methods employed [53].

Accordingly, studies on the effects of bots, for instance, on the 2016 US presidential elections, come to very different conclusions. One analysis found highly active bots that spread conservative posts, thus making them more visible and potentially influential [44] (see also the text on [Russian propaganda](#) ↻ ). Another study concluded that such bots only rarely reach new users and that their influence on opinion formation is thus limited [54]— but it is important to note that confirming people in their already held beliefs is also a media effect. Numerous scholars predict that in the future, artificial intelligence that uses large language models (such as ChatGPT) will make it even harder to identify bots, exclude them, or assess their impact.

## What can be done about them?

Identifying bots is not only important for researchers who want to describe the phenomenon and measure its effects. The respective methods are also used to automatically detect inauthentic accounts in order to delete them. This requires the balancing of competing goals [55]:

- Since not all bots are malicious and some are very useful, it is not viable to completely ban any form of automated account. Very strict limitations, for example, through screening all posts for suspicious patterns before they are published, could also impede free speech. This is due to the mistakes that are bound to happen in filtering out posts or accounts that are in line with regulations. Authoritarian regimes could also use legal frameworks to prohibit the anonymous posting of undesired content.
- With the *Medienstaatsvertrag*, the current media regulation enacted in 2020 in Germany, content providers with a large reach need to flag their automated accounts in social media as such. This is intended to increase transparency for users. However, people or organizations running illegal or malicious bots will hardly follow this stipulation.
- In addition to these permanent rules for bots, there are also suggestions to limit certain types of usage, for instance, the number of posts permissible per day or hour. To protect opinion formation in the weeks leading up to an election or other vote, this could be limited to only the critical phase directly before the ballot.

Profile: Dark ads

## Can opinion formation be manipulated through social media advertising?

### What are they?

The term “dark ads” refers to advertisement, typically on digital platforms, that are only shown to a small number of people. At first, this might seem to contradict the usual purpose of advertising – to attract attention. However, ads in traditional mass media often broadcast their message to people who will not be interested in the respective product anyway. With online media—but even more so digital platforms such as search engines, social networking sites, or video platforms – it is technically possible to target certain user groups much more specifically (and only pay for these instances). This makes it difficult for third parties, such as journalists, policymakers, researchers, or regulators, to gain an overview of what advertisements are actually spread, since “dark ads” remain hidden, as their name suggests.

### Since when do they exist and how widespread are they?

The technical groundwork for showing different ads to different users of digital platforms was laid in the 1990s and methods have been refined a lot since then. Today, Google and Facebook are market leaders in online advertising and can target very small groups with different versions of ads ⇔ . The exact number of personalized ads, however, is unknown. In this regard, Trott et al. [26] criticize the database of political advertising that Facebook offers for a number of countries as purposefully obscure and not helpful for external oversight.

### Which effects have been confirmed?

Compared to filter bubble effects ⇔ of news, there is a dearth of studies on the patterns of available digital ads and their consequences on political attitudes or actual voting behavior. Experiments on microtargeting ⇔ have resulted in ambiguous findings that can usually only assess short-term effects

that occur directly after being exposed to ads. With regard to the Cambridge Analytica scandal surrounding the Brexit referendum and US presidential elections of 2016, very different conclusions were drawn on the effectiveness of the “dark ads” that were part of the respective campaigns. Former campaigners laud their own successes [27], while others claim that the effects were much smaller in reality. Since both votes were very close races, however, even small effects on the voting behavior (or abstention) of the respective different camps can potentially have contributed to the final result.

## What can be done about them?

As explained for [microtargeting](#) ⇔ , legislation in the European Union limits how users’ data can be used for profiling and microtargeting. Trott et al. [26] additionally suggest better education of consumers about “dark ads.” They have developed a tool that provides users an overview of the degree of customization in the ads they see online. Both [Facebook with its ad library](#) and [Google with its reports on ad transparency](#) allow searching for different versions of political advertisements. In addition, some political actors have themselves started to provide an overview of which ads they use in a campaign for different target groups (e.g., [the Green party in Germany](#)).

Profile: Digital divide

## Do social media deepen societal rifts?

### What is it?

Digitalization affects numerous parts of everyday life, including infrastructure, public administration, commerce, education, work life, leisure activities, media use, and interpersonal communication. However, the degree to which these aspects of life are actually changing varies as digitalization does not spread evenly. The existing systemic differences between and within countries, as well as between groups of people, are called digital divides. In addition to possible differences between users and non-users, researchers also investigate how different types of users handle digital technology (*digital use divide*) and how these differences in usage affect them.

### Since when does it exist and how widespread is it?

Since its early days, the Internet has been lauded for providing free and equal access for all humans around the globe, at least in theory. However, the small number of initial users were distributed very differently globally as well as socially. It was first and foremost affluent Western countries and within them universities and research institutions mainly in technical and scientific fields that began adopting the Internet in the 1970s and 80s. When usage became much more widespread in the 1990s, the systematic differences in adoption of the new technology became a topic of interest. Pippa Norris published an influential eponymous book on the *digital divide* [37]; a recent overview of the many facets of the topic can be found in van Dijk’s book of the same name [38].

Current usage data continue to document large differences between countries with regard to Internet adoption (see, e.g., [the yearly reports of the International Telecommunication Union, an agency of the United Nations](#)),

and even in countries with widespread Internet usage such as the US and European countries, certain groups are either scarcely using digital technologies or even not at all: A below-average income and low level of formal education still correlate with non-usage, which is most pronounced in many countries for older women with low formal education who live in rural areas.

## Which effects have been confirmed?

Digital divides in usage can result in **deepening pre-existing differences over time**. Both in terms of comparisons between different countries and individual users, those with more resources at their disposal are able to make use of digital offerings to their advantage, which will further increase their lead over others (for example, with regard to accessing services, knowledge, or means of communication). On the level of countries, relevant aspects include the available bandwidth or the quality of digital equipment in schools and universities; whereas for individuals, studies on different areas of life (e.g., integration in the workforce, political participation, or health information) show that increased use of the internet as well as higher digital literacy have positive effects for users. In short, this means that people who do not use the internet at all or are unsure how to best use it for their needs benefit less and lag (even further) behind the users.

