Rathenau Instituut

Scrolling towards the ballot box

The role of recommendation algorithms in election interference

Report

Authors

Luuk Ex, Quirine van Eeden, Wouter Nieuwenhuizen, Mariette van Huijstee, with contribution of Joost Gerritsen

Illustrations
Laura Marienus/Rathenau Instituut

Cover photo
People queuing at a polling station in The Hague in 2021. Photo: Shutterstock

Please cite as:
Rathenau Instituut (2026). Scrolling towards the ballot box – The role of recommendation algorithms in election interference. (Authors: Ex, L., Van Eeden, Q., Nieuwenhuizen, W., & Van Huijstee, M., with contributions from: Gerritsen, J.)

Foreword

In 2024, Romanian elections were declared invalid after social media posts were allegedly manipulated to favour a particular candidate. In the Netherlands, too, there are concerns about foreign interference in elections. The Standing Committee on Digital Affairs asked us to investigate the role that the recommendation algorithms of major social media platforms might play in election interference.

These algorithms form a key link in determining which posts people see online. In doing so, the algorithms act as a kind of megaphone. They determine which voices are amplified and which are not. Content that people keep looking at and that elicits a response (likes, comments, shares, viewing time) is amplified. People then stay on a platform longer, which is good for advertising revenue.

There are various ways in which malicious foreign actors can cleverly exploit these recommendation algorithms to influence the political debate in the Netherlands. For example, they can use hundreds of fake accounts simultaneously to send out a similar message, thereby misleading people about the popularity of the content. Or they can pay influencers to amplify a message.

Many different measures have already been taken to tackle interference via platforms, both by the legislator and by the platforms themselves. However, our study shows that these are only partially effective. We propose three courses of action to reduce the risk of interference.

The first concerns platform design. For example, alternative recommendation algorithms may limit opportunities for interference. The second course of action concerns strengthening the information position of supervisory bodies, researchers, journalists and politicians. The third approach concerns how we can increase society's resilience so that it is better able to withstand attempts at interference, regardless of what the platforms do.

We hope that these courses of action will help the House of Representatives to mitigate the risks of interference in elections via social media platforms.

Prof. Dr. Ir. Eefje Cuppen
Director of the Rathenau Instituut

Summary

At the request of the House of Representatives Committee on Digital Affairs, the Rathenau Instituut conducted research into the role that recommendation algorithms of major social media platforms can play in election interference.

We analysed academic literature, legal literature and grey literature such as journalistic sources, reports and court records. In addition, we interviewed representatives from a number of platforms (in writing). We also spoke to policy officers involved in elections, platform policy and interference. Furthermore, during an expert session, we interviewed and consulted experts from fields including political communication, computer science, law and media studies.

Interference
Interference is the exertion of influence by or on behalf of a foreign state actor that is contrary to the sovereignty, values or interests of the country being influenced. Such actions are coercive, covert, deceptive or corrupting. Interference can pit people against one another, influence the outcome of elections and sow doubt about the democratic process.

Interference via social media platforms
We focused on a relatively new form that is becoming increasingly accessible and sophisticated: interference via the recommendation algorithms of social media platforms.

By interference via social media platforms, we mean the orchestrated use of multiple accounts, pages or platforms by foreign state actors and/or their representatives to increase or decrease the visibility of certain content. They do this by exploiting the recommendation algorithms of social media platforms. In doing so, the actors involved attempt to deliberately mislead the recipient of the content regarding the identity, motives, actions or connections of the actors involved and/or the popularity of the content. The ultimate aim is to influence public opinion or elections.

Recommendation algorithms
Recommendation algorithms are a collection of systems that sort and rank content based on data, with the aim of recommending potentially interesting content to users. Recommendation algorithms are a solution to the problem of information overload on the internet: from all the available content, attempts to show you something you might find enjoyable, interesting or otherwise relevant.

Platforms and their developers favour content that keeps users on the platform for longer and encourages them to visit frequently. Since 2016, many major platforms have switched to ranking content based on expected engagement (likes, comments, time spent, etc.). After all, the more engagement there is, the higher the advertising revenue.

The consequence of recommendation algorithms based on engagement is that certain content can be made more visible (*amplified,* boosted).

The tools of interference
Incitement or manipulative content disseminated to influence voters is designed to encourage user engagement, such as liking, sharing or rewatching a video. Malicious (foreign) actors can exploit the opportunities offered by engagement-based recommendation algorithms to achieve a wide reach.

This study shows that the range of tools available for interference activities on social media platforms is extensive and diverse, and that the means of interference are becoming increasingly accessible and sophisticated. Social media platforms are constantly changing, which means the toolkit is changing too. As social media platforms increasingly use recommendation algorithms, malicious actors are capitalising on this. For example, they use hyperactive accounts and coordinate their posts and hashtags.

Existing measures are only partially effective
The legislator has developed tools to tackle various aspects of interference. Examples include increasing the accountability of platforms, intervening in their design and enhancing transparency regarding circulating content. And platforms have guidelines, known as *community guidelines*, regarding which content and activities are prohibited on their platforms.

However, experts point out that it is precisely simple measures, such as removing fake accounts, that platforms do not seem to be taking. Recommendation algorithms could also be adjusted to make them less susceptible to interference activities.

Furthermore, it is difficult for independent researchers and journalists to gain a clear understanding of the nature, scale and effectiveness of interference via social media platforms. This is because research depends on the cooperation of the platform companies. As a result, it is unclear which content is circulating more widely and who it is reaching. This makes it difficult to assess whether the measures taken by platforms are effective. Recent examples in the Netherlands and Europe make it clear that recommendation algorithms offer a range of

opportunities for interference activities and thus pose a risk to online political debate.

Three courses of action to reduce the risk of interference
To reduce the risks of interference, this report outlines three courses of action: (1) changing platform design, (2) strengthening the information position, (3) increasing resilience. A combination of these courses of action is desirable.

The first approach requires adjustments to recommendation algorithms and a focus on more fundamental changes to the landscape of social media platforms. The second approach concerns strengthening the information position of regulators, supervisory bodies, researchers, journalists, politicians and policymakers. The third approach calls for the continued protection of electoral processes and current legislation, and for investment in independent researchers and resilient citizens.

The three courses of action:

1. **Changing platform design**
    a. Stimulate platforms to adapt recommendation algorithms.
    b. Intervene in governance and revenue models.
    c. Focus on decentralisation and interoperability.

2. **Strengthen information access**
    a. Enable access to platform data for government, supervisory bodies and researchers.
    b. Institutionalise independent, transparent and proactive detection of interference.

3. **Increase resilience**
    a. Strengthen electoral procedures.
    b. Protect existing legislation and clarify where necessary.
    c. Invest in independent journalism.
    d. Invest in digital citizenship and media literacy.

Contents

Glossary

Recommendation algorithms: Recommendation algorithms are a set of systems that sort and rank content based on data, with the aim of recommending potentially interesting content to users.

Amplification: There are various definitions of the concept of amplification. When we use the term amplification, we refer to what algorithm researchers Thorburn, Stray and Bengan call 'relative algorithmic amplification': 'A change in the distribution of content under a recommendation algorithm, compared to an alternative, whilst user behaviour remains constant' (Thorburn et al., 2023).

This definition makes it clear that, when discussing amplification, it must always be made clear that this refers to amplification *relative* to an alternative scenario, and that one must assume that users' behaviour or preferences remain the same.

Content moderation: Content moderation is the process of determining the conditions under which content, such as text, photos and videos, is permitted or prohibited within online environments (Rathenau Instituut, 2025a). This process involves the assessment of user-generated content on the platform.

Engagement: In this report, we use the definition provided by algorithm researchers (Stray et al., 2024, p. 13): 'A set of user behaviours that arise during normal use of the platform and are assumed to be associated with value for the user, the platform or other stakeholders.' In the context of social media platforms, this often refers to: likes, shares, comments, time spent, and other interactions that indicate how valuable or relevant the content is to the user or the platform.

User retention: Tracking and measuring the amount of time a user spends on a social media platform. It is measured by looking at how long a user remains active during a session, and whether the user remains active on the platform in the long term.

Interference: Interference refers to foreign influence exerted by or on behalf of a foreign state actor that is contrary to the sovereignty, values or interests of a state. Such actions are coercive, covert, deceptive or corrupt in nature. In the context of this investigation, these activities are aimed at influencing the elections in the Netherlands. Interference in the context of elections means that a state undertakes activities to influence an election in the Netherlands by favouring or disadvantaging a party participating in that election.

Interference via social media: By interference via social media platforms, we mean the orchestrated use of multiple accounts, pages or platforms by foreign state actors and/or their representatives to increase or decrease the visibility of certain content. They do this by exploiting the recommendation algorithms of social media platforms. The actors involved attempt to deliberately mislead the recipient of the content regarding the identity, motives, actions or connections of the actors involved and/or the popularity of the content. The ultimate aim is to influence public opinion or elections.

Social media: In this study, we focus in particular on the social media platforms operated by Google (YouTube), Meta (Facebook and Instagram), Microsoft (LinkedIn), Snap (Snapchat), TikTok and X (formerly Twitter), as they have the largest user bases in the Netherlands.

1 Introduction

Citizens' freedom to form their own opinions is a prerequisite for democracy. Social media platforms are playing an increasingly important role in shaping public opinion on political and social issues. They are spaces where political parties campaign and engage directly with citizens. Furthermore, citizens engage in political discussion and form their own judgements. The content circulating on social media platforms is therefore part of the public debate. Recommendation algorithms on social media platforms play a key role in determining which posts people see online. In this context, the WRR refers to the opinion power of platforms: social media platforms, and more specifically their recommendation algorithms, are increasingly acting as gatekeepers in the provision of information to a growing number of Dutch people (Scientific Council for Government Policy (WRR), 2024).

The role of social media platforms in public debate also entails risks. They are not only venues for the dissemination of political information, the exchange of information and campaigning, but also channels through which foreign actors may attempt to influence public opinion. The European Union has previously identified instances where foreign actors attempted to influence elections via online channels (European External Action Service, 2023).

In 2024, a round of elections in Romania was declared invalid after traffic on social media platforms were found to have been potentially manipulated to favour a particular candidate. In the Netherlands, too, there are concerns about foreign interference in elections (AIVD, 2025; *Parliamentary Papers II, 36 552, no. 12*, 2025). At the time of the 2025 parliamentary elections, various covert and manipulative activities were observed on social media platforms, leading to growing concerns about interference (HEIO Consortium et al., 2026).

At the request of the Standing Committee on Digital Affairs, the Rathenau Instituut investigated how recommendation algorithms on social media platforms – which play a significant role in determining which posts we see online – can contribute to covert foreign elections interference. The findings of this research are set out in this report.

The central question in this study is:
What role do recommendation algorithms on major social media platforms play in election interference?

This main question has been addressed through sub-questions, which are discussed in successive chapters of this report.

The sub-questions are:
1. What is interference via social media platforms in the context of elections? (Chapter 2)
2. How do the recommendation algorithms of major social media platforms work? (Chapter 3)
3. How can recommendation algorithms be used to interfere in elections? (Chapter 4)
4. What efforts are already being made to prevent interference via recommendation algorithms? (Chapter 5)
5. What options are available to further prevent and/or address interference via recommendation algorithms? (Chapter 6)

To answer the main and sub-questions, we conducted a literature review, interviews, an expert session and a review process. An overview of the parties consulted can be found in Appendix 1. A detailed description of the methodology is provided in Appendix 2. Appendix 3 contains additional information on legislation and policy.

1.1 Scope

This study does not focus on the actual impact of interference on election results, but on the opportunities that recommendation algorithms on social media offer for deliberately misleading the online public debate in an attempt to influence public opinion and elections. In doing so, we are addressing a growing social and political concern: how do we protect free elections and, by extension, democracy in the digital age?

In this report, we describe the workings of recommendation algorithms in detail and depth. We also provide examples of how recommendation algorithms operate on major online platforms today, based on scientific research conducted in collaboration with the platforms or with their users. A systematic comparison of the differences between recommendation algorithms on major social media platforms is not part of this study. A different type of research is required to gain insight into the differences between platforms.

This study focuses on the role of recommendation algorithms in interference. It is important to realise that there are also avenues for interference in which the recommendation algorithm plays no part. Consider, for example, threatening a politician via private messages, infiltration, or influencing other media such as newspapers or television programmes. One might also consider recent developments surrounding chatbots. A growing number of people appear to be using chatbots as a voting aid. It is possible to influence the output of chatbots, but this falls outside the scope of this research.

1.2 Reading Guide

This report answers the main question by addressing the five sub-questions listed above. Each chapter examines one of these sub-questions, after which we answer the main question in the conclusion.

In Chapter 2, we examine the concept of interference in general and in the context of social media in particular. We also discuss the potential consequences of interference.

In Chapter 3, we discuss in detail what recommendation algorithms on social media are, what it means that they are now largely based on engagement, and what the implications of this are for the dissemination of content on social media.

In Chapter 4, we examine the tools that, thanks to recommendation algorithms on social media, are available to malicious (foreign) actors.

In Chapter 5, we outline the efforts currently underway to combat interference via social media. We provide an overview of existing legislation and regulations, describe the measures taken by platforms to combat interference, and address the concerns that experts still have, despite these efforts, regarding the potential for interference in our elections.

In Chapter 6, we answer our central research question and describe the policy options identified by the research to further limit interference via social media.

2 A Closer Look at Interference

What is interference? What does interference look like on social media platforms in the context of elections? How is interference identified, and what are its potential consequences? These are the questions we address in this chapter.

We will first examine what foreign interference entails. As there is no widely accepted academic definition, we will use a description that is commonly employed in European practice. We will then briefly explain why interference is problematic from the perspective of international law.

We will then discuss how interference manifests itself on social media and introduce a definition for this type of interference. We will explain why defining this type of activity is a sensitive issue and illustrate what this type of interference looks like using a case study: the 2024 Romanian presidential elections.

We go on to explain why it is difficult for various (public) actors other than platforms to identify online interference. We illustrate the importance of understanding how information circulates on online platforms using the case study of fake accounts from abroad on X surrounding the recent general election.

The process in the Netherlands for identifying interference appears to have a number of weaknesses. We discuss these in the following section. Finally, we address concerns about attempts at interference via social media during elections.

2.1 What is interference?

There is no universally accepted scientific definition of interference (Berzina & Soula, 2020; European Commission, 2023). Policymakers likewise face difficulties in defining interference, as any definition must capture its diverse and evolving forms while respecting citizens' political rights and freedoms. (see also 2.2) (Berzina & Soula, 2020).

A definition of foreign interference widely used in European practice is: 'a type of foreign influence activity carried out by or on behalf of a foreign actor at state level that does not fall within the scope of diplomatic relations, and which is distinguished from such relations by the fact that it operates in a coercive, covert, deceptive or corrupting manner and acts in contravention of the sovereignty, values and

interests of the European Union' (Directorate-General for Research and Innovation, 2022, p. 2).

Examples of such activities include putting pressure on political representatives, providing (covert) financial support, putting pressure on individuals in strategic positions, carrying out hybrid attacks, or spreading disinformation (Directorate-General for Research and Innovation, 2022).

Interference in the context of elections means that a state undertakes activities to influence an election in the Netherlands by favouring or disadvantaging a party participating in that election (Ohlin & Hollis, 2021).

If the perpetrator of such interference in elections is a state, or is supported by a state, this may constitute a breach of the principle of non-intervention under international law and Article 2(4) of the Charter of the United Nations (1945; McWhinney, 1966).

Article 2(4) stipulates that UN members must respect each other's territorial integrity and political independence. It prohibits the use of force or any other means to undermine the independence and integrity of states. It is therefore based on the principle that states are sovereign and must not interfere in internal political affairs. In the context of elections, according to the American legal scholar Ohlin, self-determination is the most useful concept: the right of a people to decide their own political destiny (Ohlin, 2017).

2.2 What is social media intervention?

In this study, we focus on a relatively new *type* of interference activity that is becoming increasingly accessible whilst at the same time becoming increasingly sophisticated (Justice for Prosperity, 2025; Ohlin & Hollis, 2021): interference via the recommendation algorithms of social media platforms.

By interference via social media platforms we mean:
The orchestrated use of multiple accounts, pages or platforms by foreign state actors and/or their representatives to increase or decrease the visibility of certain content. They do this by exploiting the recommendation algorithms of social media platforms. The actors involved attempt to deliberately mislead the recipient of the content regarding the identity, motives, actions or connections of the actors involved and/or the popularity of the content. The ultimate aim is to influence public opinion or elections.

This definition is largely inspired by the definition put forward by political scientists Thiele and colleagues.(Thiele et al., 2025)[1] By deliberately misleading, we mean that the interfering party deliberately attempts to mislead the recipient about the origin of the content they are shown, by making themselves as untraceable as possible online and concealing other critical aspects of the operation.

It is also important to note that the actor in question attempts to present a misleading picture of how popular certain content is online. The actor does this by increasing or decreasing visibility by manipulating recommendation algorithms using the tools described in Chapter 4. It should be noted, however, that we cannot know how popular the content would have been had the actor not carried out this activity (see Chapter 3 on relative amplification).

The element of *deliberate deception* is a necessary prerequisite for describing interference via social media. It is this element that distinguishes an interference operation from a typical political campaign. After all, political parties also wish to increase their online visibility in an organised manner and will, in doing so, have to rely increasingly on social media's recommendation algorithms (see Chapter 3).

A clear definition is important, because classifying – or failing to classify – a political campaign as interference in the run-up to the elections can have far-reaching consequences. Fundamental rights such as freedom of expression may be violated, ultimately leading to the (undesirable) effect of increasing mistrust in institutions. The ability to classify a campaign as interference can be a powerful tool that may ultimately be abused by anti-democratic forces.

Thiele and colleagues' definition refers to deliberate misleading activities on social media, regardless of who carries them out. For the purposes of our research, we

[1] 'The coordinated activity of multiple accounts or pages on social media platforms aimed at influencing public opinion or public debate by increasing or decreasing the visibility of specific content or actors using the capabilities of these platforms, and which deliberately misleads the public regarding the identity, motives, actions or connections of the actors involved and/or about the popularity of the content disseminated (own translation, (Thiele et al., 2025, p. 4).'

have added to their definition that deliberate misleading takes place in the context of elections and is carried out on behalf of or by foreign state actors.

Experts disagree on whether research into deliberate deception should focus on the role of foreign actors. Firstly, because detection is difficult, and secondly because concealing their operations is inherent to the nature of interference activities. Moreover, the fact that deliberate deception cannot always be attributed to a specific foreign actor does not make it any less problematic. In this study, however, the decision was taken to focus on foreign state actors.

An example of interference via social media platforms can be found in the case study below concerning the 2024 Romanian presidential elections.

Box1 Case study of interference via social media: Romanian elections

In November 2024, Calin Georgescu won the first round of the Romanian presidential elections. This came as a surprise to many. The candidate had no campaign budget and was barely visible on television during the election campaign. A sudden spike in his online visibility, particularly on the social media platform TikTok in the run-up to the elections, suggested that he had mounted a particularly successful social media campaign. Whether the campaign led to a different election result is difficult to determine.

One of the key findings, according to the Romanian security services, and later also from several NGOs such as EDMO, Digihumanism, Expert Forum and DFRlab, as well as the French security services, was the deployment of a coordinated campaign aimed at increasing Georgescu's visibility on platforms including TikTok. Georgescu's TikTok account gained 2,500% more followers during the election month.

The campaign involved instructing thousands of Romanians via Telegram channels to spread pro-Georgescu messages. Another tactic was the mass posting of pro-Georgescu hashtags under messages that were already popular, including posts by rival candidate Lasconi. Bots were likely used for this as well. Influencers were also recruited to promote Georgescu. Influencers created videos that appeared neutral, containing indirect references to the candidate, without stating that the content was political. The videos were inundated with comments from pro-Georgescu accounts. These may also have included fake accounts.

Georgescu denies being behind the campaign, although there are indications to the contrary. In their released documentation, the Romanian security services state that a foreign actor is behind it and further assert that Russia has a long history of interfering in electoral processes in other countries.

Some experts draw parallels between the influence campaign for Georgescu and that for Stoianoglo, a Moldovan presidential candidate who also received support via paid Romanian influencers. Furthermore, according to Romanian investigative journalists, Russia has at least indirectly contributed to the success of pro-Russian candidates through disinformation campaigns run by the Russia-linked marketing agency AdNow in the run-up to the elections.

See also the article on the Rathenau Instituut website (in Dutch): Why is Romania an important example of political interference?

2.3 Possible consequences of interference

There are various concerns regarding the potential consequences of interference in the conduct of elections. Below, we provide an overview of the concerns mentioned in the literature consulted:

1. Pitting people against one another to exacerbate divisions in society. (section 2.3.1)
2. Altering the election results. (section 2.3.2)
3. Sowing doubt about the democratic process. (section 2.3.3)

2.3.1 Pitting people against one another

Does intervention result in existing social divisions being exacerbated and people becoming radicalised? This is a cause for concern among academics (Vliegenthart et al., 2024). Social media may potentially cause affective polarisation: people are less positive towards those with differing views (Miltenburg et al., 2022).

In Chapter 4, we discuss various ways in which malicious foreign actors ensure that content which plays on emotions and pushes the boundaries of what is permitted on platforms, as well as extreme content posted by other users on a platform, gains greater visibility than other content. The recommendation algorithm can act as a

flywheel, making this type of content more visible (we explore this in detail in Chapter 3).

As a result, people viewing content on the platform may get a distorted picture of the average sentiment on the platform.

Amplifying specific content can also have an impact on the public debate taking place outside these platforms. For example, when mainstream media report on content found on social media, that content receives renewed attention. The specific concern here is that when extreme views that gain significant reach on social media are repeated in mainstream media, they are perceived as more mainstream. This can lead to a shift in the norm towards more extreme viewpoints. The fear is that this shift in the norm will influence how politicians express themselves. After all, online culture can be mirrored in the House of Representatives. (Data School, Utrecht University, 2023)

2.3.2 Changing the voting result

One of the concerns regarding interference is the manipulation of election results. In recent years, there have been particular concerns about the spread of misinformation and disinformation via social media. The 2016 US presidential election is often cited as an example. During these elections, it was observed that some of the political advertisements originated from Russia and were targeted at voters in swing states, with the aim of getting more people to vote for the Republican candidate Donald Trump. Incidentally, various researchers have pointed to the limited and selective reach of the messages disseminated (in particular, they reached Republicans who would have voted for Trump anyway) and the focus during the elections on domestic news coverage. Consequently, it is assumed that interference via social media had no effect on election results(Honingh & Van Ham, 2024, p. 156). Nevertheless, these concerns remain because future disinformation campaigns could turn out differently (Ecker et al., 2024). The annulled elections in Romania are an example of this. These were annulled because it was suspected that the results had been altered by interference via social media (see Box 1 on the Romanian elections).