How different uses of digital technology affect **political opinions** is summarized in the texts on [filter bubbles ⇄](#) , [echo chambers ⇄](#) , and [rabbit holes ⇄](#) .

## What can be done about it?

Among other things, the International Telecommunication Union (ITU) of the United Nations is explicitly charged with leveling the global digital divide with regard to access to the Internet and other technologies. The ITU supports actors who want to build infrastructures in countries with low levels of telephone or Internet access. It also works on regulating satellite and radio communication as well as standardization of communication technologies.

Large private companies such as Meta (formerly Facebook) and Starlink are currently building technologies that will provide Internet access to many people in the Global South (e.g., Meta's 2Africa optical cable network around the African continent and Elon Musk's satellite-based Starlink system). These initiatives are subject to ambiguous discussion both in science and in public discourse, as they involve single individuals or companies gaining a lot of power over basic infrastructures. This allows them, for instance, to transfer their own content faster or more reliably (which violates net neutrality) or to neglect the necessary regulation of their platforms in less affluent countries ([Amnesty International demands reparations from Facebook](#) for the Rohingya, a minority persecuted in Myanmar, about whom hate speech and incitements to violence were spread mostly without interference from moderators of the platform).

Within countries, media literacy programs are being initiated to help level the digital divide and support people with low digital literacy to start using the internet for the first time, or with greater personal efficiency. However, the same countries might introduce digital platforms for administrative transactions or social services that specifically target groups of people with low levels of Internet use or digital literacy. Accessibility needs to be taken seriously to ensure that already disadvantaged groups are not further marginalized.

Profile: Echo chambers

## Do social media only show users what they are already thinking?

### What are they?

Echo chambers only repeat what is already spoken—this is the basic assumption behind the metaphorical use of the term in the context of social media and other digital platforms that allegedly mainly mirror the pre-existing attitudes of their users. In Cass Sunstein's original formulation of the echo chamber hypothesis [28], this was explained first by the much larger offering of content and opinions in the digital realm compared to mass media. In addition, parts of this broad landscape of content are much more partisan than traditional journalistic outlets and also tend to provide hyperlinks to like-minded rather than opposing sites. Lastly, users can be much more selective in online environments and forward content more easily to others, in both cases potentially aligned with their opinions and worldviews. Taken together, these factors can result in people from right across the political spectrum receiving confirmation for their individual stances and becoming more radical over time because like-minded people might inoculate each other against opposing views to the point where the latter might not even exist. Since the original publication, the echo chamber hypothesis has been extended to also include filter bubble effects  $\Leftrightarrow$ , since algorithmic curation can further reinforce the process [29].

### Since when do they exist and how widespread are they?

“Birds of a feather flock together” is a saying in English and many other languages. And research shows that people tend to have particularly close relationships with people who are similar to themselves. This also applies to social media, where, for example, people communicate more often with other users who share similarities [30]. Likewise, the phenomenon of group polarization is well established in psychology [31]: in a discussion of like-minded people, the group consensus will shift over time in the direction of the pre-existing opinions of the group members, while discussions in mixed groups will tend towards the middle of the range of present opinions. The echo chamber hypothesis thus builds on typical social and psychological patterns. However, it is merely a *hypothesis* rather than the result of specific studies on the entire chain of effects itself.

Other limitations also have to be noted with regard to the significance of hyperlinks between like-minded websites: it is self-evident that such networks of partisan websites exist; however, it is much less clear how often and by whom they are actually used. Likewise, the echo chamber hypothesis ignores the fact that many well-established legacy news media (in Germany, first and foremost the TV news show “Tagesschau,” as well as *Der Spiegel*, local newspapers, or regional radio stations) continue to enjoy a wide reach (both online and/or offline).

### Which effects have been confirmed?

In everyday **usage of social media**, echo chambers are much less frequent than Sunstein [28] assumes. When such platforms display, for example, that a post

has received many likes, users tend to click on it regardless of whether or not it aligns with their political attitudes [32]. And users come into contact with a more diverse range of content than people who do not use social media [5][33]. Additionally, people who follow political accounts, for example on Twitter/X, tend to do so across the political spectrum [34]. The fact that people communicate more often with others who resemble them is thus only one part of the overall picture of communication in the digital world. In fact, a study by Goel et al. [30] shows that people are connected online to significantly more people than would be manageable without digital platforms. Thus, for most people, social media tend to broaden the horizon instead of enclosing them in echo chambers where their own attitudes are repeated to them over and over again.

As is the case with filter bubbles ⇔ , studies that document **effects of echo chambers on opinions** are far and few between. For one, media usage is not as partisan as originally assumed. Second, networks of contacts on social media at least in part reflect with whom people come together or communicate in other contexts, independent of digital platforms. It is therefore difficult to determine the specific effect of using one or more platforms on attitudes or behavior. Studies of conversations or their topics reveal, for example, that even in times of smartphones and social media, people continue to talk to others and that there are still many topics that are relevant to many different groups of people at the same time, all of which safeguards against echo chamber isolation [35][36].

## What can be done about them?

Most users do not currently have to take action to avoid getting trapped in an echo chamber. As a preventative measure, it appears sensible to strengthen societal institutions that bring together a variety of people, such as schools, sports clubs, or public events. With regard to media and communication, content diversity should be pursued. Public-service broadcaster ZDF, for instance, documents on its blog how the recommendation algorithm of its video repository is designed to suggest to users a broad overview of the available content, instead of always recommending the same types of content to them. Users may also want to review how their use of digital platforms and their activities in their spare time may bring them in contact with mainly similar people and content or, conversely, with a variety of perspectives. As with rabbit holes ⇔ , it is clearly possible to use social media to purposefully only interact with a very homogeneous group of people and only a small range of media content. Based on the current state of research on echo chambers, such usage patterns are, however, far from the norm.