Because of its multi-party system, it is often thought that the Netherlands is less vulnerable. The kind of political intervention seen in the United States, where an attempt is made to make one party the largest, would not work in the Netherlands. In the Netherlands, there are many more parties competing to become the largest. Furthermore, the winning party in the Netherlands must share power within a coalition. As a result, the winner has relatively less power than in the United States.

It was also assumed that the Dutch-speaking world offers some protection against interference, as fewer people worldwide speak the language and there are therefore fewer people capable of carrying out interference activities (Irwin & van Holsteyn, 2021). However, with current translation software, it is entirely possible for foreign actors to create Dutch-language content. It has also become easier to interpret existing content and select which content is suitable for amplification.

Moreover, it might actually be worthwhile to bring about a shift in the distribution of seats in Dutch politics. If, for example, a smaller party is supported, its number of seats could grow relatively quickly, say from one to three. This would result in relatively more speaking time in the House of Representatives than if a large party were to gain two additional seats.

2.3.3 Sowing doubt about the democratic process

Even if interference were to persuade few voters to change their vote, there are concerns that people would lose confidence in the fairness of the election process. Figures on Dutch people's confidence in the electoral process do not currently suggest this (Garnett et al., 2025). Nevertheless, concerns are being raised. Interference that casts doubt on the fairness of an election can undermine the democratic process in general (Honingh & Van Ham, 2024).

2.4 Inadequate control over interference

To identify potential interference, election task forces (in Dutch: verkiezingstafels) have been set up in the Netherlands ahead of the elections. These task forces comprise ministries, security services, the National Cyber Security Centre (NCSC), the Netherlands Authority for Consumers and Markets (ACM), local authorities and the Electoral Council. They identify risks and take measures. The security services play an important role in detecting behaviour by state actors insofar as it threatens national security (*Parliamentary Papers II, 36 552, no. 12*, 2025).

Where the circulation of content that poses a threat to elections is concerned, the Ministry of the Interior and Kingdom Relations (BZK) can now invoke its so-called election flagger status. If a platform receives a report from an election flagger, it must assess, as a matter of priority, whether a post could pose a risk to the integrity of an electoral process. Incidentally, the Ministry of BZK does not report this on its own initiative, but following media reports or reports from independent fact-checkers or other government bodies.

For example, on 17 October 2025, the Ministry of the Interior and Kingdom Relations reported to X the circulation of incorrect online instructions for voters who wanted to vote for the political party GreenLeft–Labour (GroenLinks-PvdA) to vote for two candidates. X subsequently stated that the content was in line with their terms of service and left the content in place. The Ministry of the Interior and Kingdom Relations also brought the post to Meta's attention. This took place on 20 October 2025. Meta removed the post on 25 October 2025 (*Parliamentary Papers II, 35 165, no. 102*, 2026).

As the circulation of information becomes increasingly concentrated on a handful of platforms where billions of citizens converge (Rathenau Instituut, 2025c), these platforms have the clearest view of what is happening online. They are best placed to know which content reaches whom, who it originates from, and who is paying for it.

That information is not readily available to everyone. Researchers have only a general understanding of how algorithms work. What users see online and the role that recommendation algorithms play in this remains largely unknown (Cooper & Chapman, 2025).

For journalists, academics, NGOs and institutional bodies such as the Ministry of the Interior and Kingdom Relations and the Electoral Council (in Dutch Kiesraad), it is not easy to determine whether interference activities are taking place via social media platforms. However, access to reliable and verifiable information is essential for such parties to fulfil their roles effectively. A determination of election interference without comprehensive and transparent substantiation may actually increase mistrust within a society. The Digital Services Act (DSA) aims to improve oversight in this area for supervisory bodies and designated investigators. We discuss whether, and to what extent, this is successful in Chapter 5.

Incidentally, there is an institutional vulnerability that could be relevant in the context of interference. The determination of the election result and the decision on whether to hold a re-vote are, in fact, made by the sitting House of Representatives.[2] However, the sitting House of Representatives has a direct interest in the decision as to whether an election should be repeated. Once the new representatives have been admitted to the House of Representatives or the municipal council, the result is irrevocable. There is no possibility of judicial appeal.

2 The House of Representatives Committee for the Examination of Credentials is responsible for submitting a report to the House of Representatives on the conduct of the election, the confirmation of the results and the admission of members. The House is not formally bound by the committee's findings, but in practice it does adopt its conclusions (Trapman, 2024b).

Furthermore, there is no provision specifying in which cases the House of Representatives may decide to hold a re-vote (Council of State, 2024; Trapman, 2024a). Case law from the European Court of Human Rights, an opinion from the Electoral Council (in Dutch Kiesraad), and an opinion from the Council of State highlight problems with the current way in which our electoral system resolves electoral disputes (*Mugemangango v. Belgium*, 2022; Council of State, 2024).

Box2 Fake accounts from abroad on X during Dutch elections

In the Netherlands, journalistic investigation from RTL Nieuws reported around the 2025 parliamentary elections on orchestrated activities by multiple accounts that were deliberately misleading. Journalists investigated over 550 accounts that interfered with political messaging before and after the House of Representatives elections. The 550 accounts sent a total of over 23,000 Tweets. The majority of these were Retweets from accounts belonging to highly active and outspoken Dutch citizens.

The accounts used fake names and the names of well-known Dutch people. The accounts posed as Dutch users by using Dutch account descriptions and by sharing posts from Dutch political figures. They were mainly operated from countries in West Africa. At least 225 accounts were found to be managed from Nigeria.

Because the fake accounts began sharing Tweets from smaller accounts, these appeared to be more popular than they would have been had the campaign not been active. As a result, these posts may have been shown to more X users than would have been the case without the fake accounts' activities (only the platform can say how many users actually saw the Tweets).

Three forms of interference can be observed in this interference campaign. Firstly, the amplification of existing divisions. Most of the interactions from the fake accounts were Retweets on topics that are the subject of political debate in the Netherlands, such as migration and the situation in Gaza. Secondly, the fake accounts amplified a specific political perspective. They did this by retweeting right-wing politicians and opinion leaders. The aim of this is presumably to give these accounts a wider reach. Political parties Forum for Democracy and PVV leader Geert Wilders were retweeted most frequently. By comparison, the CDA and VVD each received a single Retweet from the network. Thirdly, content was disseminated to sow doubt about the democratic

---

process. The post that received the most Retweets from the fake accounts was one suggesting electoral fraud.

Two experts stated in the RTL Nieuws report that they suspect Russia is involved. The manner in which the interference activities were orchestrated bore similarities to an interference campaign described by CNN that was organised in 2020 from Ghana and Nigeria. (Clarissa Ward et al., 2020). *(Foreign troll armies amplified political and inflammatory messages surrounding the elections, 2025)*

2.5 Conclusion

In this chapter, we have introduced a general definition of interference. Interference is a foreign influence activity carried out by or on behalf of a foreign actor at the state level that does not fall within the scope of diplomatic relations, and is distinguished from such relations by the fact that it is coercive, covert, deceptive or corrupting in nature and acts in contravention of a state's sovereignty, values and interests. Foreign interference in the context of elections means that a foreign state undertakes activities to influence an election in the Netherlands by favouring or disadvantaging a party participating in that election.

If the perpetrator of such interference in elections is a state, or is supported by a state, this may constitute a breach of the principle of non-intervention under international law and Article 2(4) of the UN Charter.

By 'interference via social media platforms', we mean the orchestrated use of multiple accounts, pages or platforms by foreign state actors and/or their representatives to increase or decrease the visibility of certain content. They do this by exploiting the recommendation algorithms of social media platforms. In doing so, the actors involved attempt to deliberately mislead the recipient of the content regarding the identity, motives, actions or connections of the actors involved and/or the popularity of the content. The ultimate aim is to influence public opinion or elections.

Just like the general definition of interference, the definition of interference on social media presents challenges. This is because some elements are difficult to prove

and because misclassification can have a significant impact on freedom of expression.

The consequences of interference can include pitting people against one another by amplifying existing divisions in society on social media platforms, giving extreme views a wider reach, influencing election results, or fuelling mistrust in the democratic process. In the next chapter, we examine the role played by social media platforms' recommendation algorithms in interference campaigns.

Journalists, policymakers, researchers and members of the public are losing sight of the information circulating on social media platforms, whilst the flow of information is becoming increasingly concentrated on these platforms. This makes it more difficult to independently identify interference and have it scrutinised by other parties. There is also a vulnerability in the fact that decisions regarding the contesting of the result and the possible organisation of a re-vote may be politically biased.

3 How do recommendation algorithms work?

In this chapter, we address the question: how do recommendation algorithms on major social media platforms work? Since around 2016, we have become increasingly accustomed to content being recommended to us on social media without us having actively asked for it. Recommendation algorithms play an increasingly significant role in what we see online. That is why it is important to understand how recommendation algorithms work.

In this chapter, we discuss what engagement-based recommendation algorithms are, and what other types of recommendation algorithms exist.

3.1 What determines what you see online?

There is only limited space on any timeline. Platform designers must therefore decide which posts to display in your timeline. And they make choices about how that content[3] is organised. In its study 'Inclusive Online', the Rathenau Instituut demonstrates how the vision, revenue model and ownership structure of online platforms have a major influence on the design choices behind recommendation algorithms and the way in which the user interface is designed (Rathenau Instituut, 2025a). The choice of an advertising revenue model, for example, facilitates data collection and personalised recommendation algorithms.[4] Decisions about which posts appear on your timeline are therefore not neutral or purely technical design choices.

Social media platforms can rank posts in various ways. To do this, they use algorithms: a set of instructions designed to solve a specific problem, such as determining which content users see from a vast array of options (Rathenau Instituut, 2022). Algorithms have a reputation for being difficult to understand. But not all algorithms are complicated. Ranking posts in reverse chronological order is

3 In the literature on recommendation algorithms, the term 'items' is often used instead of 'content'. This is because items can be broader in scope: user accounts can also be recommended, and this does not fall under the category of content. In this report, we specify when we mean content and when we mean accounts.

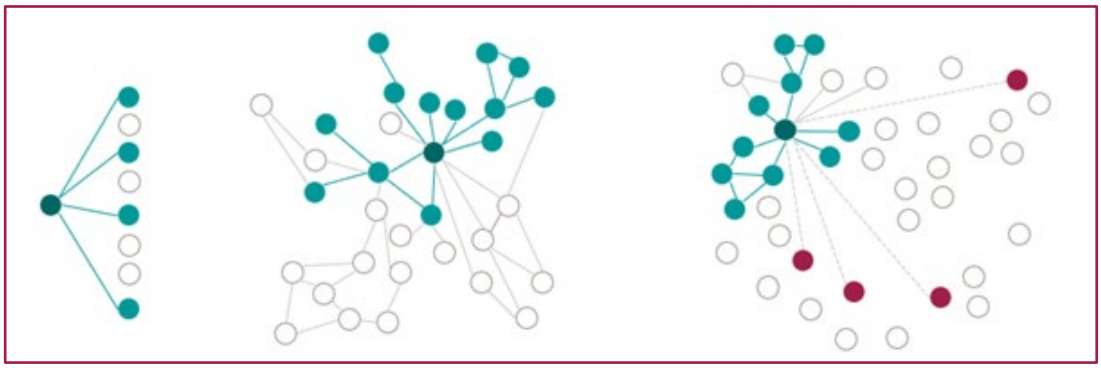
4 The decision to monitor, track and profile users online is closely linked to the dominant current revenue model of online environments: the data revenue model. In combination with displaying personalised advertisements, platforms tend to encourage users to create accounts, download an application to keep users within their own ecosystem, and encourage users to use their real names. This makes users easier to track and profile. For further explanation, see the report *Inclusief online* (Rathenau Instituut, 2025a).

an example of an algorithm that is fairly easy to grasp: display the most recent content from everyone a user follows, and sort this content chronologically, with the newest content at the top.

3.1.1 Ranking based on subscription, network and recommendation algorithms

We distinguish between ranking based on three models: subscription, network and recommendation algorithms (see also the figure below).

Figure1 Ranking of posts



Ranking based on subscription (left), network (centre) and recommendation algorithms (right). This classification is based on research conducted in 2016 by researchers at Stanford University and Microsoft Research (Goel et al., 2016). Figure: Laura Marienus/Rathenau Instituut

In the early days of social media, content was primarily organised based on subscription and network models. You would see posts from accounts you followed yourself, in chronological order, occasionally interspersed with advertisements (left in Figure 1 above). Or you would see content shared by people in your network, for example through Retweets (centre).[5] Nowadays, most platforms combine subscription- and network-based organisation with recommendation algorithms, whereby posts from accounts outside your network are also, or primarily, displayed. This increases the likelihood that a user will see content from people they had never indicated they wished to follow.

Well-known examples of ranking based on recommendation algorithms (right) include the ‘For You’ pages on TikTok, Instagram and X, or ‘Discover’ on Snapchat. Ranking based on subscriptions (who you follow) and what people in your network share still plays a role on these platforms, but to an increasingly limited extent.

5 Currently, European legislation requires large platforms to continue offering this non-personalised version of your timeline. Article 38 of the Digital Services Act (DSA) requires platforms to offer users at least one recommendation system that is not based on profiling.

TikTok has even grown thanks to this model, in which virtually all visibility is generated via recommendation algorithms (Narayanan, 2022).

Recommendation algorithms, as well as dissemination within a network, can contribute to content going viral on social media platforms. The concept of going viral is borrowed from the way we talk about viruses spreading. Going viral online is therefore something different from simply being popular. Virality is about how content spreads via other people within a particular network (Narayanan, 2022). This is reminiscent of the models scientists used during the coronavirus crisis to predict how infections would spread. In the same way, something can be popular on the internet purely because someone has a large following, meaning that something can easily be picked up by many of those people (diffusion through subscription). That is not the same as going viral (Goel et al., 2016).

So, anyone who manages to go viral reaches a larger and, above all, new audience. In Chapter 4, we'll look at the tactics people use to try to influence this mechanism.

3.1.2 Differences between social media platforms

There are differences between major social media platforms in how they determine what you see. These differences can be found in the user interface: what kind of interaction does the social media platform encourage or discourage? Consider, for example, the introduction of frictionless design such as the endless scroll (where, as a user, you can scroll down endlessly and there is no limit to the amount of content you can see), or the ability to comment on public content. But there are also differences in the way content is organised: how significant is the role of content from your own network? And on the basis of what data is it determined what you are presented with?

With recommendation algorithms playing an increasingly prominent role, control over what users see is shifting more and more towards the designers of these algorithms. Not all online environments operate in this way: on the social network Mastodon, for example, you still only see posts from people you follow yourself, or those shared by people in your network. A platform such as Substack, where you can subscribe to newsletters from creators, combines all three forms of ranking (McKenzie & Monga, 2024; Substack, 2025).

Table1 Three ways to sort online

	Method 1: Ranking by subscription	**Method 2: Sorting by network shares**	**Method 3: Ranking by recommendation algorithms**
What do users see?	Only content from channels you follow	Content from accounts you follow, including content from others shared by these accounts	Content that recommendation algorithms predict you will interact with (e.g. watch a video, like a post or forward it to someone else)
What are some well-known examples?	Subscriptions to newsletters, podcasts or RSS feeds	Twitter and Instagram until 2016, Mastodon, the optional 'Following' page on TikTok, Instagram and Facebook	YouTube, LinkedIn, Pinterest, default option on Instagram, X and TikTok
How does content gain a wider reach?	Because people actively indicate that they want to subscribe	Because people share the post within their network	Because recommendation algorithms predict that a post is relevant to you

Table: Rathenau Instituut

3.2 What are recommendation algorithms?

Recommendation algorithms are a set of systems that sort and rank content based on data, with the aim of recommending potentially interesting content to users. These recommendations are based on factors such as the content of the post, your previous behaviour and interactions within your network (Huszár et al., 2022b).

Recommendation algorithms are used, for example, to determine which new series to recommend to you, or which pair of shoes resembles something you have bought before (Ricci et al., 2022). But they are best known for their use on social media platforms. Think about what your timeline looks like on social media platforms: are the latest posts at the top, or the most 'relevant' ones? The provision of 'relevant' content is always based on recommendation algorithms. Recommendation algorithms therefore fulfil a function in ranking and sorting the vast amount of content available on the internet: they are a technical solution to the problem of information overload (Lin et al., 2024).

Please note: within the academic field of machine learning, the combination of models and systems used to recommend content (from data collection to monitoring, filtering, ranking and training) is often referred to as *recommender*

systems. A recommendation algorithm, then, refers to the specific technique used as part of a *recommender system* to make predictions. These are therefore different terms, which are often used interchangeably outside the scientific community. In this study, we use 'recommendation algorithms' as a synonym for 'recommender system'.

Recommendation algorithms can also be partly self-learning, based on feedback. For example, a recommendation algorithm can learn from user interaction and, based on that, suggest different content to a user.

The recommendation algorithm learns independently, but it does have a goal set by humans. If the goal is to increase the time a user spends on a platform, the recommendation algorithm will learn from content that a user spends a lot of time on and recommend this type of content again.

Ultimately, social media platforms recommend content that aligns with the objectives the platform aims to achieve through its recommendation algorithm (KGI Expert Working Group on Recommender Systems, 2025).

The more time people spend on a social media platform, the more advertising revenue they generate for the companies behind these platforms. It therefore pays to optimise recommendation algorithms in such a way that they increase user engagement. Engagement is a set of user behaviours that arise during normal use of the platform and are assumed to be associated with value for the user, the platform or other stakeholders (Stray et al., 2024). In the context of social media, this often refers to: likes, shares, comments, time spent, and other interactions that indicate how valuable or relevant the content is to the user or the platform.

Researchers at the Vector Institute in Canada define a recommendation algorithm as a function *f* that predicts the value of an item *i* (a piece of content) for user *u*, represented by $\hat{r}_{ui}$. The function *f* is often based on historical data. Θ represents the parameters of the model, based on previous data (Raza et al., 2026). See also Figure 2, below.

Figure2 Mathematical representation of a recommendation algorithm

$$\hat{r}_{ui} = f(u, i; \Theta)$$

Source: (Raza et al., 2026).

Recommendation algorithms contribute to this ranking and sorting by predicting the 'value' that content holds for a user. The purpose of this prediction may vary from platform to platform. For an online shop, it concerns the likelihood that you will add a product to your basket or proceed to checkout. On social media, it may concern

the likelihood that you will like or share a post. Recommendation algorithms are trained on large amounts of data for these various purposes.

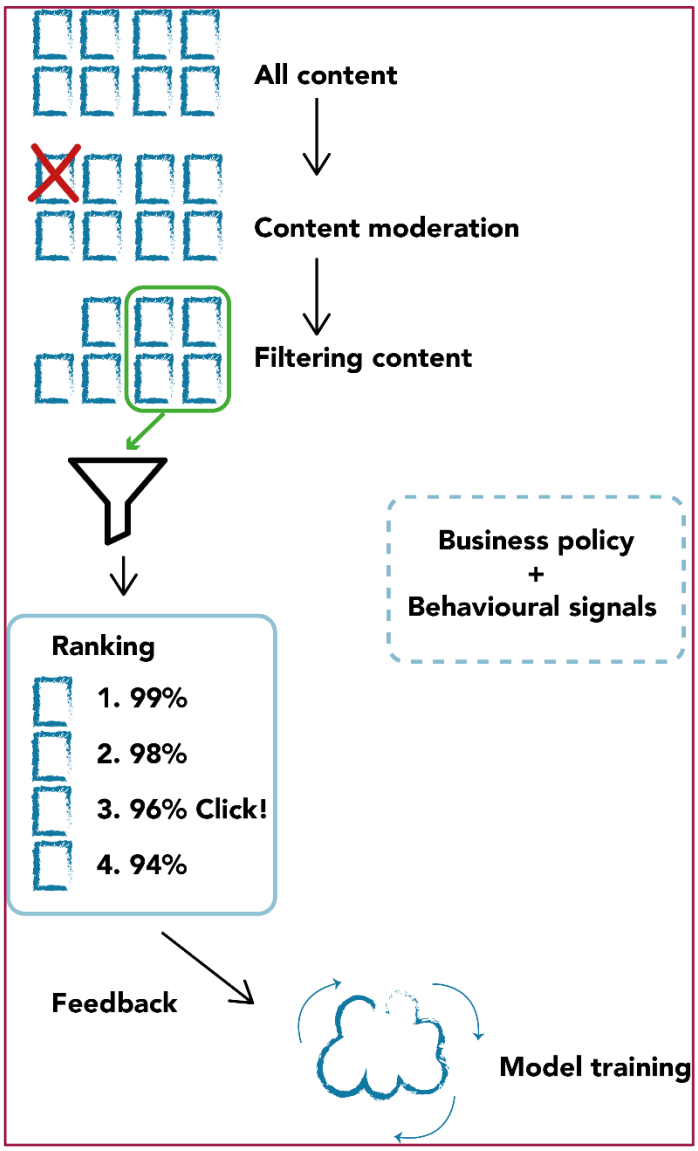
3.3 Five steps to personalised recommendations

Before content is recommended to users, several steps are followed.

We distinguish five key steps that follow one another:
1. Content moderation
2. Training the recommendation algorithm
3. Filtering content
4. Ranking filtered content
5. Learning from feedback

These steps (see also Figure 3 and the explanation below) are a simplification. In reality, recommendation algorithms (often referred to as *'recommender systems'* in this context) differ in their approach and implementation. We have excluded personalised advertisements from this roadmap, as different systems are used for those recommendations. For an explanation of how platforms generate personalised advertisements, see the report *De prijs van gratis internet* (Rathenau Instituut, 2025b).

Figure 3 Roadmap: from all content to recommended content



A step-by-step guide to how platforms transition from 'all content' to 'recommended content'. It shows how different probability scores are used in a user's timeline to determine which content appears in what order. The next step is feedback, which leads to model training. Figure: Rathenau Instituut, based on (Raza et al., 2026; Stray et al., 2024).

3.3.1 Step 1: content moderation

The first step in recommending content is content moderation. Content moderation is the process by which online environments determine the conditions under which content, such as text, photos and videos, is permitted or prohibited (Rathenau Instituut, 2025a). This process involves the assessment of user-generated content on the platform.

In its report *Inclusive Online,* the Rathenau Instituut demonstrates that online environments adopt different philosophies, styles and values when moderating content. (Rathenau Instituut, 2025a) Platforms, for example, opt for greater efficiency, or conversely, quality. They are punitive, or more educational, according to researchers in human-computer interaction in a 2023 academic paper (Jiang et al., 2023).

3.3.2 Step 2: model training

A key step in recommending content is the (ongoing) training of recommendation models. On major social media platforms, this is a process that involves numerous workflows and requires the collaboration of many engineers. Training a recommendation model is based on large amounts of data, such as posts previously uploaded to the platform and users' interactions with those posts. This data is analysed in a training pipeline, where the model is trained on large amounts of data and specific objectives, such as finding similar content. The model's outputs are then validated, and further learning takes place based on these results.

Techniques such as computer vision (the interpretation and analysis of visual data by computers) and natural language processing (the processing and classification of textual data) are used to train recommendation models.

3.3.3 Step 3: filtering content

The third step in content recommendation is to make an initial rough selection of content to display. This is done using a retrieval model, and this step is also referred to as filtering. The aim is to separate the wheat from the chaff and filter out posts that are likely to be ranked highly later on (Stray et al., 2024; TensorFlow, 2023).

We identify three key ways in which recommendation algorithms filter content on social media. The three filtering methods are:
1. Content-based filtering
2. Collaborative filtering
3. Hybrid filtering

Content-based filtering
Content-based recommendation algorithms recommend content that is similar to what a user has liked in the past (Roy & Dutta, 2022). A recommendation algorithm

that suggests films to watch works by identifying similar characteristics in films that a user has previously rated positively. Films with similar characteristics (director, genre, themes, etc.) will then be more likely to be recommended.[6]

This method of recommending content does not rely on interaction between users, but focuses purely on a user's interaction with the content itself. For example, by liking something or rewatching it frequently.

Collaborative filtering
Collaborative filtering works by analysing the interactions of similar users. Your user profile (including, for example, your age, viewing habits, the device you use to log in and other preferences) is compared with that of other users. The recommendation algorithm then suggests content based on the characteristics of other users who are similar to you. For example, if you are interested in content about fashion and similar users also view a lot of content about train travel alongside fashion, you are more likely to be recommended content about train travel. Even if that content appears to have nothing to do with fashion.

Hybrid filtering
A combination of different types of filtering is often used to recommend content. For example, a system that compares users may learn from the results of a system that compares types of content. Alternatively, one system may attempt to improve the results of the other, based on different data. Many contemporary social media platforms use a combination of content-based filtering and collaborative filtering alongside various forms of machine learning, such as reinforcement learning or large language models (Lin et al., 2024; Raza et al., 2026).

Reinforcement learning is a form of machine learning in which a mathematical model attempts to achieve a reward (such as a user liking an item) through trial and error and by learning from feedback. For example, an agent recommends videos to a user with the aim of getting the user to watch the video. Every time a user clicks away from the video, the agent learns: this was not a good recommendation. Over time, the recommendations will better align with the agent's goal: selecting videos that people watch.

Sometimes there is almost no data available about a user's preferences, particularly when someone signs up to a platform for the first time, or when someone logs in without an account. In machine learning, this is often referred to as

6 *Content-based filtering* works on the basis of vectorisation: the mathematical representation of text or images. With simple forms of vectorisation, such as the mathematical model Word2vec developed by Google in 2012, the relationship between words can be calculated. The words *'cat*' and *'dog*' are then mathematically close to each other, just like the words *'Amsterdam'* and *'Netherlands*'. In this way, content can be searched for that resembles what someone has previously liked (Mikolov et al., 2013).

the cold-start problem. Historical data is lacking to present you with ‘similar’ content. Platforms can then use other data to show you content. For example, content that has recently been popular in your country or region.

3.3.4 Step 4: Ranking filtered content

In the previous step, filtering provided a rough selection of content to display. Step 4 (ranking) is all about ranking and organising content on a user’s timeline. This is done based on probability: what is the likelihood that you, as a user, will find the content relevant? Platforms attempt to predict various forms of interaction. This prediction is determined by behavioural signals, such as whether you like something, watch a video to the end, or share it. Behavioural signals can also have a negative effect from which lessons can be learnt. Do you frequently ignore videos that are similar to one another? Then next time, they will receive a lower probability score.

Platforms use various types of data to rank content. On major social media platforms, this ranking is based on engagement, but it can also be based on other signals, such as feedback from user surveys. Recommendation algorithms can also be trained to rank content with objectives other than engagement, such as bringing people closer together.

When ranking content, platforms distinguish between quality scores, behavioural signals, weights and business rules.

Quality scores
Platforms may attempt to rank what they consider to be low-quality content, such as spam and clickbait, lower in their search results. Platforms use various variables to factor quality into their rankings. For example, Google pays people to rate content to indicate whether a search result is of high quality, and uses this information to predict the quality of search results (Google, 2025c). LinkedIn predicts when posts contain ‘knowledge and advice’ and uses that information in its recommendation algorithm (Feifer, 2023).

Platforms may, for example, also choose to assign lower quality scores to content from new accounts if their risk analyses indicate that new accounts frequently share low-quality content. Such a decision could, however, be detrimental to all new users on a platform.

Behavioural signals and other metrics

Users generate a large number of behavioural signals, both consciously and unconsciously. These can be conscious signals (such as clicking ‘like’) as well as unconscious signals (such as how long you linger on a specific photo). Researchers from the Expert Group on Recommender Systems at the Knight Georgetown Institute also highlight the difference between impulsive signals (sharing a link without having clicked on it yourself first) and deliberative signals (writing a long comment) (Knight Georgetown Institute, 2025).

Content creators can also encourage people to generate signals (such as the well-known ‘like and subscribe’) because they expect this to result in their content being recommended to more users.

In Table 2 and Table 3 on the following page, we provide examples of the predictions used to rank content.

Other information can also be used to predict which content to recommend to users, such as whether a post contains a link, or a person’s popularity within a network. For examples of this, see Table 3 on the next page.

Table 2 Predictions and behaviour

Prediction	Behavioural signals influencing the prediction
How likely you are to watch less than three seconds of a Reel*	• How often the post was skipped within two seconds of being opened • How often this reel has been viewed with the sound on • How many times users in Discover have watched at least three seconds of the reel • How often the reel has been skipped • How often users have watched at least three seconds of the reel, anywhere on Instagram
The likelihood of you clicking on one of the large, rectangular Reels in Explore to view it in full screen on the Reels tab	• The time you've spent watching Reels • How many posts and Reels you've viewed in Explore • How many large Reels you've clicked on in Explore to view them in full screen on the Reels tab
How likely you are to scroll to the next Reel	• The number of posts and Reels you've viewed in Explore • Which posts and Reels you've viewed in Explore • How often you scrolled to the next Reel when viewing Reels in Explore

Examples of predictions and behavioural signals that play a role in recommending content on Instagram Reels. *Reels are short videos created by other Instagram users that users can view in various places within the Instagram app. Source: Meta (30 July 2025). Selection by the Rathenau Instituut.

Table 3 Examples of metrics

Category	Metric
Content creator	• Centrality in the network • Popularity of the poster outside the platform
User-poster interaction	• Followers of the poster • Blocking the poster
Content	• Content contains a link • Content is new • Content is qualitative in nature (human-assessed)

Source: Examples of other types of metrics that may play a role in recommendation algorithms. Based on (Cunningham, 2023).

Weights

Platforms decide which weights to assign to certain behavioural signals. This means they choose which behavioural signals to prioritise or de-prioritise for the recommendation algorithm's predictions. Does the platform consider posting a comment to carry more weight than sharing a post? Or that watching a video in full is more important than consistently watching all the videos from a single account?

Content on the platform is assigned a ranking score based on the prediction and the corresponding weight. The formula is as follows: $score = (weight_1 \times prediction_1) + (weight_2 \times prediction_2) + ... + (weight_n \times prediction_n)$

Table4 Predictions and weights on Twitter

Prediction	Weight
The probability that the user will mark this tweet as a favourite	0.5
The probability that the user will retweet the tweet	1.0
The probability that the user will reply to the tweet	13.5
The likelihood that the user will open the author's profile and like or reply to one of their Tweets	12.0
The probability that (for a video tweet) the user will watch at least half of the video	0.05
The probability that the user will reply to a tweet and that the author of the tweet will engage with that reply	75.0
The likelihood that the user will click on the link in this tweet and reply to or like a tweet	11.0
The probability that the user will click on the thread of this tweet and remain there for at least 2 minutes	10.0
The probability that the user will react negatively to the tweet (requesting to 'see less' of the tweet or author, or blocking or muting the author)	-74.0
The probability that the user will click on 'report tweet'	-369.0

All predictions and *weights* that played a role in 2023 in ranking Tweets on Twitter's 'For You' timeline (April 2023). (Twitter team, n.d.)

Business rules

Not only behavioural signals and their weights, but also business rules can influence the ranking. For example, X has determined that 50% of users' personalised timelines consists of content from accounts you do not follow (Twitter team, n.d.). This applies to everyone, regardless of whether you follow 4 or 4,000 people.

Meta has revised its policies on political content several times in recent years (Stepanov & Gupta, 2021). In 2021, it began experimenting with reducing 'political content' in the feeds of users in Canada, Brazil and Indonesia.[7] Based on these experiments, Meta concluded that people wanted to see less political content in their timelines. To achieve this, Meta adjusted the weightings of comments and likes when ranking political content. As a result, people saw less content that they 'did not find valuable', according to Meta.

On 7 January 2025, Meta announced that it would no longer automatically de-prioritise political content on Facebook, Instagram and Threads.[8] Instead, an option was introduced to see less political content. On Facebook, users can choose not to be recommended any political content at all. The examples above demonstrate just how much influence a platform's design choices can have on what people see online.

3.3.5 Step 5: learning from feedback

From statements made by platforms in the past, we also know that they are constantly tweaking which form of engagement carries the most weight. For example, in 2012 YouTube announced that, from then on, total watch time would be weighted more heavily than the number of people clicking on a particular video (Google, 2012). YouTube had noticed that many people were using eye-catching thumbnails to entice viewers to click on a video, whilst the actual content of the video was not always watched by many people. YouTube attempted to resolve this by adjusting its recommendation algorithm so that videos that were actually watched were given greater weight than those that were merely clicked on.

Platforms use A/B tests – controlled online experiments – to test the effect of new design choices on groups of users (Quin et al., 2024). This is done by formulating a hypothesis ('this adjustment to the recommendation algorithm will encourage people to open the app more often') and presenting two groups of users with variant A and variant B. These are, as it were, live experiments that provide platforms with insight into the effect of their design choices on user behaviour and goals that are important to the platform.

7 These tests were subsequently extended to the United States, Costa Rica, Sweden, Spain and Ireland.

8 Meta defines 'political content' broadly: it encompasses content relating to governments, elections and social issues. These tests were subsequently extended to the United States, Costa Rica, Sweden, Spain and Ireland.

Control over what you see
Social media platforms offer various explicit and implicit options for influencing what you see. Depending on how much weight platforms give this in their recommendation algorithm, you can follow or unfollow accounts yourself. Sometimes you can also hide specific posts, give a thumbs-up or respond to a direct question about whether you found a recommendation relevant (Stray et al., 2024). The degree of user control does vary by platform. On Instagram, for example, feedback is actively sought via a pop-up on the screen. Users can use this to indicate what they think of a post. The actual influence (weight) of these mechanisms on the predictions of the recommendation algorithm is unknown.

Research shows that young people in the Netherlands aged between 16 and 26 consider their influence on algorithmic selection to be minimal (Swart, 2021). Many young people say they do not make the effort to actively try to steer the recommendation algorithm towards their preferences.

3.4 Why platforms switched to engagement

Since 2016, many major platforms have switched to ranking content based on predicted engagement. Reportedly, they all did so on the basis of internal research showing that user retention was higher when content was sorted by engagement (Cunningham et al., 2025). User retention is measured both in the short term (how long a user remains active during a session) and in the long term (whether the user returns later).

In 2020, Facebook and Instagram conducted an internal experiment comparing users with a semi-chronological timeline and those with a timeline ranked by engagement. What did they find? On Facebook, people with a semi-chronological feed spent 20% less time on the platform than those who were shown the engagement-ranked feed, and on Instagram, 10% less (Guess et al., 2023). Twitter found similar results: users who, since 2016 (as part of a scientific experiment), had a chronological timeline rather than one using a recommendation algorithm, viewed on average 38% less content per day (Bandy & Lazovich, 2023).

When does a recommendation algorithm actually work? From a machine learning perspective, a recommendation algorithm works when it achieves the goal of showing you ‘relevant’ content. This is measured using behavioural signals, such as whether you click on a recommendation. Even if the content you click on doesn’t make you happy, or if it makes you feel sad or angry, the mathematical objective of the recommendation algorithm has still been achieved. For developers of

recommendation algorithms, it is only an undesirable outcome of the system if the wrong things are recommended.

The creators of recommendation algorithms often assume that behavioural signals such as likes and the amount of time someone spends on a platform are indicators of what people want. Yet psychological research shows that people's impulsive behaviour is not always an indicator of what they want in the long term. Behavioural economists Kleinberg, Mullainatha and Raghavan view this as a mismatch between what platforms think they know about users based on what users do (liking posts, opening the app) rather than what users actually want (such as being happy) (Kleinberg et al., 2024).

Companies invest a great deal of money and time in further developing machine learning techniques to make technically more accurate predictions at the lowest possible cost (in terms of computing power and energy). In a 2019 scientific paper, machine learning researchers demonstrate how their technical improvements to a machine learning model contributed to a 0.37% increase in engagement on YouTube (Yi et al., 2019). On a large scale, even these seemingly small improvements can generate significant revenue. In 2024, Meta earned nearly $160 billion a year, largely from advertising revenue. In 2016, this figure stood at just $26 billion: an increase of 515%.

3.5 What is content amplification?

In debates about the impact of recommendation algorithms on what people see online, amplification plays a prominent role. Sometimes amplification is used as a synonym for dissemination. Sometimes something is referred to as amplification when people are shown content that does not match their preferences. Amplification is therefore not always a clearly defined concept. So how do you determine whether a particular post has been amplified by a platform?

In our research, we adopt the definition put forward by algorithm researchers Luke Thorburn, Jonathan Stray and Priyanjana Bengani:

Relative algorithmic amplification: A change in the distribution of content under a recommendation algorithm, compared to an alternative recommendation algorithm, whilst user behaviour remains constant (Thorburn et al., 2023).

This definition makes it clear that when discussing amplification, one must always refer to amplification *relative* to an alternative situation, and that one must assume that users' behaviour or preferences remain the same. If a platform had been

designed differently, the content would not have been amplified. For this interpretation, a counterfactual is important: an alternative scenario. For example, the amplification of a post by a recommendation algorithm, rather than on a chronological timeline. Another example is the amplification of a post after the platform has adjusted its recommendation algorithm.

Other forms of content ranking, such as the reverse chronological timeline, are also not neutral (see section 3.1.1). In such timelines, for example, content from people who post very frequently and in large quantities automatically gains greater visibility. However, you can attempt to investigate whether certain posts receive greater relative amplification from recommendation algorithms based on engagement, compared to a chronological timeline. Alternatively, you can compare the effect of different business rules (see section 3.3.4) on relative amplification.

The definition we use in this study also reflects the complex relationship between recommendation algorithms and what people are shown: various factors influence one another. Users, in turn, adapt their behaviour to the design choices underlying the recommendation algorithm. This, in turn, alters their interaction with content. And that, in turn, can lead to adjustments to the recommendation algorithms.

3.6 Which content receives greater relative amplification?

Social media platforms rank content based on engagement because it increases user retention. This can lead to benefits for users, such as discovering niche content they would not have sought out themselves. Or, from the sender's perspective, it can result in a wider reach. (Narayanan, 2023). At the same time, this ranking based on engagement may also conflict with public values, such as safety and user autonomy.

There is much debate amongst academics regarding the definitions and effects of so-called echo chambers and filter bubbles. The terms are often poorly defined and therefore difficult to study empirically, argues researcher Alex Bruns in Internet Policy Review (Bruns, 2019). The debate often revolves around the question: do recommendation algorithms interfere with a person's own personal choice of the content they would like to see? And is it therefore possible that, as a result of recommendations, someone might see all sorts of content that they did not initially want to see? Filter bubbles, for example, are said to artificially amplify people's confirmation bias (the unconscious search for information that confirms one's own views) through the way recommendation algorithms work (Pariser, 2012).

Researchers are trying to understand the type of content and the types of people who seem to benefit more from relative amplification than others. For example, they are attempting to measure the ‘fairness’ of recommendation algorithms. They are also trying, based on user data or access to platform data, to answer the question of which content receives more relative amplification than other content.

To investigate which content receives greater relative amplification than other content, and what role recommendation algorithms play in this, researchers use various methods. Commonly used methods include user-provided data, the use of accounts created by the researchers, and research involving access to platform data. Different methods can yield different types of insights.

Data donations enable a detailed analysis of individual users’ media consumption patterns. The use of research accounts helps researchers to study the recommendation algorithm and determine whether recommended content aligns with user signals, such as accounts followed. Researchers point out that it is very difficult to identify causal relationships between recommendation algorithms and user outcomes (such as changes in political preference). (Stray et al., 2024).

3.6.1 Emotional content receives greater relative amplification

Psychological research suggests that content is more likely to go viral if it triggers a particular emotion. This encourages people to share, like or comment on a post (engagement). And that, in turn, causes the recommendation algorithm to spread the post more quickly to other people (Berger & Milkman, 2012). For example, psychology researchers at the University of Cambridge and New York University observed that hostile content about political opponents was significantly more likely to be shared on Facebook and Twitter between 2016 and 2020 (Rathje et al., 2021). In another study, researchers found that Tweets by US senators containing negative emotions and moral outrage were shared or favourited relatively more often on Twitter between 2013 and 2021 (Mercadante et al., 2023).

Scientists suspect that recommendation algorithms effectively capitalise on human bias towards so-called PRIME content, where the acronym stands for (Brady et al., 2023): PRestigious, In-group, Moral, Emotional.

Experiments conducted by scientists involving users of major social media platforms have also revealed which content is amplified by recommendation algorithms. For example, in 2025, American researchers demonstrated how X engagement-based recommendation algorithm amplified Tweets containing negative emotions such as anger, sadness and fear (Milli et al., 2025). The

recommendation algorithm also amplified Tweets that caused users to feel worse about people with different political preferences. Furthermore, people indicated that they did not appreciate the recommended political Tweets (Milli et al., 2025). The researchers therefore suggest that, in certain respects, Twitter's recommendation algorithm did not align with user preferences.

Social media platforms themselves draw attention to borderline content. This is a term used to describe content that falls within the grey area of what is and isn't permitted on the platforms.[9] The term borderline content is not a strictly defined concept, but rather an umbrella term for various types of content.

Platforms have themselves pointed out in the past that borderline content, just like PRIME content, attracts more user attention than non-borderline content and is therefore recommended more frequently by recommendation algorithms (Gillespie, 2022; Macdonald & Vaughan, 2024).

An internal study at Facebook showed that content with the highest reach, in the top 1–2% percentile, was more often rated as 'bad for the world' than as 'good for the world' in user surveys (Haugen, 2022). We also know from research that engagement is often negatively correlated with content quality. This means that content with the highest predicted engagement more often contains content that scores low on quality: spam, clickbait, misleading headlines or misinformation (Cunningham, 2023). Platforms can attempt to mitigate this effect by assigning quality scores to content (see section 3.3.4).

3.6.2 Popular content receives greater relative amplification

Researchers are using scientific experiments to understand what kind of recommendation algorithms are most likely to contribute to the spread of, for example, misinformation. Fernández, Bellogin and Cantador used Twitter data to investigate whether adjusting algorithms affects the potential spread of misinformation (Fernández et al., 2021). They demonstrate that changes to the way the algorithm is configured influence the spread.

Recommendation algorithms that are largely based on *collaborative filtering* – i.e. those that recommend content that other similar users like – suffer from *popularity bias* (Fernández et al., 2021). In other words, content that is already popular or has

9 Four years ago, in a blog post on content moderation, YouTube described *borderline content* as: *'videos that don't quite cross the line of our policies for removal but that we don't necessarily want to recommend to people*'. Meta describes *borderline content* as content that is not prohibited by the terms of service, but comes very close to being so. (Macdonald and Vaughan, 2024, p. 350).

gone viral becomes even more popular. Computer scientists also refer to this as the 'rich-get-richer' effect (Bellogín et al., 2017).

3.6.3 Hyperactive users receive greater relative amplification

Research by Papakyriakopoulos, Serrano and Hegelich shows that recommendation algorithms can be influenced by hyperactive users (Papakyriakopoulos et al., 2020a). They base their research on political discussions in Facebook groups in 2016 and demonstrate how users who share an excessive amount of content receive relatively more engagement per post than other users and also appear to have an excessive influence on recommendations generated by recommendation algorithms. This is because content from hyperactive accounts can appear excessively in the training data of recommendation algorithms, thereby enabling them to influence future recommendations for other users. The authors therefore highlight the inherent risk of ranking using collaborative filtering and neural networks.[10]

3.6.4 Right-wing political content appears to be gaining more relative amplification

Researchers are attempting to investigate whether platforms' recommendation algorithms are biased when suggesting political content to users. Various studies of Platform X (formerly Twitter) have shown that content from right-wing politicians receives greater relative amplification. Scientists Ye, Luceri and Ferrara used 120 accounts created specifically for the study to examine the recommended content on X during the 2024 US presidential election and found that new accounts were relatively more likely to be recommended content from *right-wing* accounts (Ye et al., 2025). They divided the accounts into four groups: a group of accounts that followed no one, a group that followed right-wing media, a group that followed left-wing media, and a group that followed a mix of right-wing and left-wing media. The accounts that followed no one were recommended relatively more right-wing content.

Economist and political scientist Germain Gauthier and his colleagues have also recently concluded that X's recommendation algorithm suggests more conservative content than liberal content. They compared this with the chronological timeline of users who took part in the study (Gauthier et al., 2026).

10 In machine learning, this is referred to as an *unbalanced learning problem*.

The findings are consistent with a slightly older study in which two million Twitter users from seven countries were examined for bias in the relative amplification of political content (Huszár et al., 2022a). Huszár et al. concluded that, in six of the seven countries, mainstream right-wing political content received greater relative amplification through the recommendation algorithm.

In a preprint study (not yet peer-reviewed at the time of writing this report), German researchers from the University of Potsdam analysed more than 500,000 TikTok videos recommended to 78 accounts created specifically for the study during the 2024 German regional elections and the 2025 parliamentary elections (Tjaden et al., 2025). Of the accounts with no political affiliation, those were most likely to be recommended videos form political party AfD.

Paul Bouchaud also conducted research into the relative amplification of MEPs on X. He found that right-wing content was generally recommended more frequently on X, but that when adjusted for users' political preferences, this effect disappeared (Bouchaud, 2024).

3.6.5 Specific users may receive greater relative amplification

Design choices regarding recommendation algorithms partly concern the scores assigned to certain content relative to other posts. The decision to weight posts differently can also be made for individual accounts. Accounts with many followers are then given an advantage over those with fewer followers. Elon Musk is said to have personally had these weights adjusted in 2023 for his own posts. Internal memos revealed that he did this after a post by Joe Biden received more engagement than one of his own. He reportedly had his staff adjust the algorithm so that his posts were visible a thousand times more often than those of other users (Schiffer, 2023).