Profile: Fake news

## What role do social media play in the spread of false information?

### What is it?

The term “fake news” is used for media content that looks like a typical news item, but that purposefully contains false information. Fake news is frequently distributed via social networking sites or messengers and often follows

formats typical of these platforms. The intent behind distributing false information, which is supposed to be shared as widely as possible, is usually either economic or political. Different factors can make news “fake”: from the [complete fabrication of events or alleged quotes from, for example, a prominent politician](#), to details, pictures, or video snippets that are not factually wrong per se, but are misconstrued through a different context, tactical omissions of information, or through misleading commentary.

Some studies also consider satirical news (such as Comedy Central’s “The Daily Show” or digital news parody *The Onion*) as a form of fake news. For many others, the distributors’ intent to mislead users with their content is a necessary criterion. Accordingly, errors published in journalistic media are usually not seen as fake news.

## Since when does it exist and how widespread is it?

The spread of false information in the style of typical news formats, usually with manipulative intent, has existed as long as mass media themselves. Its purpose is either to benefit financially from the received attention or to influence people’s opinion or behavior. The popularization of the Internet since the 1990s has opened many convenient avenues for creating and spreading fake news. In 2016, the Brexit referendum as well as the US presidential election brought attention to [social media’s role as platforms for the wide distribution of “fake news.”](#) During the Covid-19 pandemic, messaging apps, and groups within them, became especially prolific spreaders of purposefully false information.

## Which effects have been confirmed?

It is difficult to determine the **frequency** or **spread of fake news** in social media in a representative way, relative to the entirety of the available content on the respective platforms. Therefore, studies typically assess the proportion of false or misleading posts among a selection of popular content on the topic of a given event. Accordingly, results for the frequency of fake news differ widely. Appel and Doser [12] found, for instance, that the most widely shared social media post about former German chancellor Angela Merkel was in actual fact a fake news item. This type of content can thus be highly successful, but it is difficult to compare this success to other posts as the amount of digitally available content is de facto unlimited.

There is more research on the **use** of, **trust** in, and **effects of fake news**. Usage data on the 2016 US elections show that nine out of ten Facebook users did not share any content identified as fake news and that only a small minority did so repeatedly. Older and more conservative users were more likely to share fake news than younger and more liberal ones [13]. The likelihood of accepting fake news as factually correct increases if the respective content aligns with a user’s own opinions [14]. Therefore, fake news may confirm existing worldviews rather than convince people of new opinions [15].

In summary, research on different events, in different countries, and at different times arrive at very different results with regard to receiving, sharing, and perceiving fake news in different segments of the usership. Existing findings can therefore hardly be generalized.

## What can be done about it?

Parallel to the public debate about fake news in social media, **fact-checking** websites and sections in journalism have been created that evaluate the

truthfulness of popular online content. The Reporters' Lab at Duke University offers a [database of fact-checking services available worldwide](#). Large platforms such as Facebook or YouTube display warnings about potentially false information next to some types of posts. For content on Covid-19, for instance, many social media platforms link websites with reliable health information.

Altay et al. [15] additionally suggest strengthening **trust in reliable sources of information**. The extensive discussion of fake news has been shown to make users more skeptical of news in general, even though the amount of factually correct news is much larger than that of fake news.

Profile: Filter bubbles

## Do social media users only see “more of the same?”

### What are they?

Filter bubbles allegedly emerge through the algorithmic curation of social media (and other digital platforms, such as search engines and news websites). This includes recommendation systems that suggest content that fits posts users have already selected (e.g., a YouTube video started through AutoPlay at the end of another that has a similar topic). Personalization algorithms tailor the selection and order of posts in social media to the person logged in (e.g., in a feed of recent posts or on a “For You” page). Both types of systems are typically designed to show a user “more of the same,” instead of providing a broad overview of available content or current topics. Especially in the case of political content, critics fear that users will access only one-sided information that fits their preexisting knowledge and opinions, so that over time, people with different levels of knowledge and different political positions could further diverge from one another.

### Since when does it exist and how widespread is it?

Eli Pariser published the eponymous book on filter bubbles in 2011 and thus coined the term. However, recommendation systems for online content based on similarities with other offerings existed long before social media became popular. They became well-known through Amazon's product recommendations (“Customers who bought this item also bought...”), where they provide orientation among the vast array of available goods. The equally diverse platform YouTube similarly implemented video recommendations in its early days. In the 2000s and parallel to the development of social media platforms, dynamic feeds became technically possible and were implemented, for example, as early as 2006 on Facebook.

Recommendation systems and personalization algorithms are integrated into many social media and other digital platforms. From the vantage point of users, these algorithms are usually opaque and do not show to what degree the displayed content has been customized for a given user or how many other users get shown the same posts. Research on the spread of this phenomenon first focused on search engines and news aggregators (e.g., Google News), but has since included recommendations on YouTube ⇄ as well.

### Which effects have been confirmed?



For researchers outside of social media companies, it is very difficult to investigate filter bubbles as Eli Pariser described them, based on recommendation systems and personalization. This starts with the degree to which content has been customized for an individual user (which could later influence their opinions). Just as the users themselves, independent researchers cannot really assess to what extent suggested content differs between users and what role platform algorithms play relative to the users' selection behavior. Studies based on simulations with artificially created accounts typically show low levels of customization but it remains unclear if this is due to a lack of authenticity of the simulated usage behavior.

Studies on possible filter bubbles in **search engines** show low levels of customization in the search results for many topics based on the search history of (real or simulated) users. Customization based on the approximate location of the IP address from which a search query is started is, however, well documented for different types of topics (e.g., the dates for school holidays that a search engine displays may be matched to the state where a user is located—a type of customization that is probably seen by many users as very convenient). Differences in search results for people with differing political opinions also exist: people with different opinions use different search terms for many political topics, independent of potentially active customization algorithms, which may result in them receiving widely different search results [16].

Additionally, with regard to customization of **online news**, algorithmic curation that takes into account previous usage behavior does result in a systematically more narrow selection of available news than a traditional news medium with editorial selection [17][18]. Usage data from the US show that across party lines, users see mostly the same online news [6].

Based on these low levels of customization, effects on **opinion formation** are not to be expected. In fact, while experimental studies have documented some in the short term [8], these are not based on everyday usage behavior [7].