Users regularly speculate on social media about whether their content is treated differently from that of others. They accuse the platform, for example, of shadowbanning: making content less visible on social media without removing it outright. The platform keeps the content available, but no longer displays it in people's timelines. In a lawsuit brought by privacy activist Danny Mekić against X, the Amsterdam court ruled in his favour in 2024: X should not have omitted his posts from other people's timelines without explaining that decision to him (*10767307 CV FORM 23-13934*, 2024).

3.7 Alternatives to engagement

Adapting engagement-based recommendation algorithms to mitigate harmful effects may conflict with other business interests, such as maximising engagement and user retention. Nevertheless, scientific research and NGO reports put forward various proposals for adapting engagement-based recommendation algorithms to limit negative effects. For example, researchers propose bridging-based ranking rather than ranking based on engagement (Lasser & Poechhacker, 2025). This is a form of ranking based on building bridges between people. Instead of incorporating user signals relating to engagement into recommendations, internet researcher Aviv Ovadya suggests recommending content to people who are, in fact, different from one another (Ovadya, 2022). Content that bridges the gap between groups of people is then given greater relative amplification by recommendation algorithms.

Recent research on X shows that adjustments to ranking methods can also have a positive impact within engagement-based recommendation algorithms. In 2025, researchers from Stanford University and other American universities demonstrated that changing the ranking of Tweets – by moving divisive Tweets further down in a user's timeline – led to milder feelings towards political opponents (Piccardi et al., 2025). This intervention therefore worked without removing any content, but purely by changing the order.

The studies do note, however, that this intervention led to a reduction (of five minutes) in the amount of time people spent on X. This would therefore run counter to the platforms' business model, which aims to maximise the time users spend on the platform in order to generate the highest possible advertising revenue.

3.8 Neutral recommendation algorithms do not exist

In the previous sections, we have shown just how many design choices lie behind recommendation algorithms. Even the choice of a very simple sorting and ranking system – the chronological feed – is not a neutral design choice. Chronological feeds that sort all the content you see based on (1) the accounts you follow and (2) rank it by time (newest to oldest) result in highly active accounts being overrepresented. Do you follow a lot of news outlets that post several times an hour, but also a number of friends who only share something once a week? There's a good chance that your friends' content will only appear sporadically in your timeline. You could solve this by giving friends' content a different weighting to that of news outlets, but that too is a design choice.

So you can't really speak of a standard or neutral algorithm and a manipulated version. Of course, it is possible for platforms – and to a lesser extent for researchers – to use A/B testing to examine the impact of certain choices on what people see. That is why it is important to reveal the design choices behind the ways in which people view content online. A chronological timeline is, in any case, easier for users to understand. This can increase their control over what they see.

3.9 Conclusion

As there is limited space on every timeline, platform designers have to make choices about which posts to display on your timeline. They also decide how that content is organised. This may be based on who you follow, or what people in your network are sharing, but also on recommendation algorithms.

Recommendation algorithms are a set of systems that sort and rank content based on data, with the aim of suggesting potentially interesting content to users. Recommendation algorithms sort and rank content based on predictions. Nowadays, these are often predictions based on engagement. For example: what is the likelihood that you will like this post, watch this video or forward this message to a friend? Recommendation algorithms are a solution to the problem of information overload on the internet: from all the available content, they attempt to show you something you will find enjoyable, interesting or otherwise relevant.

Platforms make many different choices regarding how they rank filtered content. Even the choice of a very simple sorting and ranking system, the chronological feed, is not a neutral design choice. With recommendation algorithms based on engagement, they make choices about which behavioural signals to include in their predictions, and what weight to assign to those signals. The way in which they carry out content moderation and how they determine the quality of certain content also falls under the choices that platforms make. Platforms also constantly test the functioning and effectiveness of their recommendation algorithms. They do this, amongst other things, to see how they can increase user engagement.

The shift towards recommendation algorithms based on engagement was driven by the fact that platforms discovered that this would help them increase user retention.

Recommendation algorithms can amplify certain content relatively: that is, give it greater visibility compared to other ranking methods. Recommendation algorithms based on engagement can lead to greater relative amplification of, amongst other things, emotional content, popular content, hyperactive users and right-wing political content. Recommendation algorithms could also operate on the basis of

signals other than engagement, and, for example, reward content that builds bridges between groups of people.

If you understand how recommendation algorithms work on major social media platforms, it is possible to capitalise on this and thereby attempt to give certain content a wider reach. In the next chapter, we will explore the opportunities available to influencers to capitalise on this.

4 Tools for interference activities

In Chapter 3, we described how the recommendation algorithm works. In Chapter 4, we shift our focus to foreign actors. How do they exploit recommendation algorithms to deliberately mislead and influence the online public debate from abroad?

Below, we describe the tools used to make certain content more or less visible. These tools are accessible to everyone. It is also possible to commission interference activities from abroad (see sections 4.2.4 and 4.2.5). Incidentally, the motive of those using these tools is not always aimed at influencing elections. Deliberate deception from abroad can also be a means of making money.[11] Social movements and activist groups also make use of these tools because their aim is to make content more or less visible (Treré & Bonini, 2024).[12]

The landscape is constantly changing. Platform companies make daily adjustments to their platforms and launch new services in the process. Malicious (foreign) actors can capitalise on this. By introducing new services, platforms also create new ways of influencing the online public debate. Examples include new services such as live streaming and the ability to monetise a social media account.

Platforms also adjust the options available on their platforms. For example, by changing the rules for content moderation. A well-known example of such an adjustment is Mark Zuckerberg's statement shortly after the inauguration of US President Donald Trump, announcing that content moderation rules relating to migration and gender were being changed to better align with 'mainstream discourse' (Hendrix, 2025). This adjustment to the content moderation rules was shared publicly. However, how platforms tighten or relax content moderation rules internally is often not disclosed in detail because platforms want to prevent malicious actors from knowing how to circumvent moderation rules (Jiang et al., 2023) .

Platforms are also adapting to comply with legislation. How these changes play out in practice depends on how users interact with the options platforms offer. Instagram, for example, allows users to indicate that uploaded content has been

11 The online news outlet 404 Media reported in 2025 that several accounts posing as US citizens and interfering in the online political debate in the United States were likely operating from other countries to make money. The news outlet referred to YouTube videos in which people explain how to make money from fake accounts. (Koebler, 2025).

12 The researchers provide an overview of networks of activists and social movements that post in a coordinated manner with the aim of making certain content more or less visible through the recommendation algorithm.

generated by AI. However, this is by no means done for all AI-generated content. Research by CampAItracker revealed, for instance, that Dutch political parties have disseminated a great deal of content without disclosing that it consisted of AI-generated posts (Simon Kruschinski & Fabio Votta, 2025).

Within these dynamic platform environments, malicious (foreign) actors employ numerous tactics simultaneously. An illustrative example can be found on the website Like.vn, which gained notoriety for the large number of likes it awarded to content from Frans Timmermans' account (see section 4.2.5). The website offers amplification services, accompanied by a note stating that it is advisable to 'order a combined engagement package, including post likes, account followers and comments on request'. According to the provider, the best results are achieved when this 'engagement boosting' is combined with 'natural engagement signals' (like.vn, 2026).

In this dynamic environment, independent researchers such as investigative journalists, academics and NGOs provide an insight into what is happening on social media platforms using publicly available sources. One example is the DISARM initiative, which offers a comprehensive, open-source description of methods of interference, complete with real-world examples (DISARM Foundation, n.d.).

Furthermore, studies conducted during the 2025 general election identified various interference activities in the Netherlands (HEIO Consortium et al., 2026; Jasper Bunskoek et al., 2025; Justice for Prosperity, 2025).

The interference activities described below deserve the attention of politicians and policymakers who wish to understand the potential for interference and mitigate its harmful consequences. Section 4.1 discusses the types of content. Section 4.2 covers the various techniques. Section 4.3 forms the conclusion of this chapter.

4.1 Content selection

People who actively participate in the political debate on social media platforms make a choice about what content they share, and which content from other users they wish to amplify. Malicious (foreign) actors must therefore first and foremost decide what content they wish to disseminate. We distinguish between three types of content within the public debate on social media platforms: legitimate content (see section 4.1.1), borderline content (4.1.2) and illegitimate content (4.1.3).

An example of legitimate content is a post about who someone is voting for or about political and social issues. This also includes politicians or political parties campaigning. Borderline content is content that borders on what is not permitted on social media platforms. This content is known to attract more attention, which means that recommendation algorithms based on engagement amplify this content relatively more than other content (see also section 3.6).

Then there is content that contravenes the law: unlawful and criminal (or illegal) content. Unlawful content is information posted online by individuals that contravenes the law, either because of its harmful consequences and/or because it seriously infringes upon the interests of others (Hoboken et al., 2020, p. 11). Unlawful is a term from civil law and relates to disputes between citizens and businesses, or to disputes in which the government acts in a commercial capacity. Criminal content refers to something that is not permitted under the provisions of the Dutch Criminal Code, such as serious threats involving a criminal offence such as murder or grievous bodily harm.

Malicious (foreign) actors may choose to create one of the three types of content themselves or to amplify existing content from other users.

4.1.1 Lawful content

Political debate takes place on social media platforms. Users share content about politicians or political ideas. In the run-up to the Romanian elections, for example, pro-Georgescu content was shared: footage of the presidential candidate’s television appearances.

Interference activities may involve participating in this political debate by disseminating messages whose content does not contravene community guidelines or legislation and regulations. We refer to interference when a foreign actor on poses as a Dutch citizen and disseminates legitimate content. In this way, the actor deliberately attempts to mislead (see also section 4.2.3 Fake accounts).

4.1.2 Borderline content

Borderline is a term used to describe content that falls within the grey area of what social media companies do and do not permit on their platforms.[13] The term borderline is not a strictly defined concept, but an umbrella term for various types of content. Examples of borderline content include: misinformation and disinformation, sexually suggestive content, graphic images, content that delegitimises elections, content that incites self-harm, spreads hate or promotes eating disorders (Macdonald & Vaughan, 2024, p. 351). Below, we describe some examples of borderline content that directly capitalise on deception.

Clone websites
Creators of clone websites set up a web address with a title that closely resembles that of an existing organisation. In this way, they capitalise on the organisation's name recognition and reputation. Using these titles, they disseminate messages from the clone website via social media accounts. In this way, the accounts attempt to entice visitors to leave the social media platform and visit the clone websites. These sites feature, for example, politically biased or misleading messages.

In 2022, clone websites of well-known media outlets from Germany, Italy, the United Kingdom and Spain were distributed via botnets on Facebook and Twitter. One of the targets was the German daily newspaper Bild. The web address of the clone page did begin with Bild, but had an addition, such as bild.vip ('Under the Hood of a Doppelgänger – Qurium Media Foundation', 2022).

In the Adversarial Threat Report 2025, Meta described the activities of accounts that created Facebook Pages posing as local media outlets. From these pages, advertisements were purchased targeting various African countries. Meta points out that these pages likely received centralised instructions, as several pages repeated exactly, or almost exactly, the same message (Meta, 2025a).

Disinformation
In addition to misleading information about a website's identity, content can also be misleading. This is often described as misinformation and disinformation.[14] In

13 YouTube described borderline content in a blog post on content moderation as: *'videos that don't quite cross the line of our policies for removal but that we don't necessarily want to recommend to people*'. Meta describes borderline content as content that is not prohibited by the terms of service, but comes very close. (Macdonald and Vaughan, 2024, p. 350).

14 There are various public and academic definitions of misinformation and disinformation (Egelhofer et al., 2020). The explanation of disinformation most commonly given recently, both in academic literature and in the Dutch public debate, is: the dissemination of claims that are largely or entirely (scientifically) unfounded, which the sender *knows* to be incorrect. The purpose of creating and disseminating such content may be to make money, for entertainment, to disrupt the democratic process, or a combination of these.(Humprecht et al., 2020). In the case of misinformation, the criterion regarding the sender's intention does not apply (Wardle & Derakhshan, 2017).

Chapter 2, we provided a recent Dutch example of this: a social media post claiming that voters wishing to vote for GreenLeft–Labour (PvdA-GroenLinks) had to tick two boxes on the ballot paper (which would render the vote invalid). (*Parliamentary Papers II, 35 165, no. 102*, 2026)

Another recently documented example is a social media post by a user who suggested that electoral fraud had taken place during the 2025 general election. This post was shared by fake accounts posing as Dutch citizens but managed from abroad (see also Chapter 2: Case study: Fake accounts from abroad on X during the Dutch elections).

4.1.3 Illegitimate content

Politicians, and female politicians in particular, are increasingly facing online hate and threats (Karlijn Saris & Coen van de Ven, 2021). Through targeted hate campaigns, comment sections on social media platforms can be used to spread a particular narrative about a candidate. Targeted hate campaigns can deter (potential) candidates from standing in elections and can deter potential voters from voting for that candidate. Spreading hate can therefore be an effective way of influencing politicians' participation and the outcome of elections (Honingh & Van Ham, 2024, p. 161).

In 2023, a smear campaign was organised from the United Arab Emirates targeting Dutch public figures. (Heck & Kouwenhoven, 2023) Fake accounts (see also section 4.2.2. Fake accounts) were used to accuse Dutch people of supporting the Muslim Brotherhood. Former GreenLeft (GroenLinks) MP Kauthar Bouchallikht and the Mayor of Arnhem, Ahmed Marcouch, were among the targets of this campaign.

Box3 AI-generated text and images

Generative AI can be used to create realistic-looking videos based on text instructions (see also the *Rathenau Scan Generative AI* report (Rathenau Instituut, 2023).

It is a new type of content that is offered on various social media platforms and can be generated on those platforms. The generated material may be lawful, borderline or unlawful in terms of content. It can also be used as a technique

---

to reach a large audience. Social media platforms may promote generative AI applications on their platforms for business reasons. Recommendation algorithms would then suggest this content to more users (Simon Kruschinski & Fabio Votta, 2025).

The generated photos and videos can be used to reinforce a particular narrative or to convince people of the authenticity of certain images. The visual material paints a political utopia or dystopia of reality. A recently described example of AI-generated material is an image in which the then party leader of GreenLeft--Labour (GroenLinks-PvdA), Frans Timmermans, was depicted in handcuffs (Feenstra & Sabel, 2025).

In the run-up to the 2025 elections, researchers tracked which AI-generated content was disseminated on Facebook, Instagram, TikTok and X by Dutch politicians, political parties and selected influencers and commentators. This shows that only a fraction of this type of content is labelled as AI-generated, even though this is mandatory under platform guidelines (Simon Kruschinski & Fabio Votta, 2025).

4.2 Techniques that increase or decrease visibility

Exerting influence via social media platforms is not just a matter of uploading a certain type of content. The content – whether created by a malicious foreign actor or originating from other users – is also disseminated. The aim of this dissemination is to reach as many people as possible, or indeed a specific target group. By engaging with content – such as liking, sharing, commenting on or viewing it – the recommendation algorithm may spread this content further. It may also be the case that the number of likes, or the number of times it is shared, creates the impression that the content is more popular than it would be if it had fewer likes.

Platforms offer various ways to distribute content. As explained in Chapter 2, recommendation algorithms play a crucial role in distribution; the following sections explore this in more detail.

4.2.1 Public posts and posts in groups

On most platforms, sharing content is referred to as posting. This can be done publicly, making posts visible to everyone, or within (private) groups. In the latter case, only people who have been admitted to a group or who have joined a group themselves will be able to see the posts. Platforms such as Facebook, Telegram and WhatsApp offer group chats. People can join a group or be invited to join one. Users who manage the group can decide for themselves whether the group's existence is made known to all users. If the group is publicly visible, the recommendation algorithm plays a role because the group is recommended to users who are not yet members.

Facebook groups are an example of an online environment where posts are sent only to group members. For example, the AI-generated content about Frans Timmermans was first shared in a publicly accessible Facebook group. This group was publicly searchable and, according to the newspaper *de Volkskrant*, had 130,000 followers in June 2025 (Feenstra & Sabel, 2025).

Interference can occur through setting up or participating in groups of this kind. Groups can be a suitable way to reach specific users, such as people interested in a particular political party.

4.2.2 Streaming

Streaming is a form of real-time communication via a platform in which an individual user has a live video connection with multiple other users. Compared to posting, this form of platform communication is a direct and unfiltered way of reaching groups of users. Major platforms that offer streaming include: TikTok, YouTube, Facebook, Instagram and Twitch.

Recommendation algorithms play a role in streaming because various platforms give live streams a prominent place. In doing so, the platforms promote this form of communication.

Influencers who use live streaming can make good use of platform guidelines to promote their live streams. Content moderation is less effective in live streams than it is for posts. Because a stream is a live connection with users, the content cannot be reviewed before it reaches the audience. The platform can only choose to take a stream offline.

You can influence live streams by leaving comments in the streams or by rewarding specific users with a tip. These reward mechanisms can create a market that makes certain types of content in the stream lucrative (HEIO Consortium et al., 2026, p. 52; *Turning Passion to Profit: WAYS TO MAKE MONEY ON TIKTOK LIVE*, n.d.)

4.2.3 Fake accounts

Fake accounts are also known as sockpuppet accounts, a reference to a sock puppet or hand puppet. Fake accounts are created to assume a fabricated or stolen identity (Rathenau Instituut, 2021).

Sock puppets can generate negative or positive sentiment regarding political parties, politicians or socio-political issues.

Activity from accounts of this kind has been observed ever since the advent of the internet. Around the time of the Dutch parliamentary elections, fake accounts were spotted on TikTok and LinkedIn posing as the account of PVV leader Geert Wilders (HEIO Consortium et al., 2026, p. 39). Journalists from RTL Nieuws identified 550 fake accounts operating from abroad that were interfering in Dutch-language messaging on X and Facebook (see also Chapter 2: Case study: Fake accounts from abroad on X during Dutch elections).[15]

The reach of such content can extend beyond the platforms themselves. For example, posts from sockpuppet accounts have in the past been picked up by mainstream media, even though these were accounts with fabricated names managed from Russia. (*All major media outlets in Norway fall for Tweets from Russian trolls*, 2020).

4.2.4 Coordinated posting

A commonly described method used to influence online debates is the coordinated posting of content. In coordinated posting, accounts work together to disseminate specific content at an agreed time. These coordinated campaigns aim to generate high levels of engagement and create virality, with the goal of making the content appear popular (Rogers & Righetti, 2025). Recently, Israeli companies launched two applications that allowed users to disseminate pro-Israeli content in an instant

15 In 2025, X (formerly Twitter) provided details for many accounts regarding the IP location and App Store from which these accounts were active (Nikita Bier [@nikitabier], 2025).

and flood social media platforms with complaints about content unfavourable to Israel. The app could be used in such a way that the content appeared to have been created spontaneously by the users themselves (Beemsterboer, 2026).

This form of interference has been observed internationally and documented in academic research. Political communication scholar Thiele and colleagues (Thiele et al., 2025) provide an overview of examples of coordinated posting that have been extensively studied. Thiele argues that campaigns vary in terms of scale, the degree to which they operate covertly, and the impact of the content. Campaigns labelled as interference by the researchers vary in scale. In large-scale interference, content is uploaded using many different accounts over a long period across multiple platforms simultaneously. These kinds of large-scale interference campaigns can be purchased from providers such as the company Internet Research Agency (Thiele et al., 2025, p. 14).

Thiele and colleagues also described campaigns targeting a specific audience, in which exactly the same message was copied and uploaded by different accounts each time. A similar attempt to influence public opinion has also recently been observed in the Netherlands. Several accounts disseminated identical Dutch-language messages within a short period of time. These messages all linked to the same web address, suggesting that the accounts were likely managed by a single individual. Incidentally, the researchers did not attribute this campaign to a foreign actor, but to a Dutch account (Justice for Prosperity, 2025).

Israeli companies recently launched two applications that allow users to disseminate pro-Israeli content in an instant and flood social media platforms with complaints about content deemed unfavourable to Israel. The app was designed to make it appear as though the content had been created spontaneously by the users themselves (Beemsterboer, 2026).

4.2.5 Bot accounts and hyperactivity

The literature describes various influence campaigns in which thousands of bot accounts (botnets) have been deployed to increase the reach of messages with political content. Thiele and colleagues (2025) provide an overview of studies in which interference involving coordinated botnets has played a role.

A frequently cited example is a botnet network operated from Russia that was active on Facebook, Reddit and Twitter in the run-up to the 2016 US presidential election. Using a thousand accounts, a wide variety of messages were disseminated over a long period across various platforms, including 10.4 million

Tweets and 116,000 Instagram posts. The Tweets addressed political news, to which other bot accounts then responded (Thiele et al., 2025).

Activities carried out by hyperactive accounts include sharing, liking and commenting on existing content. The aim is to increase the visibility of specific accounts or the content being commented on across platforms. This can be done directly, for example, by commenting on posts from a politician's account. However, research by RTL Nieuws shows that it also happened indirectly: by commenting on accounts that deal with political issues or political parties, but are not themselves affiliated with a political party. The latter is presumably done to operate less conspicuously (Jasper Bunskoek et al., 2025).

As described in section 3.6.3, research shows that hyperactive accounts can be successful because recommendation algorithms reward activity. This can lead to the over-representation of a particular type of content. Individuals or groups of organised users can capitalise on this by creating a hyperactive account.

Engagement on demand
In the run-up to the election, a network was identified that had liked a large number of posts on X by Frans Timmermans (Justice for Prosperity, 2025, p. 17). The accounts that had liked the posts had a distinctive signature in their usernames, which made it possible to trace them back to a Vietnamese provider.

The Vietnamese provider's website states that campaigns can be ordered for social media platforms: Facebook, TikTok, Instagram, YouTube, X, Threads and LinkedIn: *'We provide packages to increase likes, follows, views, and comments on social media platforms such as Facebook, TikTok, Instagram, YouTube, etc., through an automated system. You simply need to send the link you wish to boost engagement for, and the system will process it according to your requirements within the agreed timeframe.'* (Own translation, like.vn, n.d.)[16]

4.2.6 Hashtag boosting

Many platforms use hashtags. The purpose of using hashtags is to tag the content of posts. For example, when users add #parliamentaryelections to content, they are indicating that the post is about this topic. By clicking on the hashtag, other users can view a list of content with the same tag. Many platforms display a list of the

16 Translated from Vietnamese using DeepL translation software. The original text on the website is: *'We offer packages to increase likes, followers, views and comments on social media platforms such as Facebook, TikTok, Instagram and YouTube... via an automated system. You simply need to send the link for which you wish to increase engagement, and the system will process it in accordance with your requirements within the agreed timeframe*.'

most popular hashtags at that moment. As a result, content with these hashtags has a wider reach than less popular hashtags.

During the Romanian elections, it was observed how popular hashtags were used to boost the reach of a particular candidate's content.

Coordinated content distribution can drown out other content, thereby reducing its visibility. Treré and colleagues refer to this as 'flooding'. They illustrate this with an example of users who identify as K-pop fans. This fan movement uploaded music videos of K-pop groups and *memes* using the then-popular hashtags #MAGA and #whitelivesmatter. The aim of the uploads was to drown out racist and hateful messages that were also being posted with these hashtags.

The authors also give an example of flooding where it was unclear who was behind the activities. In Mexico in 2014, a social movement emerged following the deaths of six students and the disappearance of 43 others. On Twitter, the hashtag #YaMeCanse was used to organise protests against the violence. Subsequently, accounts became active that introduced noise by using the hashtag to disseminate content that contravened the platform's guidelines (pornographic material, advertisements and violent images). This created so much noise amongst the posts under the hashtags that it became difficult for the social movement to organise protests, after which the hashtag disappeared from the platform's trending list (Treré & Bonini, 2024, p. 313).

4.2.7 Influencer marketing

One way to indirectly influence the online public debate on social media platforms is to pay influential users with a large following to share content. An example of this is paying influencers.

This form of dissemination has been described by Goanta of Utrecht University, who specialises in law, economics and policy. During the presidential elections in Romania, she examined influencers on TikTok who were found to be producing many similar videos. It was striking that they all asked their followers to describe an ideal presidential candidate, whilst highlighting the qualities of candidate Georgescu. An advertisement on the FameUp platform for influencers revealed that this was a paid assignment carried out by the influencers (*FameUp – FameUp – Over 500+ Nano and Micro Influencers Activated in 24 Hours*, n.d.).

An analysis of influencer accounts revealed a wide variety of influencers: from content creators specialising in fashion, beauty, travel and fitness. In only seven of

the forty videos analysed did influencers disclose that the content was part of a paid marketing campaign (Goanta, 2024).

This example from Romania shows that, within influencer marketing, it is sometimes decided to pay influencers who have relatively few followers but a specific target audience. This is known as micro- or nano-influencing. In this way, advertisers can allocate their resources more effectively. This helps them avoid spending money on a message aimed at the wrong target audience.

Influencers can be funded without them knowing where the money comes from. On TikTok, for example, someone can donate money to an account without contacting the influencer. This makes it possible to exert covert influence over the content that influencers share (Goanta, 2024).

4.3 Conclusion

Platforms are increasingly using recommendation algorithms to rank content. These algorithms make it easier for users to reach people who have not explicitly indicated that they want to see their content. This increases the ability to influence what content reaches people. For example, when content goes viral. But even if content does not go viral, it can still achieve a wide reach via recommendation algorithms because it is shown to people who have never indicated that they follow the poster.

As a result, malicious (foreign) actors have a wide range of options for influencing who sees what content. They can choose to post content and reach as large a group of users as possible via recommendation algorithms, or attempt to target a specific group of users.

Through activities such as managing multiple hyperactive accounts, coordinated posting and hashtag boosting, attempts are made to create a flywheel effect.

Platforms are constantly evolving. As a result, the opportunities for interference are also changing. This makes it a challenge for researchers and supervisory bodies to stay up to date.

The tools are available to everyone. Interference activities often become intertwined with domestic activities, blurring the line between foreign influence and domestic activities.

5 Existing efforts to prevent interference via recommendation algorithms

In this chapter, we examine what efforts are already being made to prevent interference via recommendation algorithms. We answer this question by analysing existing legislation and regulations. Following a description of the main features of legislation and regulations, we briefly outline the existing regulatory framework based on insights from an expert session organised by the Rathenau Instituut (see Appendix 1) and academic research. We then provide a brief overview of the efforts made by platforms, based on their own reports and (written) interviews. We also assess these efforts in the light of the expert session and the literature.

5.1 Existing legislation and policy

To answer the sub-question, an analysis was carried out of the most relevant legislation and policy relating to interference and/or recommendation algorithms. It is important to note that the laws each have their own scope and objectives and therefore approach the problem of interference via recommendation algorithms from different angles and in both direct and indirect ways.

Relevant laws and policies are:
- AI Regulation
- General Data Protection Regulation (GDPR)
- Digital Services Act (DSA) and the Code of Practice on Disinformation
- European Media Freedom Act (EMFA)
- European Democracy Shield (EDS)
- Unfair Commercial Practices Directive
- Audiovisual Media Services Directive
- Government-wide strategy for the effective tackling of disinformation
- Regulation on transparency and targeted political advertising (VPR)

We will not discuss the Electoral Act here. That Act was briefly mentioned in Chapter 1.

A detailed analysis can be found in Annex 3. In the sections below, we discuss the main points of the legislation.

A great deal has been done to limit interference via social media platforms. Legislators and policymakers are clearly aware of the problem and have developed various tools to address it.

In terms of policy, one key initiative is the *government-wide strategy for effectively tackling disinformation*. This national approach aims to counter the spread of disinformation at national level, strengthen citizens' resilience and enhance knowledge development in this area. In addition, there is the European Democracy Shield (EDS), designed to support the efforts of European Member States to strengthen their democracies. As part of the EDS, a European Centre for Democratic Resilience has been established with the aim of collaborating and exchanging information in areas including disinformation and foreign information manipulation and interference (FIMI).

With regard to the most relevant legislation, three key themes can be identified that are relevant to preventing and combating interference:

- Increasing the accountability of platforms. (section 5.1.1)
- Intervening in the design of platforms. (section 5.1.2)
- Transparency regarding circulating content. (section 5.1.3)

5.1.1 Increasing the accountability of platforms

A key theme in the relevant legislation is the emphasis on increasing the accountability of platforms. This is achieved primarily through the Digital Services Act (DSA).[17] The DSA aims to protect users of digital platforms from the dissemination of illegal content, whilst respecting their fundamental rights. To this end, the legislation sets out responsibilities and accountability requirements for providers of intermediary services, including social media platforms.

The DSA requires platforms to limit the dissemination of illegal content as much as possible. Various types of content (both offline and online) are illegal in European Member States, including the Netherlands. Examples include: discriminatory content, threatening content and terrorist content. In addition, the DSA contains provisions regarding complaints procedures and due diligence obligations for platforms. For example, platforms must respond appropriately once illegal content has been posted.

17 Other legislation, such as the Regulation on Transparency and Targeted Political Advertising (VPR) and the General Data Protection Regulation (GDPR), also impose accountability obligations on platforms. We discuss these briefly below and in more detail in Appendix 3.

Key provisions relating to interference are Articles 34 and 35, which require Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) to adopt a proactive approach. These include, amongst others, YouTube, Facebook, Instagram, LinkedIn, TikTok, Twitch, Google Search, Snapchat, X, WhatsApp and Bing.

Under the DSA, very large platforms and search engines must identify systemic risks and take reasonable, proportionate and effective measures. For example, to prevent negative impacts on electoral processes. This could include the circulation of misleading information about candidates, voting data or fabricated expressions of support for candidates. The platforms are free to determine for themselves how they implement these measures. Platforms must report on their efforts to the European Commission.

The European Commission is currently investigating the functioning of X's recommendation systems (European Commission, 2025a; O'Carroll, 2025). The European Commission is also investigating TikTok's recommendation algorithms. Firstly, in relation to 'coordinated inauthentic manipulation or automated exploitation of the service'. And secondly, regarding TikTok's policies on paid political content and political advertisements (European Commission, 2025e).

The DSA sets out proposals to mitigate systemic risks posed by platforms. These include adjusting recommendation algorithms to temporarily limit the spread of misleading or harmful content. The DSA also makes a number of proposals. For example, to monitor viral trends on other platforms, deploy moderation teams familiar with the local language and culture, and temporarily ban AI-generated content in political advertisements.

The European Commission has drawn up guidelines for platforms specifically in relation to electoral processes. These include removing financial incentives for platforms and influencers to spread disinformation, hate speech and extremist content via advertisements (demonetisation) (European Commission, 2024a). These guidelines also stipulate that measures must be taken against botnets, inauthentic accounts or other misleading uses of the service (anti-manipulation).

Platform providers must specify, in clear and understandable language, the key parameters used in their recommendation systems. Users should have the option to modify or influence these parameters. Under the DSA, users must be offered, as standard, a feed containing content solely from accounts they follow (a

chronological timeline), in addition to a timeline generated by recommendation algorithms.[18]

In addition, platforms and search engines must provide designated supervisory authorities and researchers with insight into how these platforms operate. Supervisory authorities may request data to assess compliance with the DSA, and platforms must, at the request of supervisory authorities, explain how their recommendation algorithms work. Designated researchers must also be granted access to data to investigate systemic risks. Platforms may only refuse a request from researchers if they do not possess the data or if access would lead to serious security or confidentiality risks.[19]

Finally, there is the Code of Conduct on Disinformation, which can be seen as a further elaboration of the rules set out in the DSA. For example, signatories to the Code of Conduct must endeavour to limit unauthorised manipulative behaviour and practices. This includes the use of bots, fake accounts and coordinated inauthentic behaviour, as well as 'user behaviour' aimed at artificially increasing the reach or perceived public support for disinformation. Reference is made to a comprehensive, public and regularly updated database of *tactics, techniques and procedures* maintained by the DISARM Foundation (DISARM Foundation, n.d.).

5.1.2 Intervening in the design of platforms

There is legislation aimed at regulating the design of platforms. For example, the DSA requires very large platforms and search engines to mitigate the risks of addictive design. An addictive design is, for example, a highly personalised recommendation algorithm. Addressing addictive design is relevant in the context of this research insofar as it involves intervening in engagement-based recommendation algorithms that malicious actors can exploit.

The European Commission's preliminary finding is that TikTok has not taken sufficient measures to mitigate the addictive aspects of its design (European

18 The NGO Bits of Freedom has won a preliminary injunction against Meta, in which it demanded that users be able to set the chronological timeline as the default, rather than the algorithmic timeline (*Meta must (for the time being) offer the chronological timeline as the permanent default option*, 2026).

19 X has recently been fined for breaching transparency obligations; among other things, it restricted access to data for researchers and provided insufficient insight into advertisements (European Commission, 2025d). The European Commission has also stated in preliminary findings that Meta and TikTok did not provide sufficient access to research data. The investigation is ongoing (European Commission, 2025c). Another investigation is also underway against Meta regarding poor access to research data and the takedown of CrowdTangle, which was intended to monitor elections (European Commission, 2024b).

Commission, 2026).[20] The Commission has also launched proceedings against Meta's Instagram and Facebook on the grounds that their design is allegedly too addictive for minors (European Commission, 2024c).

Under the AI Regulation, addictive design in an AI system may even be regarded as a prohibited practice if it exploits users' age, disability or socio-economic background. It also follows from the GDPR that the collection and processing of a person's data with the aim of making the user addicted is likely to be incompatible with the principle of lawfulness. The outgoing government has indicated that it will closely monitor developments in the European Commission's aforementioned paragraph and, alternatively, focus on the Digital Fairness Act, which centres on regulating addictive design on social media platforms (*Parliamentary Papers II, 26 643, no. 1391*, 2025). The coalition agreement also announced plans to tackle addictive algorithms (D66, VVD and CDA, 2026).

Furthermore, attention is being paid to manipulative design. The DSA prohibits manipulative techniques. Addictive design may also fall under manipulative techniques, but this remains to be seen (*Parliamentary Papers II, 26 643, no. 1391*, 2025). X has recently been fined by the Commission for misleading users with blue ticks (European Commission, 2025d).

There is a ban on AI systems that use manipulative techniques. Whether social media platforms fall under the category of manipulative systems is a matter of debate. At the time of writing, a legal case is underway brought by the SOMI Foundation against TikTok, alleging that the platform specifically manipulates young people and misuses sensitive personal data, which is used by the platform to drive its own recommendation algorithm (*Dutch claims foundation launches legal action against TikTok and X (formerly Twitter)*, 2025).

Under the GDPR, the processing of personal data must not be disproportionate, unexpected or misleading (the principle of fairness). According to the law, these constitute manipulative techniques. The European Parliament's research service, the EPRS, expects that the recommendation system used by platforms such as TikTok – which, according to the research service, utilises people's behavioural data to manipulate them – is incompatible with the principle of fairness under the GDPR (European Parliamentary Research Service, 2025).

20 There was also a (now concluded) investigation by the European Commission into, among other things, the addictive elements of the TikTok Lite Awards Programme, which TikTok has since withdrawn (European Commission, 2024d).

5.1.3 Transparency regarding circulating content

A third common thread, alongside increasing accountability and intervening in the design process, is that the legislator aims to enhance the transparency of online content. The laws primarily impose transparency requirements on content creators, but platforms themselves must also comply with transparency obligations. The laws may enable third parties (citizens, researchers, public authorities) to conduct research, for example into interference activities. For instance, under the AI Regulation and the DSA Regulation, AI-generated images that are virtually indistinguishable from real ones must be labelled.

The role of influencers, such as vloggers and bloggers who receive payment for (political) messages, has also been taken into account. The European Media Freedom Act (EMFA) defines, among other things, what constitutes a media service. Professional influencers may also fall under this definition. Under the EMFA, very large platforms and search engines must incorporate a feature enabling media service providers to make a series of declarations: that they are media providers, that they are editorially independent of Member States, political parties and third countries, that they comply with regulatory requirements, and that they do not offer AI-generated content without human review.

Furthermore, the Unfair Commercial Practices Directive requires influencers to be transparent about the advertising they produce and the benefits they derive from it. The use of fake accounts to generate fake likes and fake followers is not permitted. The Audiovisual Media Services Directive also stipulates that misleading or aggressive practices and the use of subliminal techniques (hidden messages) by anyone disseminating advertising, including in videos and live streams, are not permitted.

Under the banner of increasing the transparency of content in circulation, there is also the Regulation on Transparency and Targeted Political Advertising (VPR). The aim of this legislation is to help citizens recognise political advertising online more easily. Interference and online manipulation were among the reasons for drafting the VPR.

The VPR imposes transparency requirements on political advertising and sets out rules governing the use of so-called targeting and optimisation techniques for the dissemination of political advertising. Under the VPR, the very large platforms and search engines also have a few specific obligations, primarily focused on public transparency regarding the advertisements and responding promptly to requests from, for example, complainants and regulators. The VPR stipulates that, from three

months before the elections take place, political advertising services may only be provided to clients who are EU citizens or have a specific connection to the EU.

Under the DSA, very large platforms and search engines must maintain registers containing information about the advertising (such as the intended target audience and the advertiser) placed on their platforms. These registers must be publicly accessible.

5.1.4 Are there plans to relax the rules?

At the time of writing this report, the European Commission has proposed a so-called Digital Omnibus with the aim of simplifying rules for businesses (*Digital Omnibus Proposal COM/2025/837*, 2025). Although the business community needs a reduction in the regulatory burden, there are concerns, including within the outgoing Dutch government, that this will lead to undesirable deregulation (*Parliamentary Papers II, 22 112, no. 4223*, 2025, no. 4223 2025).

One of the possible changes in this regard is to postpone the implementation of requirements for high-risk systems and to allow providers to exempt themselves from the obligations associated with a high-risk system without having to declare this in a designated public database. As a result, AI systems that could be used to influence elections might remain under the radar of the regulator.

Another change in this regard is that the definition of personal data may need to be interpreted less broadly. As a result, organisations would be less likely to be required to comply with the GDPR, and in such cases, data controllers would no longer be required to adhere to the principle of fairness. This is because certain data would no longer fall within the definition of personal data. Furthermore, the training and operation of AI systems using personal data (including sensitive data such as political information) would be permitted without the need for consent.

Box4 Geopolitical context: platform regulation under fire

Since the inauguration of the second Trump administration in the United States, European platform regulation has come under fire. President Trump, Vice-President Vance, and more recently the Judiciary Committee of the US House of Representatives, have levelled strong criticism at European

---

regulations, particularly the DSA. They accuse the European Commission, the ACM (the DSA regulator in the Netherlands), and several experts we spoke to for this report (as they attended a roundtable on interference via social media platforms organised by the ACM) of censorship, view the European Commission as the party that, through the DSA, controls public debate in America, and are threatening to take action regarding trade relations or military cooperation if Europe tightens its grip on platforms.

This American perspective is at odds with the European view that American platforms actually exert a major influence on public debate in Europe, that they bear responsibility for the negative effects of their services, and that rules are therefore necessary. Furthermore, it overlooks the fact that the DSA deliberately does not prescribe what may or may not be said online (that is the role of national laws, for example by prohibiting discrimination, hate speech or *deepfakes*), but instead imposes an accountability obligation on platforms to clarify how they comply with laws, how they should handle user reports and mitigate systemic risks. With these developments, platform regulation has found itself at the heart of the geopolitical power struggle. (Christina Lu, 2026; Committee on the Judiciary of the U.S. House of Representatives, 2026; *Statement by the European Commission on the U.S. Decision to Impose Travel Restrictions on Certain EU Individuals*, 2025)

5.1.5 Experts: Legislation does not appear to address all aspects of the tools for interference, and platforms' willingness to comply is high

Although much has been regulated, it appears that the legislation does not sufficiently address the tools used by actors seeking to influence the use of AI. It is striking that AI systems used to influence the outcome of an election or referendum, or voting behaviour, are not prohibited under the AI Regulation, but fall into the high-risk category. According to experts, younger generations in particular make frequent use of these systems when making everyday decisions. According to the Dutch Data Protection Authority, eligible voters increasingly used AI chatbots to help them decide how to vote in the run-up to the previous elections (Dutch Data Protection Authority, 2025).

Furthermore, experts are warning of the rise of AI-generated and personalised news summaries (Schneier & Sanders, 2025). It is not yet clear which category

these summaries fall under in the AI Regulation, as the Regulation – in addition to high-risk AI systems – regulates the *models* used for generative AI tools, rather than the *tools* themselves. It cannot be ruled out that these tools may constitute high-risk AI systems. Finally, one expert highlighted the need for greater attention to be paid to interference via chatbots, which are currently subject to limited regulation under the Regulation. Under the law, it must now simply be clear to the user of a chatbot that they are conversing with a chatbot.

The willingness of platforms plays a key role in the success of legislation on interference. As described in this chapter, regulators have a large number of investigations currently underway.

In addition, experts have criticised the efforts made by Google and Meta in the run-up to the Regulation on transparency and targeted political advertising. With the introduction of this regulation, Google and Meta have indicated that they will no longer facilitate political advertising on their platforms, partly because, according to both companies, this would lead to significant complexity and legal uncertainties for advertisers and platforms in the EU (*Google to stop political advertising from September*, 2025). Google has also deleted its seven-year archive of European political advertisements, which further hinders researchers' ability to investigate interference (404 Media, 2025). Meta further stated that it would no longer permit 'social issues' either.

Experts argue that these decisions have an impact on freedom of expression. They also believe that banning political advertisements encourages politicians and others to disseminate inflammatory content in order to reach their target audiences via recommendation algorithms (Eijsvoogel, 2025; *No more political advertising on Facebook and Instagram in the middle of the campaign period*, 2025). The aim of this law – to make it easier for citizens to recognise political advertising online – seems, as a result of these platforms' choices, to be slipping further out of sight.

5.2 Efforts by social media platforms

Social media platforms are taking various measures to prevent interference in elections. In this section, we discuss *community guidelines* and content moderation.

For this study, we approached Meta (Facebook and Instagram), TikTok, Snap (Snapchat), Google (YouTube), Microsoft (LinkedIn) and X to arrange online interviews (see Appendices 1 and 2 for the methodological rationale). Of these, Meta and TikTok responded in writing. We conducted an oral interview with Snap. Google and Microsoft chose not to respond to our request, and X did not respond.

For this section, we have drawn on interviews and written responses, as well as publicly available information from social media platforms.

5.2.1 Community guidelines

Social media platforms have community guidelines setting out what is and is not permitted on their platforms (see section 3.3.1). These guidelines cover both content and activities, such as methods of dissemination. Under Article 14 of the DSA, platforms are also obliged to enforce their own guidelines.

Community guidelines serve as the basis for moderation in online environments. They are the rules of conduct or house rules that determine which behaviours are and are not tolerated. For example, they set out rules specifying what language and images users are and are not permitted to use on a platform. Popular online environments almost always employ some form of centralised content moderation, whereby platforms set the rules and decide how to enforce them (Rathenau Instituut, 2025a).

Content moderation can result in content being removed from the platform, but also in the reach of a post being restricted based on its content. Platforms apply different rules in this regard. Snap, for example, applies stricter rules to content that may be recommended by its recommendation algorithms than to content that can be shared by people within their own network (Snapchat, n.d.). Meta's Community Standards specify the content that is prohibited when it comes to violence or hateful behaviour. In addition, Meta states that the reach of certain forms of violent content is restricted to people over the age of eighteen and a warning is displayed. Meta automatically places users with a teen account on the strictest setting regarding sensitive content (Meta, n.d.).

All major social media platforms have guidelines relating to interference tools (Chapter 4): the sharing of misinformation, unlabelled AI-generated content, hate speech, and the use of fake accounts. Platforms therefore have explicit policies in place to counter interference activities.

Under the DSA, very large online platforms (VLOPs) are required to publish annual transparency reports detailing how they have carried out content moderation (see Table 5 below). These reports include, among other things, the number of posts removed for breaching the guidelines.

To provide an overview of public sources in which platforms report on their efforts, Table 5 (below) lists references to transparency reports and guidelines relating to the tools for countering interference described in Chapter 3.

Table 5 Overview of efforts reported by platforms to counter interference

Organisation	*Community guidelines*	Specific measures against interference in elections
Google (YouTube)	Transparency report, general guidelines, and specifically: policy against misleading information about elections, policy against fake engagement, and policy against impersonation	Google wrote about preparations for the US elections
Meta (Facebook and Instagram)	Transparency report, general guidelines and specific policies against hateful conduct, policies against inauthentic behaviour and policies aimed at combating disinformation	Election integrity reports blog on global elections on Meta platforms regarding the prevention of foreign interference
Microsoft (LinkedIn)	Transparency report, general guidelines, specific guidelines on being trustworthy, guidelines on inaccurate or misleading content	Not known
Snap (Snapchat)	Transparency report, general guidelines, specific guidelines against harmful, false or misleading practices, and policy against hateful content	Reflections on Dutch elections
TikTok	Transparency report, general guidelines, specific policies on integrity and authenticity, including Edited Media and AI-Generated Content (AIGC)	Global elections hub, blog on safeguarding the integrity of the Dutch parliamentary elections
X (formerly Twitter)	Transparency report, general guidelines, specific guidelines on authenticity and guidelines on hateful conduct	Not known

The table provides a non-exhaustive overview of public documents in which platform companies describe their efforts to combat election interference via their social media platforms. Compiled by: Rathenau Instituut

5.2.2 Content moderation

Social media platforms use a combination of human and automated content moderation to proactively detect harmful content. For some content, it is immediately clear to platforms whether something has been manipulated, is harmful or is inaccurate. This is done automatically using AI systems and/or with the help of content reviewers. For some content, it is necessary to have knowledge of the local language and culture in order to assess it. Table 6 provides an overview of the

number of Dutch-speaking content moderators employed by major online platforms, as reported by the platforms themselves.