## What can be done about them?

Users do not have to fear the filter bubble as Eli Pariser originally described it, as the phenomenon is much less marked than he assumed. Usage studies that exceed social media or search engines additionally show that most people use diverse sources of information, including television, radio, and newspapers (digital or printed). A biased filtering of content across many different media outlets – that are often competitors – is unlikely; however, people who purposefully want to consume only one-sided information can freely do so in a fully digital media environment. Yet, this type of filtering is based more on selection behavior than on personalization or recommendation systems.

It is possible that the public debate around Pariser's best-selling book as well as the later discussion of possible rabbit holes ⇔ have contributed to platforms such as Google and YouTube changing their algorithms and reducing customization [19]. In this sense, a critical public debate is helpful in limiting the potential negative effects of digital platforms. Since 2023, the Digital Services Act of the European Union obligates large online platforms to inform users of the recommendation systems they use.

## What is it?

Microtargeting is addressing very small groups of (potential) customers via advertising, typically on digital platforms. This is attractive for advertisers who, for example, want to market a product or service that is only intended for very few people. Microtargeting also allows advertisers to show different messages to different target groups. Such practices are heavily criticized in the context of political campaigning as a potential manipulation of voters, because it can mean that parties or candidates send out contradictory political messages while the recipients have no way of knowing that the campaigners make very different promises to different groups of people ↔ .

## Since when does it exist and how widespread is it?

Since the 1990s, the spread of the Internet and other electronic systems (e.g., electronic cash registers in stores or electronic telephone systems) in many countries has created very detailed data about the everyday behavior of consumers. Many social media platforms additionally contain information about preferences, hobbies, or social contacts the users posted themselves. These large amounts of data can be analyzed in much more detail than was previously possible, which means that the meaning of “micro-” in microtargeting has shifted over time to imply ever smaller target groups. Large digital datasets of behavioral information can also be combined with survey data to tailor advertising messages to very small target groups.

With regard to politics, Panagopoulos, among others, has named the US presidential election in 2000 as the first election in which microtargeting was employed [20]. At the time, the Republican party aimed their campaign at the core of their electorate to mobilize them, rather than trying to address voters across the entire population. Barack Obama’s campaigns in 2008 and 2012 subsequently went much further and specifically targeted potential voters based on location, demographics, and previous voting behavior. Since the Brexit referendum and the US elections in 2016, microtargeting that includes so-called psychographic analyses, which show social media users political messages tailored to their party preference as well as their personality, has been employed.

In Germany, jurisdiction around privacy and data protection limits microtargeting much more strictly than in the US. However, German law also allows campaigners to show very specific advertising on social media to people who, for example, might have liked a given party (or other potentially relevant topics or accounts) [21].

## Which effects have been confirmed?

From the advertisers’ or their clients’ point of view, microtargeting does not necessarily aim at raising the effectiveness of a campaign, but rather at optimizing the efficiency of the invested means. Its core tenet is that advertising is targeted specifically at persuadable individuals instead of blanketing the general public with broadcasting or billboards, given that most people will not be convinced by such messaging to change their behavior. Barack Obama’s 2012 reelection campaign, for instance, focused heavily on contested states or electoral districts to address undecided voters and mobilize core Democrat voters [22]. On social networking sites like Facebook, the campaign used advertising options that displayed a number of different messages to very small target groups. After disappointing results during the 2010 mid-term elections, the incumbent president was reelected two years later, a success which has been repeatedly attributed to the systematic and

continuous data collection as well as fine-grained and frequently updated analyses of his campaigners.

The state of research on the effects of psychographics in microtargeting is so far ambiguous. In a Dutch study, voting intention of partisans could be strengthened through advertising that matched their personalities [23]. A US study, however, found no difference in voting intention after showing participants personality-congruent versus -incongruent messages [24]. In a German study, matching political advertising to people's attitudes was a relevant factor, while fitting the message to their personality traits did not systematically change their voting intention [25].

## What can be done about it?

Zarouali et al. [23] see microtargeting based on political attitudes and personality traits as highly manipulative because a close fit between an advertising message and a user could override the latter's usual ways of bracing themselves against ads. Accordingly, they advocate for legally binding measures of protection. In the European Union and thus in Germany, the General Data Protection Regulation as well as the Digital Services Act set clear boundaries for microtargeting. Digital platforms are forbidden from showing users advertisements based on profiles that include, among others, political opinions, ethnic background, or sexual orientation. In addition, transparency norms for customizing advertising have been established (see also measures against filter bubbles ↔ and rabbit holes ↔). Whether or not these legal regulations will effectively protect users from manipulative and highly specific targeted messages remains to be seen—potentially in comparison to other countries that continue to allow (political) microtargeting.

Profile: Rabbit holes

## Are users becoming more radical through YouTube's recommendations?

### What are they?

“Rabbit holes” is the term applied to supposed chains of content recommendations that become more and more extreme over time, leading users from innocuous to radical content. The metaphor is based on the fictional character Alice, who falls ever deeper into a literal rabbit hole and ends up in wonderland. The assumptions behind this alleged phenomenon are similar to those of filter bubbles ↔: recommendation algorithms suggest content that fits previously used content. The debate about rabbit holes focuses mainly on YouTube. Its effects are also described in similar terms to filter bubbles: it is assumed that users' opinions can become more and more radical through watching extreme videos, leading up to political extremism. Accordingly, the main concern is extreme political content, including so-called alternative or radical offerings. Occasionally, other fringe communities are also discussed with regard to possible rabbit holes.

### Since when do they exist and how widespread are they?

The phenomenon of rabbit holes has been receiving attention since about the mid-2010s. Before that time, scholars had investigated deviant content and the respective, often tight-knit online communities built around it, first in the blogosphere and then on YouTube. A prominent example of this research investigates false information about vaccinations as well as skeptics and opponents of vaccinations. Parallel to the rise of right-wing populism in different Western countries, the focus has subsequently shifted to politically extreme content. The *New York Times* and other media have [prominently published op-eds](#) and [features on the topic](#). The accidental discovery of recommendations networks among YouTube videos showing children with little clothing raised even more awareness of the phenomenon. Apparently, people with pedophile tendencies were able to use the recommendation algorithm in order to link together and recommend to one another videos of, among others, playing children that parents had innocently posted. YouTube has since changed its algorithm several times to ensure that this type of abuse (and also radicalization) cannot occur.