Table6 : Dutch-speaking content moderators according to self-reports

Organisation	Number of Dutch-speaking content moderators	Number of users in the Netherlands
Facebook and Instagram (Meta)	111	Facebook 10,300,000 Instagram 11,900,000
LinkedIn (Microsoft)	8	5,300,000
Snap (Snapchat)	23	6,148,873
TikTok	100	6,700,000
X (formerly Twitter)	0	8,242,130
YouTube (Google)	75	*42,100,000

Dutch-speaking content moderators and user numbers in the Netherlands as of 1 August 2025 according to platform reports. * Google does not distinguish between unique monthly users, so this figure is difficult to compare. Source: (Google, 2025a, 2025b; LinkedIn, 2025; Meta, 2025b, 2025c; Snap, 2025a; TikTok, 2025; X, n.d.).

In the (written) interviews with Meta, TikTok and Snap, they provided examples of initiatives they have taken in the run-up to the 2025 general election. Whilst not exhaustive, we provide here an overview of the type of efforts these platforms claim to be making.

Enforcement of their own guidelines

TikTok states that it has removed ‘more than 4,000 videos’ that breached guidelines on ‘civic integrity, misinformation and AI-generated content’ both before and after the 2025 Dutch parliamentary elections. TikTok states that more than 96% of those videos were removed before anyone reported them. TikTok also says it has removed more than 1,000 accounts for impersonating Dutch election candidates and government officials.

TikTok also highlights initiatives it has taken that are specifically aimed at influencer marketing. TikTok states that political fundraising and political campaign fundraising are specifically prohibited on the platform. This also means that paid advertisements related to politics or elections are prohibited, and that creators may not be paid for such content.

Snap states that during the period leading up to, during and following the 2025 Dutch parliamentary elections, a handful of posts were removed in accordance with its policy on the dissemination of false information. Among the reported content (16

items) were ‘an AI-generated post and one that was an attempt at humour’. The removed posts were ‘visible for a short (limited) time, and then removed’. Snap states that there was only one instance of political disinformation, and that this post was removed before it could reach a wider audience.

Every day, Snap takes 'a sample of Snaps that have passed through all moderation checks, to see what’s in them, what’s going on, and whether their policy has worked.'

Snap states that it will allow political and news content from publishers with whom Snap collaborates, as well as from verified ‘Snap Stars’, to appear on Spotlight. Only this content may be shared with a wider audience. Snap refers to this as ‘gatekeeping’: only certain parties are permitted to communicate about elections.

Snap states that content must not be misleading or incorrect: the platform does not tolerate ‘a bit of misinformation’. Snap also states that it does not label or downrank misinformation.

Fake accounts and coordinated actions

Platforms are highlighting various initiatives aimed at combating fake accounts. TikTok states that it is pushing for account verification to provide users with better information about who is behind an account, so that its reliability can be assessed.

Meta uses automated systems to detect ‘inauthentic behaviour’ and fake accounts. In its responses to our questions, Meta states that detection technology ‘blocks millions of attempts to create fake accounts every day and blocks millions of newly created fake accounts within minutes of their creation’. Meta also states: ‘Between July and September 2025, we removed 698 million fake accounts, 99.9% of which were identified and dealt with by us before users could report them’.

Snap states that it has never observed any coordinated actions. Snap believes that its current measures are so effective that coordinated actions cannot take place on its platform.

Measures relating to elections

TikTok and Meta have announced that they are training content moderators and teams specifically in relation to the elections. TikTok is inviting speakers with expertise in ‘election integrity, media literacy and disinformation’ to deliver internal presentations ahead of the elections. Meta states: 'There are several multidisciplinary teams within Meta that focus on election-related topics and work together to identify, assess and mitigate specific risks surrounding the elections.

New risks are also monitored and assessed so that relevant mitigations can be applied.'

Meta is working with fact-checking partners Agence France-Presse (AFP) and Deutsche Presse-Agentur. These partners assess and classify viral misinformation, including applying the ‘Edited’ label to content that has been altered in a way that could mislead people, for example through the use of AI. Meta states that content which violates Meta’s policies will be removed.

Meta also states that it will direct users to reliable information about the election date and refer people to the UK government’s website for further details about the elections.

Snap states that it has ‘a highly cross-functional team comprising various disciplines, which comes together in the event of a crisis’. Snap also states that it ‘holds regular policy briefings for its moderation teams, which may also cover developments relating to elections’.

5.2.3 Experts: greater transparency needed for independent observers

What users see online and the role that recommendation algorithms play in this remains largely unknown to independent observers (Cooper & Chapman, 2025). Greater transparency is needed to investigate how recommendation algorithms work in practice.

Participants in our expert session indicated that they lack, among other things, the following information to be able to conduct proper research into the risks of interference, manipulation and deception on social media platforms:

- Information about the reach of (political) content. Who has seen the content? Even before it is removed by a platform.
- Information about their monitoring processes: how do platforms detect manipulative networks of accounts, and how do we know if this is effective enough?
- Information on the origin of accounts.
- More precise information about the reach and origin of advertisements.
- Insight into how recommendation algorithms work and the weights and business rules behind the recommendations.

The experts also point out that researchers' access to platforms – a requirement under the DSA – leaves much to be desired. Requests are currently often rejected, and it is frequently unclear what information platforms hold that could be requested (see also section 5.1).

5.2.4 Experts: enforcement of *community guidelines* falls short

The enforcement of community guidelines falls short. This is the conclusion reached by researchers at the Hybrid Election Integrity Observatory, a collaboration between Post-X Society, AI Forensics, Trollrensics, the University of Amsterdam and Justice for Prosperity. The research collective argues that content moderation during the 2025 Dutch parliamentary elections offered 'insufficient protection'. They base this on platforms' failure to act on reports of illegal content and their failure to remove content that contravened their *community guidelines*. The researchers argue that platforms' community guidelines 'exist primarily for PR purposes' and 'do not contribute to user safety' (HEIO Consortium et al., 2026, p. 59).

There are also signs in other countries that content moderation falls short during election periods. The Fimi Defenders for Election Integrity (FDEI) consortium, comprising ten organisations working together, states in its report on the Czech parliamentary elections that TikTok removed 187,000 fake accounts, 2.9 million fake likes and 2 million fake followers prior to the Czech parliamentary elections (GLOBSEC et al., 2025). Despite this, a network of pro-Russian TikTok accounts still managed to garner between 5 and 9 million views without being taken offline.

5.3 Conclusion

A great deal has been done to prevent the risk of interference via social media platforms. The legislator is clearly aware of the problem and has developed various tools to prevent interference. In doing so, the legislator focuses on those within its sphere of influence (influencers, platforms) rather than on the interfering actors themselves. Key themes in the regulatory framework are aimed at increasing the accountability of platforms, intervening in the design of platforms and achieving greater transparency regarding circulating content.

Major online platforms are working to prevent election interference via their platforms through the use of guidelines and content moderation. As explained in section 5.1.1, the DSA also requires them to take measures to protect the integrity of elections.

Nevertheless, findings from NGOs and academics – even with limited access to platform data – show that platforms do not always succeed in preventing interference activities or reducing the tools available for such purposes. Examples include coordinated actions (such as liking or sharing posts), sharing content that contravenes platforms' codes of conduct, delayed responses to reports, and facilitating the dissemination of illegal content. The experts we consulted in our expert session (see Annexes 1 and 2) also point out that the risks of interference via recommendation algorithms are significant and diverse, and that further measures are both necessary and possible.

The following chapter sets out policy options for politicians and policymakers to further mitigate the risks of interference via social media platforms' recommendation algorithms.

6 Conclusion and policy recommendations

In this chapter, we first answer the central question of this study: what role do recommendation algorithms on major social media platforms play in election interference? We do this on the basis of all the preceding chapters. We then address the final sub-question: what (policy) measures are possible to prevent and/or tackle interference via recommendation algorithms?

6.1 Conclusion

Research question: what role do recommendation algorithms on major social media platforms play in election interference?

Malicious foreign state actors can exploit social media recommendation algorithms by increasing or decreasing the visibility of certain content. Interference occurs when these actors attempt to deliberately mislead the recipient about the content and/or its popularity. The ultimate aim is to influence public opinion or elections. In doing so, they know how to exploit the social media landscape.

Social media platforms have increasingly relied on engagement-based recommendation algorithms, thereby creating an online environment where content that plays on emotion, deception and incitement can achieve a wide reach. Malicious foreign actors capitalise on this by creating this type of content and attempting to give it a wider reach.

Technically, recommendation algorithms can be configured differently, but fundamental changes run counter to the business model of platform companies: platforms benefit from engagement-based recommendation algorithms, as these increase the amount of time users spend on a platform. The tools for interference activities are therefore effectively built into the platforms.

Much has been done to prevent or combat the risk of interference via social media platforms. The legislator is clearly aware of the problem and has developed various instruments to tackle different aspects of interference. Platforms have also drawn up rules. Nevertheless, experts still point to the need for additional measures and enforcement. To reduce risk, we propose courses of action in three areas: changing

platform design, strengthening the information position and increasing resilience. The following section deals with this.

6.2 Areas for action

We structure the courses of action based on three approaches that emerge from our literature review, interviews, expert session and analysis thereof

The three approaches are:
1. Changing the platform design. (section 6.2.1)
2. Strengthening the information position. (section 6.2.2)
3. Increasing resilience. (section 6.2.3)

Figure 4 summarises the courses of action. We explain them in more detail in the sections that follow.

Figure 4 Options for action perspectives to reduce interference

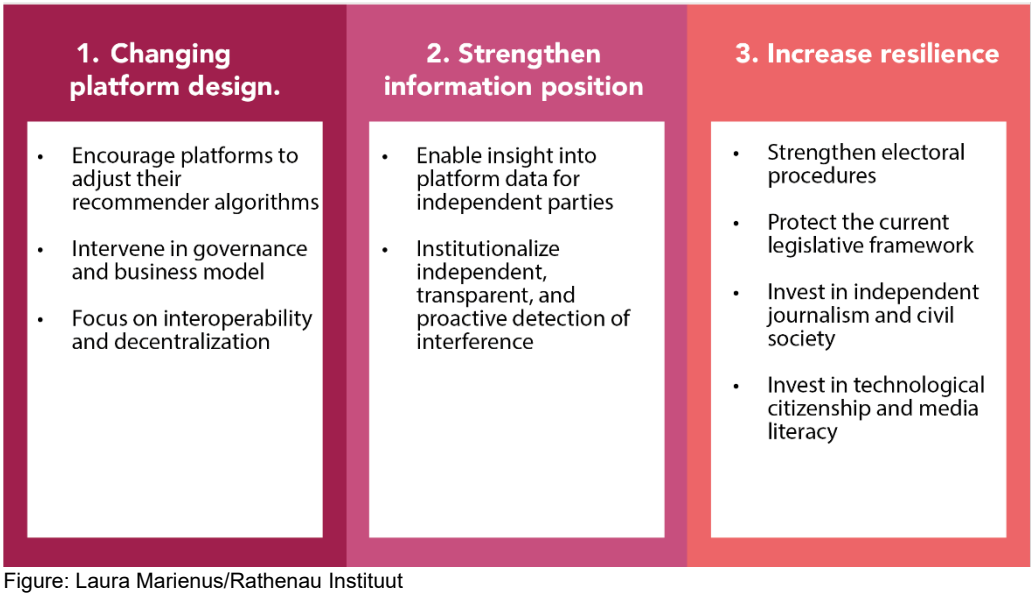


Figure: Laura Marienus/Rathenau Instituut

6.2.1 Changing platform design

The design choices that platforms make regarding the structure of their recommendation algorithms, content moderation and platform architecture have a significant influence on what people see and can share online. Platforms base these design choices on factors such as their vision, governance and revenue

model (Rathenau Instituut, 2025a). The way in which recommendation algorithms are designed provides malicious actors with a toolkit to interfere in our online public debate. Governments may therefore consider encouraging or obliging platforms to adapt their platform design. The 2026–2030 coalition agreement appears to be steering towards such changes, as it states: ‘Addictive, polarising and anti-democratic algorithms will be banned’ (D66, VVD and CDA, 2026).

When considering possible courses of action, we distinguish between intervening in recommendation algorithms and intervening in governance and revenue models.

Stimulate platforms to adapt recommendation algorithms
Governments and regulators can ask social media platforms about the measures they are taking to mitigate the risks arising from their recommendation systems. The risk reports that major platforms are required to produce under the DSA provide a starting point for this.

We have drawn up a table of potential interventions in recommendation algorithms that platforms could implement (see Table 7 below). Some of these interventions may already have been partially implemented by platforms. Others have only been tested in scientific experiments or suggested by experts.

The table shows that there is still a great deal that can be changed in recommendation algorithms, with potential positive effects for users, public debate and reducing the risk of interference in elections.

Parliament could ask platforms to state, for each of these interventions, whether they have considered and tested the experts’ suggestions and what their reasons are for not implementing the changes. Insisting on or mandating changes to increase user control could also be included in future legislation, such as the European Digital Fairness Act, which is currently under negotiation.

Table 7 Research-based suggestions for potential adjustments to recommendation algorithms

Adjustment	What is it?	Examples
Ranking based on bridging groups of people (Lasser & Poechhacker, 2025)	Platforms design their recommendation algorithms with different objectives, such as increasing trust between groups of people, rather than the goal of maximising engagement	Recent scientific research into *bridging-based ranking.*
Circuit breakers	Additional checks before information can go viral.	Snap states that content is first reviewed by human moderators

Adjustment	What is it?	Examples
(Snap, 2025b) (Pathak & Spezzano, 2024)		before it is recommended to a wider audience.
Quality scores (Cunningham, 2023)	Content of a lower quality, such as images generated by generative AI, is given less weight by the recommendation algorithm.	
Diversification algorithms (Bengani, 2023)	Diversification algorithms can ensure a diverse set of content when filtering and ranking content.	Platforms claim to do this, but it is unclear how much weight diversity carries in their recommendations. Content diversification can take many different forms: by topic, geographical origin, source, media type and popularity.
Giving greater weight to user feedback on content (Cunningham et al., 2025; Stray et al., 2024)	By asking users more explicit questions, for example through user surveys, platforms gain a better understanding of users' preferences. This is in contrast to recommendation algorithms, which currently often rely on inferred preferences, such as likes and viewing time.	When people are shown a lot of similar content, particularly during election periods, platforms could actively ask users whether this content is valuable to them and how it affects their attitudes towards elections.
Downranking of content (Piccardi et al., 2025)	Instead of removing content, platforms could give less weight to borderline or anti-democratic content in their recommendation algorithms.	A recent experiment by researchers shows that ranking *anti-democratic* and *polarising* posts lower on X had a positive effect on how users view people with different political views.
Manually approve all content (including comments) (Ribeiro et al., 2023)	By manually approving comments and content, the number of comments that do not comply with the rules appears to decrease. This can improve the quality of the conversation on social media.	In Facebook groups, moderators have the option to manually approve all comments.
Community notes (Chuai et al., 2024)	By using the contextual information that users provide with a post as a signal for the recommendation algorithm, the spread of misleading content could be reduced. Further research into this is needed.	Research shows mixed results regarding the effect of community notes on the spread of information on X.
Adjusting reach based on account reliability assessment (Fernández et al., 2021; Stray et al., 2024)	In the run-up to and during elections, platforms could adjust the weightings of their recommendation algorithms based on internal investigations into accounts. Accounts with a low	New accounts and accounts that become active in the period immediately before an election could be recommended less frequently during election periods,

Adjustment	What is it?	Examples
	trust score could be recommended less frequently.	or only after a human check on the type of content being shared.
Informing users after interaction with content (Truong et al., 2024)	Platforms can warn users who frequently engage with accounts that share low-quality or removed content, and help them recognise misleading content.	

Table compiled by the Rathenau Instituut

When making design changes, a balance often needs to be struck between the pros and cons. Changes can also affect users' fundamental rights, or they may harm the economic interests of platforms (Lubin et al., 2024).

Take, for example, limiting the visibility of new accounts. This can be effective in making it harder to deliberately mislead people using a network of fake accounts in the run-up to elections. The downside is that *all* new accounts receive less reach.

Or consider steering towards recommendation algorithms that are less based on engagement, meaning hyperactive accounts get less reach, but users also spend less time on social media platforms.

It is therefore necessary to be able to monitor the effectiveness and impact of measures, so that they can be adjusted where necessary. Fortunately, social media platforms lend themselves well to this form of research-based regulation, as they potentially offer effective ways of testing the impact of measures (Lubin et al., 2024).

It is unclear how the proposed changes outlined below would play out and whether the benefits would outweigh any potential negative side effects. This is because researchers are currently unable to scrutinise platforms. This also highlights the need to strengthen the information position of researchers, amongst others, vis-à-vis platforms (section 6.2.2).

Intervene in governance and revenue models

The scale on which today's online platforms operate is enormous. Consequently, the ownership structure and governance (how an online environment is managed) of major online platforms mean that decisions made by a small group of companies and their owners have a significant influence on what people worldwide see and are able to share online (Rathenau Instituut, 2025a). They have the power to determine which types of content are relatively amplified compared to other content, and their design choices create opportunities or limitations for interference on their platforms.

In the early days of the internet, the power to shape online environments was distributed differently than it is today. In his book The Modem World: A Prehistory of Social Media, Kevin Driscoll describes, for example, how bulletin board systems differed in a number of ways from how we use the internet today (Driscoll, 2022). Online environments were often local, with significant influence from the communities that used them.

Nowadays, large platform companies receive by far the lion's share of advertising revenue, which is also used to fund journalism. As a result, the range of journalistic content is dwindling (Digital News Report Netherlands 2025, 2025, p. 40). This is despite the fact that democracy actually benefits from a diverse and free flow of information.

A radical approach would be to intervene in the advertising-driven revenue model of social media platforms. That revenue model is a key reason for the rise of engagement-based recommendation algorithms. The longer people stay on a platform, the more advertising revenue is generated. Intervening in the revenue model naturally has its drawbacks too, such as an increase in costs that are passed on to users. Nevertheless, this trade-off should be on the table. The Digital Fairness Act offers starting points for this.[21]

Focus on decentralisation and interoperability

The government can stimulate a diverse and decentralised landscape of online environments. In decentralised environments, control no longer lies with a single party, but is in the hands of various (smaller) parties. This gives users greater control and autonomy. The report Inclusive Online (Rathenau Instituut, 2025a, p. 101) explores in greater depth the promotion of alternatives to dominant social media platforms.

Alternative platforms offer no guarantee against interference. However, a pluralistic and decentralised landscape of online environments does strengthen the government's room for negotiation vis-à-vis the current dominant platforms. Alternatives also increase freedom of choice for users. Furthermore, they reduce the impact of one or a few platforms.

Social media platforms currently control the entire process of distributing and viewing content on their platforms. What if multiple parties could be involved in this process? People could then choose different recommendation algorithms from

21 See also the Rathenau Instituut's submission of 23 October 2025 to the European Commission regarding the Digital Fairness Act: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14622-Digital-Fairness-Act/F33095380_en

other parties. Or different forms of content moderation. Scholars refer to this idea as middleware: breaking down social media platforms so that different parties can also play a role in the governance of platforms (Hogg & DiResta, 2024).

By allowing other parties to offer recommendation algorithms as well, users of Instagram, TikTok or Snap, for example, could choose to have their feed curated by a national newspaper, their local sports club or another group of people they trust. These groups could then set their own parameters: what type of content should be prioritised?

The same may apply to content moderation, which is currently carried out centrally by large platforms. Content moderation could also be carried out more by the communities themselves. This could enhance the legitimacy and quality of decisions regarding content (Zuckerman, 2023). The introduction of middleware could also generate new revenue models for parties outside the large centralised platforms.

It is recommended that, at European level, an investigation be carried out into how and whether middleware can provide better protection against attempts at interference via social media. Various organisations, including the economic think tank Bruegel, recommend extending interoperability requirements to make it easier for citizens to switch providers (Scott Morton, 2024).[22]

6.2.2 Strengthening the information position

Our research shows that the information available to regulators, researchers, journalists, politicians and policymakers regarding platforms leaves much to be desired. Parties other than platforms (such as the government, the academic community or the media) are unable to determine with sufficient independence whether interference occurred during the elections. The more knowledge available to external experts, the better and more promptly interference can be detected. After all, such public oversight is also essential because the definition of interference can easily be misused for political purposes (see Chapter 2).

Under the heading ‘strengthening the information position’, we provide insight into the type of information that external experts on platforms need to gain a grasp of interference activities, and we propose an independent detection body that can itself proactively tackle interference activities.

22 Specifically, consideration could be given to broadening Article 7 of the Digital Markets Regulation to cover other online environments in the revision.

Enable access to platform data for government, regulators and researchers
We need a better understanding of how content appears in our news feeds. This includes (clarification of) content moderation policies, the quality scores assigned to content, the algorithms used to rank content, and the weightings that platforms apply in their recommendation algorithms (see Chapter 3). With the help of this type of information, external parties can better monitor the measures that platforms actually take and the extent to which these are effective. The same applies to the content posted on platforms. Platforms, together with content creators, also play a role in making such information transparent. Various measures can improve the information position of policymakers, politicians, regulators, researchers and journalists.

Under the DSA, platforms must make their content accessible to regulators and independent researchers. The level of access to systems varies, with regulators being granted the greatest access. Experts indicate that the information platforms have shared so far is still too limited to gain a proper understanding of the workings of recommendation algorithms and interference activities (see Chapter 5).

Below, we set out proposals that could enhance the information available to third parties and contribute to the achievement of the objectives of various provisions of the DSA, such as Articles 34, 35, 37 and 40.

The coalition agreement may also provide starting points for requiring platforms to demonstrate that their platforms do not contribute to interference in elections, the dissemination of misleading content, or the sowing of doubt about democratic processes.

Real-time spot checks
Understanding the role and functioning of recommendation algorithms helps to gain a clear picture of systemic risks. This can be achieved by publishing real-time samples of the most widely disseminated public content and the content that generates the most engagement. Insight can also be enhanced by conducting real-time representative samples of the public content consumed during a typical user session on the platform at a random moment (Cooper & Chapman, 2025).

Content dissemination
Understanding the number of people and the types of groups to which content that has already been removed has reached can be helpful.

Design of recommendation algorithms

Insight into the technical specifications of the architecture of platforms' recommendation algorithms (*recommender systems*) is important (Panoptykon et al., 2024). Consider:

a. What is the type of algorithm and what are the hyperparameters? What is the size of the neural network used?
b. What is the input data and what is the relative importance of different types of input data? For linear regressions: what are the weights?
c. How is the input data presented, and what is the logic behind the embedding spaces used?
d. What is the loss function of the models used?
e. Is training data labelled by humans? Which parties are involved in this labelling process and in what way?
f. What efforts are being made to build machine learning systems that are understandable to humans? What are the results of the algorithmic interpretation tools developed?