### Which effects have been confirmed?

It is self-evident that radical and otherwise deviant content is available on many digital platforms. Like many other platform providers, YouTube holds the position that a wide range of content and opinions can and shall be posted. At the same time, the platform's community guidelines of course also state limits for what types of content are permissible.

Due to the sheer amount of available content and personalized video recommendations, it is difficult to assess, first, **how often radical content is actually watched** and, second, to what degree users' opinions become more radical because they followed **chains of increasingly extreme recommendations**. A US study combining tracking data with a survey of users' opinions found that "alternative" or extremist content together make up only 2% of all video recommendations [19]. These were mostly suggested to users who were already watching videos from the respective category or had subscribed to these types of channels. An analysis of the survey showed that these users were far more likely to hold sexist and racist attitudes than others. Based on long-term patterns of usage, especially channel subscriptions, it is more plausible to assume that people with radical opinions choose to watch radical content rather than to think that the small proportion of recommendations for radical videos sways a large number of users in their attitudes. Providing people with hateful views **confirmation** of their opinions through YouTube videos can also pose a problem for the safety and peaceful coexistence of the members of a given society. But it appears to be the case that people who are already deep in the proverbial rabbit hole tend to receive radical content recommendations that keep them there, rather than that average users are being pulled into said holes in large numbers.

### What can be done about them?

Similar to [filter bubbles](#) ⇔ , public attention and scandal have led YouTube to change its algorithm to prevent chains of more and more radical content recommendations. Since the European Union's Digital Services Act (DSA) went into effect in 2023, very large online platforms have to provide users information about the workings of their recommendation systems and offer options to customize the latter (Article 27). Accordingly, [TikTok](#), as well as [Facebook and Instagram](#), allow users to opt out of personalization of content. The long-term effects of these legal and technical changes will have to be monitored by future research.

Profile: Russian propaganda (by Elizaveta Kuznetsova)

## How does Russian propaganda affect opinion formation in other countries?

### What is it?

Russian propaganda refers to a variety of activities performed by actors directly or covertly sponsored by the Russian government to influence public opinion abroad. Particularly, Russia has been shown to “muddy the waters” of public discourse and to sow distrust in Western institutions [39]. The Kremlin’s propaganda methods often employ blatantly false claims and conspiratorial content but can also include factually correct statements presented in a misleading context and therefore aimed at deceiving the public. More subtle techniques include satire and irony that are used to relativize facts. In Europe, Russian propaganda particularly targets vulnerable communities dissatisfied with the status of European democracies, on both the far-right and the far-left ends of the political spectrum.

### Since when does it exist and how widespread is it?

Russian propaganda is not a new phenomenon. Having its roots in the Soviet period, many contemporary methods employ elements developed during WWII and the Cold War. However, both the narratives and the techniques have evolved substantially and Russian propaganda strategies have adjusted to the political communication environments worldwide. Specifically, Kremlin information influence is now more contextual and adaptive compared to its Soviet predecessor. At the same time, it utilizes technological advancements in automation and more recently AI. Kremlin propaganda intensified in the period shortly preceding and after Russia’s full-scale invasion of Ukraine [40].

Exposure to Russian propaganda is particularly likely in online information environments. The Kremlin has been known to use a wide range of computational propaganda tools, such as bots ↔ and trolls that are present on most social media platforms [41][42]. Additionally, Russian propaganda sources include traditional media outlets operating online, such as RT and Sputnik, but also alternative news media that are not directly linked to the state. At the same time, there is a plethora of social media accounts, often anonymous, present on Telegram, spreading Russian propaganda messages, particularly related to far-right networks.

### Which effects have been confirmed?

Several studies conducted in Russia have pointed out the effect of national television on political preferences and voting behavior in the country [43]. However, there are substantial methodological difficulties in measuring effects of Russian propaganda in democratic contexts. This is primarily connected to the complexity of the mechanisms of opinion formation. Moreover, Russian propaganda often exploits pre-existing vulnerabilities in democratic societies and disentangling its effects from other factors presents a challenge. Lastly, propaganda is a strategic process aimed at exposing publics to persuasive content over long periods of time. Therefore, one-off experimental studies are unable to capture longitudinal dynamics.

That said, we do have growing evidence of different communities online being increasingly exposed to Russian propaganda. Studies showed a substantial presence of bots spreading information about the 2016 US presidential election among conservative users on Twitter [44]. Audience studies conducted in Europe, meanwhile, have revealed that users of so-called alternative media consistently include Russian propaganda sources in their information diets [45]. There is also evidence that some Russian propaganda narratives receive particular resonance in conservative communities worldwide. For example, the infamous disinformation story promoted by the Kremlin claiming that the US military was funding the development of biological weapons in Ukraine has been organically propagated by anti-vax English-speaking users on Twitter [46].

## What can be done about it?

Although there have been multiple proposals on preventing Russian propaganda from influencing societies in democratic countries, we do not yet have a clear solution. This is firstly connected to the lack of knowledge on the actual effect of Russian propaganda on foreign audiences. Secondly, most of the strategies focus on how to tackle false information ⇔ without being able to counter more diffuse forms of influence. So far, the existing suggestions include:

- Media literacy. Researchers have pointed out specific psychological traits that make individuals more vulnerable to misinformation, including a lack of critical thinking and a tendency to depend on intuition. They have also demonstrated how accuracy prompts and tips on digital literacy can be effective tools in fighting against false information [47].
- Inoculation. Another recommendation for enhancing the quality of information involves countering misleading stories, for instance, by employing strategies like inoculation and pre-bunking (instead of de-bunking after the fact) [48], which have proven effective in building resilience against false information [49].
- Automated removal of content. Within environments where algorithms mediate the flow of information, advancements have been made in creating tools that automatically detect and eliminate the dissemination of false information. However, these are still in development and are not to be fully relied on [50][51].



Dr. Elizaveta Kuznetsova

Elizaveta Kuznetsova is the leader of the research group “[Platform Algorithms and Digital Propaganda](#)” at the Weizenbaum Institute. Her primary focus is on AI and algorithms on platforms, digital propaganda, and international media. She is interested in legacies of historical practices of propaganda and political communication as well as new developments in the digital world.

[Go to profile ↗](#)

## References

- [1] J. Kaiser und A. Rauchfleisch, „Birds of a feather get recommended together: Algorithmic homophily in YouTube's channel recommendations in the United States and Germany", *Soc. Media Soc.*, Bd. 6, Nr. 4, S. 205630512096991, 2020 <https://doi.org/10.1177/2056305120969914>
- [2] J. Behre, S. Hölzig, und J. Möller, „Reuters Institute Digital News Report 2023: Ergebnisse für Deutschland", *Arbeitspapiere Hans-Bredow-Inst.*, 2023 <https://doi.org/10.21241/SSOAR.86851>
- [3] W. Schulz, „Prekäre Mobilisierung und defizitäre Kommunikation: Was von der Kampagne bei wem (nicht) ankam", in *Die (Massen-)Medien im Wahlkampf*, C. Holtz-Bacha, Hrsg., Wiesbaden: Springer VS, 2023, S. 317–338 [https://doi.org/10.1007/978-3-658-38967-3\\_12](https://doi.org/10.1007/978-3-658-38967-3_12)
- [4] R. Kösters und O. Jandura, „Politische Kommunikation in heterogenen Lebenswelten. Kommunikationspraxis in politischen Milieus und Bedingungen ihrer Integration", *SCM Stud. Commun. Media*, Bd. 7, Nr. 2, S. 129–185, 2018 <https://doi.org/10.5771/2192-4007-2018-2-1>
- [5] M. Scharnow, F. Mangold, S. Stier, und J. Breuer, „How social network sites and other online intermediaries increase exposure to news", *Proc. Natl. Acad. Sci.*, Bd. 117, Nr. 6, S. 2761–2763, 2020 <https://doi.org/10.1073/pnas.1918279117>
- [6] A. M. Guess, „(Almost) everything in moderation: New evidence on Americans' online media diets", *Am. J. Polit. Sci.*, Bd. 65, Nr. 4, S. 1007–1022, 2021 <https://doi.org/10.1111/ajps.12589>
- [7] B. Stark, M. Magin, und P. Jürgens, *Ganz meine Meinung? Informationsintermediäre und Meinungsbildung – Eine Mehrmethodenstudie am Beispiel von Facebook*. Düsseldorf: Landesanstalt für Medien Nordrhein-Westfalen, 2017 [https://www.medienanstalt-nrw.de/fileadmin/user\\_upload/lfm-nrw/Foerderung/Forschung/Dateien\\_Forschung/L194-Ganz-meine-Meinung\\_LfM-Doku55.pdf](https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Foerderung/Forschung/Dateien_Forschung/L194-Ganz-meine-Meinung_LfM-Doku55.pdf)
- [7] B. Stark, M. Magin, und P. Jürgens, *Ganz meine Meinung? Informationsintermediäre und Meinungsbildung – Eine Mehrmethodenstudie am Beispiel von Facebook*. Düsseldorf: Landesanstalt für Medien Nordrhein-Westfalen, 2017 [https://www.medienanstalt-nrw.de/fileadmin/user\\_upload/lfm-nrw/Foerderung/Forschung/Dateien\\_Forschung/L194-Ganz-meine-Meinung\\_LfM-Doku55.pdf](https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Foerderung/Forschung/Dateien_Forschung/L194-Ganz-meine-Meinung_LfM-Doku55.pdf)
- [8] E. Kubin und C. von Sikorski, „The role of (social) media in political polarization: A systematic review", *Ann. Int. Commun. Assoc.*, Bd. 45, Nr. 3, S. 188–206, 2021 <https://doi.org/10.1080/23808985.2021.1976070>
- [9] R. Heiss, C. von Sikorski, und J. Matthes, „Populist Twitter posts in news stories: Statement recognition and the polarizing effects on candidate evaluation and anti-immigrant attitudes", *Journal. Pract.*, Bd. 13, Nr. 6, S. 742–758, 2019 <https://doi.org/10.1080/17512786.2018.1564883>
- [10] T. Flew, *Regulating platforms*. in *Digital media and society series*. Cambridge, UK: Polity, 2021
- [11] D. Stegmann, L. Zieringer, B. Stark, und C. Reinemann, „Meinungsvielfalt, Meinungsmacht, Meinungsbildung. Zum (ungeklärten) Verhältnis zentraler Begriffe der deutschen Medienkonzentrationskontrolle", *UFITA Arch. Für Medien. Medienwiss.*, Bd. 86, Nr. 1, S. 38–70, 2022 <https://doi.org/10.5771/2568-9185-2022-1-38>
- [12] M. Appel und N. Doser, „Fake News", in *Die Psychologie des Postfaktischen: Über Fake News, „Lügenpresse", Clickbait & Co.*, M. Appel, Hrsg., Berlin: Springer, 2020, S. 9–20
- [13] H. Allcott und M. Gentzkow, „Social media and fake news in the 2016 election", *J. Econ. Perspect.*, Bd. 31, Nr. 2, S. 211–236, Mai 2017 <https://doi.org/10.1257/jep.31.2.211>
- [14] U. K. H. Ecker und L. C. Ang, „Political attitudes and the processing of misinformation corrections", *Polit. Psychol.*, Bd. 40, Nr. 2, S. 241–260, 2019 <https://doi.org/10.1111/pops.12494>
- [15] S. Altay, M. Berriche, und A. Acerbi, „Misinformation on misinformation: Conceptual and methodological challenges", *Soc. Media Soc.*, Bd. 9, Nr. 1, S. 205630512211504, 2023 <https://doi.org/10.1177/20563051221150412>
- [16] A. G. Ekström, G. Madison, E. J. Olsson, und M. Tsapos, „The search query filter bubble: Effect of user ideology on political leaning of search results through query selection", *Inf. Commun. Soc.*, S. 1–17, Juli 2023 <https://doi.org/10.1080/1369118X.2023.2230242>
- [17] M. Haim, A. Graefe, und H.-B. Brosius, „Burst of the filter bubble? Effects of personalization on the diversity of Google News", *Digit. Journal.*, Bd. 6, Nr. 3, S. 330–343, 2018 <https://doi.org/10.1080/21670811.2017.1338145>
- [18] J. Möller, D. Trilling, N. Helberger, und B. Van Es, „Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity", *Inf. Commun. Soc.*, Bd. 21, Nr. 7, S. 959–977, 2018 <https://doi.org/10.1080/1369118X.2018.1444076>
- [19] A. Y. Chen, B. Nyhan, J. Reifler, R. E. Robertson, und C. Wilson, „Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels", *Sci. Adv.*, Bd. 9, Nr. 35, S. eadd8080, 2023 <https://doi.org/10.1126/sciadv.add8080>
- [19] A. Y. Chen, B. Nyhan, J. Reifler, R. E. Robertson, und C. Wilson, „Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels", *Sci. Adv.*, Bd. 9, Nr. 35, S. eadd8080, 2023 <https://doi.org/10.1126/sciadv.add8080>
- [20] C. Panagopoulos, „All about that base: Changing campaign strategies in U.S. Presidential elections", *Party Polit.*, Bd. 22, Nr. 2, S. 179–190, 2016 <https://doi.org/10.1177/1354068815605676>
- [21] O. Papakyriakopoulos, S. Hegelich, M. Shahrezayeh, und J. C. M. Serrano, „Social media and microtargeting: Political data processing and the consequences for Germany", *Big Data Soc.*, Bd. 5, Nr. 2, S. 205395171881184, 2018 <https://doi.org/10.1177/2053951718811844>
- [22] S. Issenberg, „How Obama's team used big data to rally voters", *MIT Technol. Rev.*, 2012 <https://www.technologyreview.com/2012/12/19/114510/how-obamas-team-used-big-data-to-rally-voters/>
- [23] B. Zarouali, T. Dobber, G. De Pauw, und C. De Vreese, „Using a personality-profiling algorithm to investigate political microtargeting: Assessing the persuasion effects of personality-tailored ads on social media", *Commun. Res.*, Bd. 49, Nr. 8, S. 1066–1091, 2022 <https://doi.org/10.1177/0093650220961965>
- [24] L. J. Krotzek, „Inside the voter's mind: The effect of psychometric microtargeting on feelings toward and propensity to vote for a candidate", *Int. J. Commun.*, Bd. 13, S. 3609–3629, 2019