Furthermore, to gain a proper understanding of the precise functioning of recommendation algorithms on specific platforms, it is necessary to understand the 'trade-offs' that platforms make in the design of their recommendation algorithms, including the data they use as input and the weights they assign in their recommendation systems (KGI Expert Working Group on Recommender Systems, 2025). Insight into the metrics that platforms use to evaluate the design of their recommendation systems would also be helpful (KGI Expert Working Group on Recommender Systems, 2025).

Results of platforms' internal A/B tests

Insight into the effects of (constant) changes to the design of recommendation systems can be gained through A/B testing. Many platforms use a holdout group for A/B testing, a control group of users who are not exposed to the changes. Long-term holdout experiments provide significant insight into the effect of design choices on users and could be published in aggregated form. A requirement to do so could also incentivise platforms to better align their design with users' needs (Cooper & Chapman, 2025).

Efforts during election periods

Insight into accessible information about platform efforts during election periods, for example through pre- and post-electoral evaluation, can help strengthen the information position.

Standardisation of access and reporting
Finally, it is important to focus on standardising access for researchers and regulators, as well as standardising the identification, mitigation and reporting of systemic risks. Effective risk mitigation requires transparent measurement criteria and evaluation methods, supported by data-driven evidence that enables outsiders to assess the risks and effectiveness of measures over time. Under the DSA, regulators can hold the very large platforms and search engines to account for the quality of their risk analyses and promote a culture of continuous learning.

Institutionalise independent, transparent and proactive detection of interference
Interference requires greater institutional embedding of the detection and combating of interference activities. The government has an election flagger status for reporting content that could harm the elections. However, there is also a need for greater government oversight of interference activities on social media. It is important that there are clear safeguards against political influence and censorship.

We propose two courses of action in this regard. The first suggestion is that the government could be given greater powers to flag content. The government holds election flagger status with the major platforms (see Chapter 2), which it deliberately exercises with restraint. However, content may only be flagged on the recommendation of third parties, not on the basis of its own investigations. It is worth considering giving the government more scope to proactively flag content that undermines the elections. It is important to strike the right balance between expanding the scope for action and safeguarding freedom of expression. This is in line with our Report to Parliament: *Combating Disinformation, Avoiding Censorship* (Rathenau Instituut, 2020).

The second suggestion is that efforts should be made to establish an independent body capable of identifying interference. This body should work closely with independent experts such as disinformation specialists, platform researchers, computer scientists, cybersecurity experts and legal experts. At the time of writing this report, the Ministry of the Interior and Kingdom Relations is conducting a feasibility study into the establishment of a disinformation detection body. This could fulfil this role.

Interference actors are resourceful. Interference activities are constantly evolving and capitalise on new developments, such as the rise of chatbots, the personalisation of news, and micro-payments to influencers via live streams that are virtually impossible to moderate. Ongoing research and monitoring are therefore essential.

There is a heavy reliance on academics and NGOs for the monitoring and detection of interference activities on social media platforms. Consideration should be given to providing structural funding to such organisations to enable them to carry out this research systematically.

6.2.3 Increasing resilience

The courses of action outlined above focus, on the one hand, on platforms and, on the other, on improving the information available regarding interference via platforms. These courses of action are heavily dependent on enforcement and the cooperation of social media companies. In the context of interference via platforms, it is also worthwhile to consider interventions that do not depend on social media platforms and that enhance the resilience of society as a whole.

Below, we explore various ways of making society more resilient to interference.

Strengthen electoral procedures
Attempts at interference can lead to disputes over election results. This makes it essential to ensure that elections are conducted fairly and are not vulnerable to political influence.

Therefore, take measures to strengthen independence in decisions regarding electoral disputes. Ensure that there are clear criteria for proceeding with recounts and re-votes. Examine whether the timeframe for identifying irregularities during elections can be extended, as proving interference via social media takes time. Systematically assess the extent to which electoral procedures are adequately equipped to deal with interference activities and adjust them where necessary.

Safeguard current legislation and clarify where necessary
Clear standards, consistent application of existing legislation and effective oversight are necessary to counter interference activities.

It is striking that AI systems used to influence the outcome of an election, referendum or voting behaviour are not prohibited, but fall into the high-risk category. It is currently unclear exactly what is meant by AI systems. Ensure that regulators have sufficient capacity to clarify which AI systems fall under the prohibited category and which systems fall under the high-risk category of the AI Regulation. It could also be investigated to what extent such systems are compatible with the GDPR and the VPR.

Support researchers, civil society organisations and supervisory authorities with examples and case studies, for instance on how AI systems can influence voting behaviour and cause social harm, to underpin the guidelines currently being drawn up. Develop and test templates for risk assessments of the use of AI in elections.

In addition, advocate for, or explore the possibilities of, a moratorium on the use of AI systems in election campaigns to gain a better understanding of the societal and political impact and to facilitate research.

Also strengthen supervision of the AI Regulation, DSA, GDPR and VPR with a view to preventing interference. Focus on the rapid implementation of the VPR in the Netherlands and ensure that platforms adhere to their own *community guidelines* regarding paid political content. In doing so, seek cooperation with other European regulators.

Invest in independent journalism

Invest in a diverse media landscape, including public service broadcasting, to prevent the erosion of independent news provision. Independent journalism serves as a watchdog in society and is one of the checks and balances in a democracy. Journalism and investigative reporting operate independently of government and the business sector and can highlight abuses in society. Independent investigative reporting thus plays a vital role in the information ecosystem.

Invest in technological citizenship and media literacy

The DSA has improved users' position vis-à-vis major social media platforms. However, it remains to be seen whether users are aware of this. Digital rights include, among other things, users' ability to report content that breaches platform rules, as well as the right to know what happens to a report. The government could help raise public awareness of their digital rights.[23]

If users of social media platforms realise the potential for interference, such interference might become less effective. People may recognise these interference activities. However, warnings about interference could also lead people to distrust all information, or cause them to have less confidence in the fairness of elections (Altay et al., 2025).

Teaching people how an open democracy works can help to strengthen democracy. How does the Netherlands ensure that elections are conducted fairly? What safeguards do newsrooms have in place? And where can accurate information about elections be found? And how do citizens form their political judgements?

23 See, for example, this campaign by Bits of Freedom to raise awareness of digital rights: https://www.jouwplatformrechten.nl/

These questions can provide topics for discussion centred on the functioning of a democratic society. In doing so, they can increase resilience against interference.

Bibliography

404 Media. (30 September 2025). *Google Just Removed Seven Years of Political Advertising History from 27 Countries*. 404 Media. https://www.404media.co/google-ad-transparency-european-union

10767307 CV FORM 23-13934, ECLI:NL:RBAMS:2024:3980 (Amsterdam District Court, 5 July 2024). https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBAMS:2024:3980

AIVD. (2025). *Annual Report 2024*. https://www.aivd.nl/documenten/jaarverslagen/2025/04/24/jaarverslag-2024

All major media outlets in Norway fall for Tweets from Russian trolls. (2020, 3 March). https://nos.nl/artikel/2325674-alle-grote-media-in-noorwegen-trappen-in-tweets-van-russische-trollen

Altay, S., Hoes, E., & Wojcieszak, M. (2025). Following news on social media boosts knowledge, belief accuracy and trust. *Nature Human Behaviour*, *9*(9), 1833–1842. https://doi.org/10.1038/s41562-025-02205-6

Dutch Data Protection Authority. (2024, 3 September). *AP fines Clearview for illegal data collection for facial recognition*. https://www.autoriteitpersoonsgegevens.nl/actueel/ap-legt-clearview-boete-op-voor-illegale-dataverzameling-voor-gezichtsherkenning

Dutch Data Protection Authority. (2025, 21 October). *DPA warns: chatbots provide biased voting advice*. https://www.autoriteitpersoonsgegevens.nl/actueel/ap-waarschuwt-chatbots-geven-vertekend-stemadvies

Bandy, J., & Lazovich, T. (2023). Exposure to Marginally Abusive Content on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *17*, 24–33. https://doi.org/10.1609/icwsm.v17i1.22123

Beemsterboer, T. B. R. P. T., & Beemsterboer, T. B. R. P. T. (2026, 17 February). Israel sees TikTok as the most important geopolitical battleground of our time. *NRC*. https://www.nrc.nl/nieuws/2026/02/17/israel-ziet-tiktok-als-het-belangrijkste-geopolitieke-strijdperk-van-deze-tijd-a4919546

Bellogín, A., Castells, P., & Cantador, I. (2017). Statistical biases in Information Retrieval metrics for recommender systems. *Information Retrieval Journal*, *20*(6), 606–634. https://doi.org/10.1007/s10791-017-9312-z

Bengani, P. (2023, 17 November). What is Media Diversity and Do Recommender Systems Have It? *Understanding Recommenders*. https://medium.com/understanding-recommenders/what-is-media-diversity-and-do-recommender-systems-have-it-b2c86be9f08f

Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Berzina, K., & Soula, E. (2020). *Conceptualising Foreign Interference in Europe*. Alliance for Securing Democracy.

Bouchaud, P. (2024). Algorithmic Amplification of Politics and Engagement Maximization on Social Media. In H. Cherifi, L. M. Rocha, C. Cherifi, & M. Donduran (Eds.), *Complex Networks & Their Applications XII* (pp. 131–142). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53503-1_11

Brady, W. J., Jackson, J. C., Lindström, B., & Crockett, M. J. (2023). Algorithm-mediated social learning in online social networks. *Trends in Cognitive Sciences*, *27*(10), 947–960. https://doi.org/10.1016/j.tics.2023.06.008

Bruns, A. (2019). Filter bubble. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1426

Bunskoek, J., van den Berg, E., & Stoker, H. (25 November 2025). *Foreign troll armies amplified political and inflammatory messages surrounding the elections*. RTL.nl.

Chuai, Y., Tian, H., Pröllochs, N., & Lenzini, G. (2024). Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter? *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW2), 1–52.

Cooper, A., & Chapman, P. (2025, 19 May). *Making Recommender Systems Work for People: Turning the DSA's Potential into Practice - DSA Observatory*. https://dsa-observatory.eu/2025/05/19/making-recommender-systems-work-for-people/

Cunningham, T. (2023, 8 May). *Ranking by Engagement | Tom Cunningham – Tom Cunningham*. https://tecunningham.github.io/posts/2023-04-28-ranking-by-engagement.html
Cunningham, T., Pandey, S., Sigerson, L., Stray, J., Allen, J., Barrilleaux, B., Iyer, R., Kothari, M., Rezaei, B., Kairam, S., & Milli, S. (2025). Ranking by engagement and non-engagement signals: Learnings from industry. *Annals of the New York Academy of Sciences*, *1551*(1), 19–32. https://doi.org/10.1111/nyas.15399

D66, VVD and CDA. (2026). *Getting started. Building a better Netherlands*. https://www.kabinetsformatie2025.nl/documenten/2026/01/30/aan-de-slag---coalitieakkoord-2026-2030

Data School, Utrecht University. (2023). *Playing with Fire – How the Interaction Between the House of Representatives and Social Media Fuels Anger.* Utrecht University. https://dataschool.nl/2023/10/06/spelen-met-vuur/

Digital News Report Netherlands 2025 (Digital News Report, p. 52). (2025). Media Authority. https://www.cvdm.nl/wp-content/uploads/2025/06/Digital-News-Report-2025-1.pdf

Directorate-General for Research and Innovation. (2022). *Tackling R&I foreign interference: staff working document*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/513746

DISARMFoundation. (n.d.). *01_AMITT_TTP_Guide.pdf*. GitHub; DISARMFoundation. Retrieved 22 January 2026, from https://github.com/DISARMFoundation/DISARMframeworks/blob/20ac4ea378578e1fecf4c9014f525bfc27a9b66c/generated_pages/techniques/T0104.001.md

Driscoll, K. (2022). *The Modern World: A Prehistory of Social Media* (First Edition). Yale University Press.

Ecker, U., Roozenbeek, J., & Lewandowsky, S. (2024). *Misinformation remains a threat to democracy*.

Egelhofer, J. L., Aaldering, L., Eberl, J.-M., Galyga, S., & Lecheler, S. (2020). From Novelty to Normalisation? How Journalists Use the Term "Fake News" in their Reporting. *Journalism Studies*, *21*(10), 1323–1343. https://doi.org/10.1080/1461670X.2020.1745667

Eijsvoogel, J. (2025, 2 October). Meta causes confusion and anger among government and NGOs with advertising ban on 'social issues'. *NRC*.

https://www.nrc.nl/nieuws/2025/10/02/meta-wekt-verwarring-en-woede-bij-overheid-en-ngos-met-advertentieverbod-voor-maatschappelijke-kwesties-a4908280

EU in contact with Telegram as it nears the threshold for EU tech rules. (28 May 2024). Reuters. https://www.reuters.com/technology/eu-touch-with-telegram-it-nears-criterion-eu-tech-rules-2024-05-28

European Commission. (2023). *Commission Staff Working Document – Executive Summary of the Impact Assessment Report* (Commission Staff Working Document (SWD) SWD/2023/663 final). European Commission. https://eur-lex.europa.eu/eli/impact_assessment/52023SC0663/EN

European Commission. (2024a). *Guidelines on mitigating systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065* ((C/2024/3014)). http://data.europa.eu/eli/C/2024/3014/oj

European Commission. (2024b, 30 April). *Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act*. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

European Commission. (2024c, 16 May). *Commission opens formal proceedings against Meta*. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664

European Commission. (2024d, 5 August). *TikTok commits to permanently withdrawing TikTok Lite Rewards*. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4161

European Commission. (2025a, 17 January). *Commission addresses additional investigatory measures to X in the ongoing proceedings under the Digital Services Act*. https://digital-strategy.ec.europa.eu/en/news/commission-addresses-additional-investigatory-measures-x-ongoing-proceedings-under-digital-services

European Commission. (2025b). *Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)* (C(2025) 5052 final). European Commission. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

European Commission. (2025c, 24 October). *Commission preliminarily finds TikTok and Meta in breach of their transparency obligations* [Text]. https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2503

European Commission. (2025d, 5 December). *Commission fines X €120 million under the Digital Services Act*. https://digital-strategy.ec.europa.eu/en/news/commission-fines-x-eu120-million-under-digital-services-act

European Commission. (2025, 17 December). *Commission opens formal proceedings against TikTok under the DSA* [European Commission]. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_6487

European Commission. (2026, 6 February). *Commission preliminarily finds TikTok's addictive design in breach of the Digital Services Act*. https://ec.europa.eu/commission/presscorner/detail/en/ip_26_312

European External Action Service. (2023). *1st EEAS Report on Foreign Information Manipulation and Interference Threats. Towards a Framework of Networked Defence*. European External Action Service. https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-DataTeam-ThreatReport-2023..pdf

European Parliamentary Research Service. (2025). *TikTok and EU regulation: Legal challenges and cross-jurisdictional insights*.

FameUp – FameUp – Over 500+ nano and micro influencers activated in 24 hours. (n.d.). FameUp. Retrieved 4 October 2025, from https://fameup.com/en/

Feenstra, W., & Sabel, P. (2025, 27 October). *Two PVV MPs anonymously attack Timmermans with fake images; GroenLinks-PvdA files a complaint*. de Volkskrant. https://www.volkskrant.nl/politiek/twee-pvv-kamerleden-vallen-met-nepbeelden-anoniem-timmermans-aan-groenlinks-pvda-doet-aangifte~b6958ea1/

Feifer, J. (26 June 2023). LinkedIn Changed Its Algorithms — Here's How Your Posts Will Get More Attention Now. *Entrepreneur*. https://www.entrepreneur.com/science-technology/linkedin-changed-its-algorithms-heres-how-your-posts/454728

Fernández, M., Bellogín, A., & Cantador, I. (2021). *Analysing the Effect of Recommendation Algorithms on the Amplification of Misinformation* (arXiv:2103.14748). arXiv. https://doi.org/10.48550/arXiv.2103.14748

Garnett, H. A., James, T. S., & Caal-Lam, S. (2025). *Electoral Integrity Global Report 2025* (PEI 11.0). Electoral Integrity Project. https://ueaeprints.uea.ac.uk/id/eprint/99835/1/Year_in_Elections_PEI_11_Report_FINAL.pdf

Gauthier, G., Hodler, R., Widmer, P., & Zhuravskaya, E. (2026). The political effects of X's feed algorithm. *Nature*. https://doi.org/10.1038/s41586-026-10098-2

Gillespie, T. (2022). Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society*, *8*(3), 20563051221117552. https://doi.org/10.1177/20563051221117552

Gleicher, N. (2019, 6 May). Removing More Coordinated Inauthentic Behaviour From Russia. *Meta Newsroom*.

GLOBSEC, Debunk.org, Institute for Strategic Dialogue, DEN Institute, & Alliance4Europe. (2025). *Election report. Assessment of Foreign Information Manipulation and Interference in the 2025 Czech Parliamentary Election*. https://fimi-isac.org/wp-content/uploads/2025/12/FRT-24_Globsec_Czech-Election-Report_12122025.pdf

Goanta, C. (2024, 17 November). Influencers & elections: An Eastern European campaign blueprint [Research blog]. *Human Ads ERC*. https://humanads.eu/influencers-and-elections-an-eastern-european-campaign-blueprint

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The Structural Virality of Online Diffusion. *Management Science*, *62*(1), 180–196. https://doi.org/10.1287/mnsc.2015.2158

Google. (10 August 2012). *YouTube Now: Why We Focus on Watch Time*. Blog.YouTube. https://blog.youtube/news-and-events/youtube-now-why-we-focus-on-watch-time/

Google. (2025a). *Information about Monthly Active Recipients under the Digital Services Act (EU)*. https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-24_2025-1-1_2025-6-30_en_v1.pdf

Google. (2025b). *EU Digital Services Act (EU DSA) Biannual VLOSE/VLOP Transparency Report*. https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-27_2025-1-1_2025-6-30_en_v1.pdf

Google. (2025c). *Search quality evaluator guidelines*. https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

Google to stop political adverts from September. (2025, 6 August). Tweakers. https://tweakers.net/nieuws/237800/google-stopt-vanaf-september-met-politieke-advertenties.html

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behaviour in an election campaign? *Science*, *381*(6656), 398–404. https://doi.org/10.1126/science.abp9364

Charter of the United Nations, (1945). https://wetten.overheid.nl/BWBV0004143/1973-09-24

Haugen, F. [producer]. (2022). *FBarchive Document odoc2025266473w32 (Green Edition)*. Cambridge, MA: FBarchive [distributor]. https://fbarchive.org/user/doc/odoc2025266473w32

Heck, W., & Kouwenhoven, A. (2023, 7 July). This shadowy company destroyed the reputations of European Muslims. *NRC*. https://www.nrc.nl/nieuws/2023/07/07/dit-schimmige-bedrijf-vernietigde-met-succes-de-reputaties-van-europese-moslims-a4169074

HEIO Consortium, Post-X Society, AI Forensics, Trollrensics, University of Amsterdam, & Justice for Prosperity. (2026). *Hybrid Election Integrity Observatory: Final Report – Monitoring the Dutch Parliamentary Elections, 29 October 2025*. HEIO Consortium. https://www.heio.nl/wp-content/uploads/2026/01/260115-HEIO-Final-Report.pdf

Hendrix, J. (7 January 2025). *Transcript: Mark Zuckerberg Announces Major Changes to Meta's Content Moderation Policies and Operations*. Tech Policy Press. https://techpolicy.press/transcript-mark-zuckerberg-announces-major-changes-to-metas-content-moderation-policies-and-operations

Hoboken, Joris van, Naomi Appelman, Anna van Duin, et al. 2020. *WODC study: Provision for requests for the rapid removal of unlawful online content*. Institute for Information Law, University of Amsterdam. https://pure.uva.nl/ws/files/87684586/3108_volledige_tekst_tcm28_464976.pdf.

Hogg, L., & DiResta, R. (2024). *Shaping the Future of Social Media with Middleware*. Georgetown University. https://doi.org/10.57928/NENS-7F25

Honingh, M., & van Ham, C. (2024). *Exploration and in-depth analysis of democratic erosion and response in the Netherlands* (Sustainable democracy). Ministry of the Interior and Kingdom Relations. https://www.kennisopenbaarbestuur.nl/site/binaries/site-content/collections/documents/2024/04/12/exploration-and-in-depth-analysis-of-democratic-erosion-and-response-in-the-netherlands/Exploration+and+in-depth+analysis+of+democratic+erosion+in+the+Netherlands+-+April++2024+final.pdf

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. *The International Journal of Press/Politics*, *25*(3), 493–516. https://doi.org/10.1177/1940161219900126

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022a). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, *119*(1), e2025334119. https://doi.org/10.1073/pnas.2025334119

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022b). Algorithmic Amplification of Politics on Twitter. *Proceedings of the National Academy of Sciences*, *119*(1), e2025334119. https://doi.org/10.1073/pnas.2025334119

Irwin, G. A., & van Holsteyn, J. J. M. (2021). Keeping Our Feet Dry: Impediments to Foreign Interference in Elections in the Netherlands. *Election Law Journal: Rules, Politics, and Policy*, *20*(1), 54–69. https://doi.org/10.1089/elj.2020.0654

Jiang, J. A., Nie, P., Brubaker, J. R., & Fiesler, C. (2023). A Trade-off-centred Framework of Content Moderation. *ACM Trans. Comput.-Hum. Interact.*, *30*(1), 3:1–3:34. https://doi.org/10.1145/3534929

Justice for Prosperity. (2025). *From botnet to public perception – Attempts to influence the 2025 Dutch elections* (p. 26). Justice for Prosperity. https://justiceforprosperity.org/wp-content/uploads/2026/01/Van-botnet-tot-beeldvorming-JfP-extern-1.pdf

Parliamentary Papers II, 35 165, no. 102. (9 January 2026). Overheid.nl. https://zoek.officielebekendmakingen.nl/kst-35165-102.html

Parliamentary Papers II, 36 742, no. 8. (20 July 2025). Overheid.nl. https://zoek.officielebekendmakingen.nl/dossier/36742

Parliamentary Papers II, 22 112, no. 4223. (2025, 12 December). https://zoek.officielebekendmakingen.nl/kst-22112-4223.html

Parliamentary Papers II, 26 643, no. 1391. (4 September 2025). Overheid.nl. https://zoek.officielebekendmakingen.nl/kst-26643-1391.html#ID-1211802-d36e130

Parliamentary Papers II, 26 643, no. 1400. (24 September 2025). Overheid.nl. https://www.tweedekamer.nl/kamerstukken/kamervragen/detail?id=2025D42593&did=2025D42593

Parliamentary Papers II, 30 821, no. 230. (2024). Overheid.nl. https://zoek.officielebekendmakingen.nl/kst-30821-230

Parliamentary Papers II, 36 552, no. 12. (2025, 4 July). Overheid.nl. https://zoek.officielebekendmakingen.nl/kst-36552-12.html

KGI Expert Working Group on Recommender Systems. (2025). *Better Feeds: Algorithms That Put People First – A How-To Guide for Platforms and Policymakers*. https://kgi.georgetown.edu/wp-content/uploads/2025/02/Better-Feeds_-Algorithms-That-Put-People-First.pdf

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2024). The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimisation. *Management Science*, *70*(9), 6336–6355. https://doi.org/10.1287/mnsc.2022.03683

Knight Georgetown Institute. (2025). *Taxonomy of User Signals*.