- [25] H. Decker und N. Krämer, „Is Personality Key? Persuasive effects of prior attitudes and personality in political microtargeting“, *Media Commun.*, Bd. 11, Nr. 3, S. 250–261, 2023 <https://doi.org/10.17645/mac.v11i3.6627>
- [26] V. Trott, N. Li, R. Fordyce, und M. Andrejevic, „Shedding light on 'dark' ads“, *Continuum*, Bd. 35, Nr. 5, S. 761–774, 2021 <https://doi.org/10.1080/10304312.2021.1983258>
- [27] E. L. Briant, „LeaveEU: Dark money, dark ads and data crimes“, in *The Sage handbook of propaganda*, P. Baines, N. O'Shaughnessy, und N. Snow, Hrsg., London: Sage, 2020, S. 532–547 <https://doi.org/10.4135/9781526477170>
- [28] C. R. Sunstein, *Republic.com*. Princeton, NJ: Princeton University Press, 2001
- [29] C. R. Sunstein, *#republic. Divided democracy in the age of social media*. Princeton, NJ: Princeton University Press, 2018
- [30] S. Goel, W. Mason, und D. J. Watts, „Real and perceived attitude agreement in social networks“, *J. Pers. Soc. Psychol.*, Bd. 99, Nr. 4, S. 611–621, 2010 <https://doi.org/10.1037/a0020697>
- [31] D. J. Isenberg, „Group polarization: A critical review and meta-analysis“, *J. Pers. Soc. Psychol.*, Bd. 50, Nr. 6, S. 1141–1151, 1986 <https://doi.org/10.1037/0022-3514.50.6.1141>
- [32] S. Messing und S. J. Westwood, „Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online“, *Commun. Res.*, Bd. 41, Nr. 8, S. 1042–1063, 2014 <https://doi.org/10.1177/0093650212466406>
- [33] R. Fletcher und R. K. Nielsen, „Are people incidentally exposed to news on social media? A comparative analysis“, *New Media Soc.*, Bd. 20, Nr. 7, S. 2450–2468, 2018 <https://doi.org/10.1177/1461444817724170>
- [34] G. Eady, J. Nagler, A. Guess, J. Zilinsky, und J. A. Tucker, „How many people live in political bubbles on social media? Evidence from linked survey and Twitter data“, *Sage Open*, Bd. 9, Nr. 1, S. 1–25, 2019 <https://doi.org/10.1177/2158244019832705>
- [35] V. Gehrau, „Issue diversity in the internetage. Changes in nominal issue diversity in Germany between 1994 and 2005“, *Stud. Commun. Media*, Bd. 2, Nr. 1, S. 129–142, 2013 <https://doi.org/10.5771/2192-4007-2013-1-129>
- [36] V. Gehrau, „Gespräche über Medien in Zeiten von Mobilkommunikation und sozialen Onlinenetzen. Eine Analyse von Befragungsdaten aus den Jahren 2015, 2016 und 2017 im Vergleich zu 1996/97 und 2007“, *Publizistik*, Bd. 64, Nr. 1, S. 37–58, 2019 <https://doi.org/10.1007/s11616-018-00465-z>
- [37] P. Norris, *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge, UK: Cambridge University Press, 2001
- [38] J. van Dijk, *The digital divide*. Cambridge: Polity, 2020
- [39] P. Pomerantsev, *Nothing is True and Everything is Possible: The Surreal Heart of New Russia*. New York: PublicAffairs, 2014
- [40] A. Litvinenko, „Propaganda on demand: Russia's media environment during the war in Ukraine“, *Glob. Media J. - Ger. Ed.*, Bd. 12, Nr. 2, S. 1–14, 2022 <https://doi.org/10.22032/dbt.55518>
- [41] Y. Golovchenko, C. Buntain, G. Eady, M. A. Brown, und J. A. Tucker, „Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Presidential Election“, *Int. J. Press.*, Bd. 25, Nr. 3, S. 357–389, Juli 2020 <https://doi.org/10.1177/1940161220912682>
- [42] D. Stukal, S. Sanovich, und J. Tucker, „Detecting Bots on Russian Political Twitter. Big Data“, *Big Data*, Bd. 5, Nr. 4, S. 31–324, 2017
- [43] L. Peisakhin und A. Rozenas, „Electoral Effects of Biased Media: Russian Television in Ukraine“, *Am. J. Polit. Sci.*, Bd. 62, Nr. 3, S. 535–550, 2018
- [44] A. Badawy, A. Addawood, K. Lerman, und E. Ferrara, „Characterizing the 2016 russian IRA Influence Campaign. Social Network Analysis and Mining“, *Soc. Netw. Anal. Min.*, Bd. 9, Nr. 1, S. 1–11, 2019
- [44] A. Badawy, A. Addawood, K. Lerman, und E. Ferrara, „Characterizing the 2016 russian IRA Influence Campaign. Social Network Analysis and Mining“, *Soc. Netw. Anal. Min.*, Bd. 9, Nr. 1, S. 1–11, 2019
- [45] C. Wagnsson, T. Blad, und A. Hoyle, „Keeping an Eye on the Other Side? RT, Sputnik, and Their Peculiar Appeal in Democratic Societies“, *Int. J. Press.*, Bd. 0, Nr. 0, S. 19401612221147492, Jan. 2023 <https://doi.org/10.1177/19401612221147492>
- [46] I. Alieva, J. D. Moffitt, und K. M. Carley, „How disinformation operations against Russian opposition leader Alexei Navalny influence the international audience on Twitter“, *Soc. Netw. Anal. Min.*, Bd. 12, Nr. 1, S. 80, Juli 2022 <https://doi.org/10.1007/s13278-022-00908-6>
- [47] A. A. Arechar u. a., „Understanding and combatting misinformation across 16 countries on six continents“, *Nat. Hum. Behav.*, Bd. 7, Nr. 9, S. 1502–1513, Sep. 2023 <https://doi.org/10.1038/s41562-023-01641-6>
- [48] C. Lewandowsky und J. Cook, *The Conspiracy Theory Handbook*. University of Bristol & George Mason University, 2020 <https://digitalcommons.unl.edu/scholcom/246/>
- [49] R. Maertens, J. Roozenbeek, M. Basol, und S. van der Linden, „Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments“, *J. Exp. Psychol. Appl.*, Bd. 27, Nr. 1, S. 1–16, 2021 <https://doi.org/10.1037/xap0000315>
- [50] J. C. Aguerri und M. Santisteban, „The algorithmic responses to disinformation: a suitable pathway?“, *Justice Power Resist.*, Bd. 5, Nr. 3, S. 299–306, Dez. 2022 <https://doi.org/10.1332/CQNF2293>
- [51] F. Saurwein und C. Spencer-Smith, „Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe“, *Digit. Journal.*, Bd. 8, Nr. 6, S. 820–841, Juli 2020 <https://doi.org/10.1080/21670811.2020.1765401>
- [52] M. Orabi, D. Mouheb, Z. Al Aghbari, und I. Kamel, „Detection of bots in social media: A systematic review“, *Inf. Process. Manag.*, Bd. 57, Nr. 4, S. 102250, 2020 <https://doi.org/10.1016/j.ipm.2020.102250>
- [53] F. Martini, P. Samula, T. R. Keller, und U. Klinger, „Bot, or not? Comparing three methods for detecting social bots in five political discourses“, *Big Data Soc.*, Bd. 8, Nr. 2, S. 205395172110335, 2021 <https://doi.org/10.1177/20539517211033566>
- [54] O. Boichak, J. Hemsley, S. Jackson, R. Tromble, und S. Tanupabrungsun, „Not the bots you are looking for: Patterns and effects of orchestrated interventions in the U.S. and German elections“, *Int. J. Commun.*, Bd. 15, S. 814–839, 2021



[55] S. Pedrazzi und F. Oehmer, „Communication rights for social bots? Options for the governance of automated computer-generated online identities“, J. Inf. Policy, Bd. 10, S. 549–581, 2020 <https://doi.org/10.5325/jinfopoli.10.2020.0549>

[56] C. E. Robertson, N. Pröllochs, K. Schwarzenegger, P. Pärnamets, J. J. Van Bavel, und S. Feuerriegel, „Negativity drives online news consumption“, Nat Hum Behav, Bd. 7, S. 812–822, 2023 <https://doi.org/10.1038/s41562-023-01538-4>

[57] D. Rozado, R. Hughes, und J. Halberstadt, „Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models“, PLoS ONE, Bd. 17, Nr. 10, S. e0276367, 2022 <https://doi.org/10.1371/journal.pone.0276367>

# Digital democracy – Do social media steer opinion formation?

👤 PD Dr. Merja Mahrt

Weizenbaum Institute

Version 📄 6/25/2024

Deutsch ↗

## Contact

PD Dr. Merja Mahrt. merja.mahrt@weizenbaum-institut.de

## Copyright note

The texts on this page are licensed under the Creative Commons License "Attribution-NonCommercial 4.0 International" (CC BY NC Open link). If you intend to make commercial use, please contact us.

CC BY NC

## Conceptualization, design, development

Atelier Hurra - Conceptualization, design  
Webseite

MADEFUL® GmbH - Projectmanagement, conceptualization, design  
Webseite

Uninspired Studio - Development, conceptualization, design  
Webseite

## Image credits

Animated illustrations (own design, CC BY-NC)

Photo of the author, Dr. Elizaveta Kuznetsova (© 2023 Weizenbaum Institute e.V)

Photo of the author, PD Dr. Merja Mahrt (© 2019 Alexander Vejnovic)