Koebler, J. (2025, 24 November). *America's Polarisation Has Become the World's Side Hustle*. 404 Media. https://www.404media.co/americas-polarization-has-become-the-worlds-side-hustle/

Kruschinski, S., & Votta, F. (2025). *Campaign Tracker Dashboard*. Campaign Tracker. https://www.campaigntracker.nl/

Lasser, J., & Poechhacker, N. (2025). Designing social media content recommendation algorithms for societal good. *Annals of the New York Academy of Sciences*, *1548*(1), 20–28. https://doi.org/10.1111/nyas.15359

like.vn. (n.d.). *Frequently Asked Questions (FAQ) - like.vn*. Like.vn 💔 - SMM Panel Services - Cheapest API Provider - Vietnam. Retrieved 3 February 2026, from https://like.vn/cau-hoi-thuong-gap

like.vn. (3 February 2026). *Buy cheap Instagram likes ❤Cheap IG like boosting and buffing services (Real Likes)*. Like.vn 💔 - SMM Panel Services - Cheapest API Provider - Vietnam. https://like.vn/tang-like-instagram

Lin, Y., Liu, Y., Lin, F., Zou, L., Wu, P., Zeng, W., Chen, H., & Miao, C. (2024). A Survey on Reinforcement Learning for Recommender Systems. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(10), 13164–13184. https://doi.org/10.1109/TNNLS.2023.3280161

LinkedIn. (2025). *Transparency Report on the Digital Services Act – August 2025*. https://content.linkedin.com/content/dam/help/tns/en/August-2025-Digital-Services-Act-Transparency-Report.pdf

Lubin, N., Mayberry, K., Moses, D., Revel, M., Thorburn, L., & West, A. (2024). Mapping the space of social media regulation. *MIT Science Policy Review*, *5*, 108–133. https://doi.org/10.38105/spr.sqyw1u2jf7

Macdonald, S., & Vaughan, K. (2024). Moderating borderline content while respecting fundamental values. *Policy & Internet*, *16*(2), 347–361. https://doi.org/10.1002/poi3.376

McKenzie, H., & Monga, S. (2024, 22 February). Upgrading Substack's recommendation network [Substack newsletter]. *On Substack*. https://on.substack.com/p/substacks-recommendations-network

McWhinney, E. (1966). *Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of their Independence and Sovereignty*. https://www.cambridge.org/core/product/identifier/S0002930000219969/type/journal_article

Mercadante, E. J., Tracy, J. L., & Götz, F. M. (2023). Greed communication predicts the approval and reach of US senators' Tweets. *Proceedings of the National Academy of Sciences*, *120*(11), e2218680120. https://doi.org/10.1073/pnas.2218680120

Meta. (n.d.). *Community Standards | Transparency Center*. Retrieved 15 January 2026, from https://transparency.meta.com/policies/community-standards

Meta. (2025a). *Adversarial Threat Report Q2/Q3* (p. 45). Meta. https://dn720004.ca.archive.org/0/items/a-blueprint-for-content-governance-and-enforcement-facebook/A%20Blueprint%20for%20Content%20Governance%20and%20Enforcement%20_%20Facebook.pdf

Meta. (2025b). *Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Facebook*.

Meta. (2025c). *Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Instagram*.

Meta must (for the time being) offer a chronological timeline as the default option. (2026, 22 January). Mr. Online. https://www.mr-online.nl/meta-moet-voorlopig-chronologische-tijdlijn-als-blijvende-voorkeursoptie-aanbieden/

In the middle of the election campaign, political advertising suddenly disappears from Facebook and Instagram: 'Tech companies are setting the rules of the game in our elections'. (2025, 1 October). EenVandaag. https://eenvandaag.avrotros.nl/artikelen/midden-in-campagnetijd-ineens-geen-politieke-reclame-meer-op-facebook-en-instagram-tech-bepaalt-spelregels-in-onze-verkiezingen-161498

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus*, *4*(3), pgaf062. https://doi.org/10.1093/pnasnexus/pgaf062

Miltenburg, E., Geurkink, B., Tunderman, S., Beekers, D., & den Ridder, J. (2022). *Citizen Perspectives 2022 | report 2*. SCP.

https://www.scp.nl/documenten/2022/12/29/continu-onderzoek-burgerperspectieven---bericht-2-2022
Ministry of the Interior and Kingdom Relations. (2025, 10 November). *Implementation Act on the Regulation on Transparency and Targeted Political Advertising*. https://wetgevingskalender.overheid.nl/Regeling/WGK026963

Mugemangango v. Belgium, No. 310/15 (European Court of Human Rights, 10 July 2022). https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22002-12906%22]}

Narayanan, A. (2022, 15 December). TikTok's Secret Sauce. *Knight First Amendment Institute*. http://knightcolumbia.org/blog/tiktoks-secret-sauce

Narayanan, A. (2023). *Understanding Social Media Recommendation Algorithms*. Knight First Amendment Institute.

Dutch claims foundation launches legal action against TikTok and X (formerly Twitter). (2025, 5 February). https://www.njb.nl/nieuws/nederlandse-claimstichting-begint-rechtszaken-tegen-tiktok-en-x-voorheen-twitter/

O'Carroll, L. (17 January 2025). EU asks X for internal documents about algorithms as it steps up investigation. *The Guardian*. https://www.theguardian.com/technology/2025/jan/17/eu-asks-x-for-internal-documents-about-algorithms-as-it-steps-up-investigation

Ohlin, J. D. (2017). Did Russian Cyber Interference in the 2016 Election Violate International Law? *Texas Law Review*. https://texaslawreview.org/russian-cyber-interference-2016-election-violate-international-law/

Ohlin, J. D., & Hollis, D. B. (Eds.). (2021). *Defending Democracies: Combating Foreign Election Interference in a Digital Age* (1st ed.). Oxford University Press, New York. https://doi.org/10.1093/oso/9780197556979.001.0001

O'Sullivan, D. (6 May 2019). *Be careful who you accept: Russian fake accounts targeted Facebook Groups | CNN Business*. CNN. https://www.cnn.com/2019/05/06/tech/facebook-groups-russia-fake-accounts

Ovadya, A. (2022). *Bridging-Based Ranking. How Platform Recommendation Systems Might Reduce Division and Strengthen Democracy*. https://www.belfercenter.org/sites/default/files/pantheon_files/files/publication/TAPP-Aviv_BridgingBasedRanking_FINAL_220518_0.pdf

Papakyriakopoulos, O., Serrano, J. C. M., & Hegelich, S. (2020a). Political communication on social media: A tale of hyperactive users and bias in recommender systems. *Online Social Networks and Media*, *15*, 100058. https://doi.org/10.1016/j.osnem.2019.100058

Papakyriakopoulos, O., Serrano, J. C. M., & Hegelich, S. (2020b). Political communication on social media: A tale of hyperactive users and bias in recommender systems. *Online Social Networks and Media*, *15*, 100058. https://doi.org/10.1016/j.osnem.2019.100058

Pariser, E. (2012). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Books.

Pathak, R., & Spezzano, F. (2024). An empirical analysis of the effectiveness of intervention strategies for countering the amplification of misinformation by recommendation algorithms. *European Conference on Information Retrieval*, 285–301.

Piccardi, T., Saveski, M., Jia, C., Hancock, J., Tsai, J. L., & Bernstein, M. S. (2025). Reranking partisan animosity in algorithmic social media feeds alters affective polarisation. *Science*, *390*(6776), eadu5584. https://doi.org/10.1126/science.adu5584

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus), (2025). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025PC0837

Quin, F., Weyns, D., Galster, M., & Silva, C. C. (2024). A/B testing: A systematic literature review. *Journal of Systems and Software*, *211*, 112011. https://doi.org/10.1016/j.jss.2024.112011

Council of State. (2024). *Information on the structure of the dispute resolution system in the electoral process.* (W04.24.00045/I). https://open.overheid.nl/documenten/9b0b1cd5-64bb-4945-8000-2f0dcfa73562/file

Council of State. (2025). *Implementation Act on the Regulation on Transparency and Targeted Political Advertising.* https://www.raadvanstate.nl/adviezen/@152389/w04-25-00196/

Rathenau Instituut. (2020). *Combating disinformation, avoiding censorship*. https://www.rathenau.nl/nl/berichten-aan-het-parlement/desinformatie-bestrijden-censuur-vermijden

Rathenau Instituut. (2021). *Online derailment – An exploration of harmful and immoral behaviour on the internet in the Netherlands*. (authors: Van Huijstee, M., Nieuwenhuizen, W., Sanders, M., Masson, E., & Van Boheemen, P.) https://www.rathenau.nl/nl/digitalisering/online-ontspoord

Rathenau Instituut. (2022). *Weighing up algorithms – An exploration of measures to protect human rights in the implementation of profiling.* (authors: Hamer J., Lemmens, A., & Kool, L.) https://www.rathenau.nl/nl/digitalisering/algoritmes-afwegen

Rathenau Instituut. (2023). *Rathenau Scan Generative AI*. (authors: Kool, L., Hamer, J., Hijstek, B., Van Eeden, Q., & Das, D.) https://www.rathenau.nl/nl/digitalisering/generatieve-ai

Rathenau Instituut. (2025a). *Inclusive Online – Towards the design of apps and environments where people can be free and safe*. (authors: Nieuwenhuizen, W., Nieuwenhuis, T., Van Eeden, Q., & Van Huijstee, M.) https://www.rathenau.nl/nl/digitalisering/naar-een-nieuwe-verhouding-tot-technologiebedrijven/inclusief-online (English translated version: https://www.rathenau.nl/en/digitalisation/reviewing-relationship-technology-companies/inclusive-online-summary)

Rathenau Instituut. (2025b). *The price of free internet – Directions for future online tracking policy.* (authors: Das, D., Wals, F., Hijstek, B., Lagendijk, V., & Kool, L., with the collaboration of Gerritsen, J. & Nieuwenhuizen, W.) https://www.rathenau.nl/nl/digitalisering/naar-een-menswaardige-digitale-technologie/de-prijs-van-gratis-internet

Rathenau Instituut. (2025c). *Behind the power of big tech – Explanatory factors for digital power.* (authors: Van Eeden, Q., Karstens, B., Roolvink, S., & Van Huijstee, M.) https://www.rathenau.nl/nl/digitalisering/naar-een-nieuwe-verhouding-tot-technologiebedrijven/achter-de-macht-van-big-tech

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118. https://doi.org/10.1073/pnas.2024292118

Raza, S., Rahman, M., Kamawal, S., Toroghi, A., Raval, A., Navah, F., & Kazemeini, A. (2026). A comprehensive review of recommender systems: Transitioning from theory to practice. *Computer Science Review*, *59*, 100849. https://doi.org/10.1016/j.cosrev.2025.100849

Ribeiro, M. H., Veselovsky, V., & West, R. (2023). The Amplification Paradox in Recommender Systems. *Proceedings of the International AAAI Conference on Web and Social Media*, *17*, 1138–1142. https://doi.org/10.1609/icwsm.v17i1.22223

Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2022). *Recommender Systems Handbook*. Springer US. https://doi.org/10.1007/978-1-0716-2197-4

Rogers, R., & Righetti, N. (2025). Coordinated inauthentic behaviour on Facebook? A typology of manufactured attention. *Platforms & Society*. https://doi.org/10.1177/29768624251369784

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9*(1), 59. https://doi.org/10.1186/s40537-022-00592-5

Saris, K., & van de Ven, C. (2021, 3 March). *Misogyny as a political weapon*. De Groene Amsterdammer. https://www.groene.nl/artikel/misogynie-als-politiek-wapen

Schiffer, Z. (2023, 15 February). *Yes, Elon Musk created a special system for showing you all his Tweets first*. The Verge. https://www.theverge.com/2023/2/14/23600358/elon-musk-tweets-algorithm-changes-twitter

Schneier, B., & Sanders, N. E. (2025). *Rewiring Democracy: How AI Will Transform Our Politics, Government, and Citizenship*. MIT Press.
It's time for the European Union to rethink personal social networking [Bruegel]. Bruegel. https://www.bruegel.org/policy-brief/its-time-european-union-rethink-personal-social-networking

Snap. (2025a, 29 August). *European Union Transparency | Snapchat Transparency*. https://values.snap.com/privacy/transparency/european-union-h1-2025

Snap. (December 2025b). *Snapchat Moderation, Enforcement, and Appeals | Community Guidelines Explainer*. https://values.snap.com/privacy/transparency/community-guidelines/moderation

Snap. (n.d.). *Content Guidelines for Recommendation Eligibility*. Accessed 15 January 2026.

Stepanov, A. & Gupta, A. (10 February 2021). Political Content in Feeds. *Meta Newsroom*. https://about.fb.com/news/2021/02/reducing-political-content-in-news-feed/

Stray, J., Halevy, A., Assar, P., Hadfield-Menell, D., Boutilier, C., Ashar, A., Bakalar, C., Beattie, L., Ekstrand, M., Leibowicz, C., Moon Sehat, C., Johansen, S., Kerlin, L., Vickrey, D., Singh, S., Vrijenhoek, S., Zhang, A., Andrus, M., Helberger, N., … Vasan, N. (2024). Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ACM Trans. Recomm. Syst.*, *2*(3), 20:1–20:57. https://doi.org/10.1145/3632297

Substack, O. (27 October 2025). Demystifying the feed [Substack newsletter]. *On Substack*. https://on.substack.com/p/demystifying-the-feed

Swart, J. (2021). Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media. *Social Media + Society*, *7*(2), 20563051211008828. https://doi.org/10.1177/20563051211008828

TensorFlow. (2023, 27 May). *Recommending movies: retrieval | TensorFlow Recommenders*. Recommending Movies: Retrieval. https://www.tensorflow.org/recommenders/examples/basic_retrieval

Thiele, D., Milzner, M., Heft, A., Gong, B., & Pfetsch, B. (2025). Attributing coordinated social media manipulation: A theoretical model and typology. *New Media & Society*, 14614448251350100. https://doi.org/10.1177/14614448251350100

Thorburn, L., Stray, J., & Bengani, P. (2023, 7 May). Making Amplification Measurable. *Understanding Recommenders*. https://medium.com/understanding-recommenders/making-amplification-measurable-2be548e5986c

TikTok. (2025). *TikTok DSA Transparency Report (January - June 2025)*. https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H1_2025/TikTok-DSATransparencyReport-January-June-2025.pdf

Tjaden, J., Wolfgram, J., Philipp, A., Weißmann, S., Bobzien, L., Kohler, U., & Verwiebe, R. (2025). *Does the TikTok feed lean right? Exposure to Political Party Content among non-partisan users during regional and federal elections in Germany* (7vdex_v1). SocArXiv. https://doi.org/10.31235/osf.io/7vdex_v1

Trapman, L. (2024a). A new system for the resolution of electoral disputes. *Nederlands Juristenblad*.

Trapman, L. (2024b). *Suffrage, elections and election campaigns*. Radboud University Nijmegen.

Treré, E., & Bonini, T. (2024). Amplification, evasion, hijacking: algorithms as a repertoire for social movements and the struggle for visibility. *Social Movement Studies*, *23*(3), 303–319. https://doi.org/10.1080/14742837.2022.2143345

Truong, B. T., Lou, X., Flammini, A., & Menczer, F. (2024). Quantifying the vulnerabilities of the online public square to adversarial manipulation tactics. *PNAS Nexus*, *3*(7), pgae258. https://doi.org/10.1093/pnasnexus/pgae258

Turning Passion to Profit: WAYS TO MAKE MONEY ON TIKTOK LIVE. (n.d.). Retrieved 13 February 2026, from https://www.tiktok.com/live/creators/en-US/article/turning-passion-to-profit-ways-to-make-money-on-tiktok-live_en-US

Twitter team. (n.d.). *twitter/the-algorithm: Source code for the X Recommendation Algorithm*. GitHub. Retrieved 16 January 2026, from https://github.com/twitter/the-algorithm

Under the hood of a Doppelgänger – Qurium Media Foundation. (2022, 27 September). *Qurium*. https://www.qurium.org/alerts/under-the-hood-of-a-doppelganger/

Vliegenthart, R., Kruikemeier, S., Turkenburg, E., & Hamilton, A. (2024). *Literature review on social media and democracy, with a particular focus on anonymity*.

Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making* (p. 107). https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Scientific Council for Government Policy (WRR). (2024). *Focus on the media. Towards new safeguards for their democratic functions* (No. 111). Scientific Council for Government Policy (WRR). https://www.wrr.nl/publicaties/rapporten/2024/10/03/aandacht-voor-media

X. (n.d.). *DSA Transparency Report – October 2025*. Retrieved 12 February 2026, from https://transparency.x.com/dsa-transparency-report-2025-october.html

X v Russmedia Digital SRL and Inform Media Press SRL, ECLI:EU:C:2025:935 (ECJ 2 December 2025). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62023CJ0492

Ye, J., Luceri, L., & Ferrara, E. (2025). Auditing Political Exposure Bias: Algorithmic Amplification on Twitter/X During the 2024 U.S. Presidential Election. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2349–2362. https://doi.org/10.1145/3715275.3732159

Yi, X., Yang, J., Hong, L., Cheng, D. Z., Heldt, L., Kumthekar, A., Zhao, Z., Wei, L., & Chi, E. (2019). Sampling-bias-corrected neural modelling for large-corpus item recommendations. *Proceedings of the 13th ACM Conference on Recommender Systems*, 269–277. https://doi.org/10.1145/3298689.3346996

Zuckerman, E. (2023, 25 October). Let the community work it out: A throwback to early internet days could fix social media's crisis of legitimacy. *Nieman Lab*. https://www.niemanlab.org/2023/10/let-the-community-work-it-out-a-throwback-to-early-internet-days-could-fix-social-medias-crisis-of-legitimacy/

Appendix 1: Parties consulted

Interviews

- Ministry of the Interior and Kingdom Relations
- Netherlands Authority for Consumers and Markets
- Catalina Goanta (Utrecht University)
- John Albert (University of Amsterdam)
- Snap (Snapchat)

Written interviews

- Meta (Instagram and Facebook)
- TikTok

Expert session

- Robert van der Noordaa (Trollrensics)
- Richard Odekerker (Trollrensics)
- Fabio Votta (University of Amsterdam)
- Jelle Postma (Justice for Property)
- Pieter van Boheemen (Post-X Society)
- Sanne Kruikemeier (Wageningen University & Research)
- Varvara Boboc (University of Amsterdam)
- Eva Hofman (De Groene Amsterdammer)
- Evangelos Kanoulas (University of Amsterdam)
- Harrie Oosterhuis (Radboud University)

Appendix 2: Research Methodology

This report has been written on the basis of a main research question that has been broken down into five sub-questions. Each sub-question has been addressed using different research methods. This appendix specifies the data collection method for each sub-question.

The research question of this report is:
What role do recommendation algorithms on major social media platforms play in election interference?

To answer the main question, sub-questions have been formulated by distilling various concepts from the main question into a sub-question:

1. What is interference via social media platforms in the context of elections?
2. How do recommendation algorithms on major social media platforms work?
3. How can recommendation algorithms be used to interfere in elections?
4. What efforts are already being made to prevent interference via recommendation algorithms?
5. What options for action are available to further prevent and/or tackle interference via recommendation algorithms?

The table below lists the sub-questions we worked with whilst collecting and analysing data. Next to the sub-questions are sub-sub-questions. These are questions that provide a focus for the researchers whilst analysing the data.

Table8 Overview of data collection method

Sub-question	Sub-questions	Data collection method
1. What constitutes interference via social media platforms in the context of elections?	• How is the concept of interference defined in the literature and in policy? What is said about, among other things: the actor (i.e. public/private) and their origin (i.e. foreign/domestic), the instrument (i.e. algorithms, disinformation campaigns), and the objective (elections, undermining the rule of law) • How can significant differences between these be explained? • What risks of interference are mentioned in academic and grey literature regarding social media recommendation algorithms?	• Desk research into academic and grey literature on: the definition of the term 'interference' and examples where interference during elections is discussed, such as the Romanian presidential elections. • Interview with legal experts and policymakers on the definition of interference. • Expert session on the definition of interference. • Desk research into academic and grey literature on: concerns mentioned in academic literature, parliamentary letters, journalistic sources and NGO reports. • Expert session to validate and identify risks and concerns.
2. How do recommendation algorithms on major social media platforms work?	• What is a recommendation algorithm? • What is an engagement-based algorithm? • What are the differences between platforms? • What role do filter bubbles and echo chambers play in the functioning of recommendation algorithms? • What is amplification? • Which content is recommended relatively more often?	• Desk research into academic and grey literature on: the technical and social aspects of recommendation algorithms on social media platforms. • Expert session to validate the functioning of recommendation algorithms. • Interviews with platforms and public sources regarding the functioning of recommendation algorithms.

Sub-question	Sub-questions	Data collection method
3. How can recommendation algorithms be used to interfere in elections?	• What tools and techniques are described as being available to actors to increase visibility on social media for a political party, candidate or issue? • Which of these described tools and techniques are: accepted, not illegal but unethical, or illegal?	• Desk research into academic literature and grey literature on: the role of recommendation algorithms in the dissemination of political content. • Interviews with platforms on the role of recommendation algorithms in virality during campaign periods. • Expert session for validation. • Analysis of legislation by in-house legal counsel.
4. What efforts are already being made to prevent interference via recommendation algorithms?	• Which policy instruments and legislation in the Netherlands and Europe relate to interference, elections and/or recommendation algorithms? • In what ways do these instruments seek to address interference activities via social media? • What efforts do platforms claim to be making? • What do experts think of the effectiveness of these efforts?	• Desk research into academic and grey literature, as well as relevant legislation and regulations concerning, among other things: a. The Digital Services Act and *Guidelines for the mitigation of systemic risks online for elections.* b. The Dutch Code of Conduct on Transparency in Online Political Advertising. c. *Regulation (EU) 2024/900 on the transparency and targeting of political advertising.* d. The Dutch Electoral Act. • Analysis of the above legislation by our regular legal expert. • Expert session
5. What courses of action are available to prevent and/or tackle interference via recommendation algorithms?	• Are existing efforts sufficient to mitigate risks? • If not, what else is needed or possible?	• Scientific and grey literature on interference via social media and ways to combat it. • Expert session to identify and validate courses of action to limit interference tools.

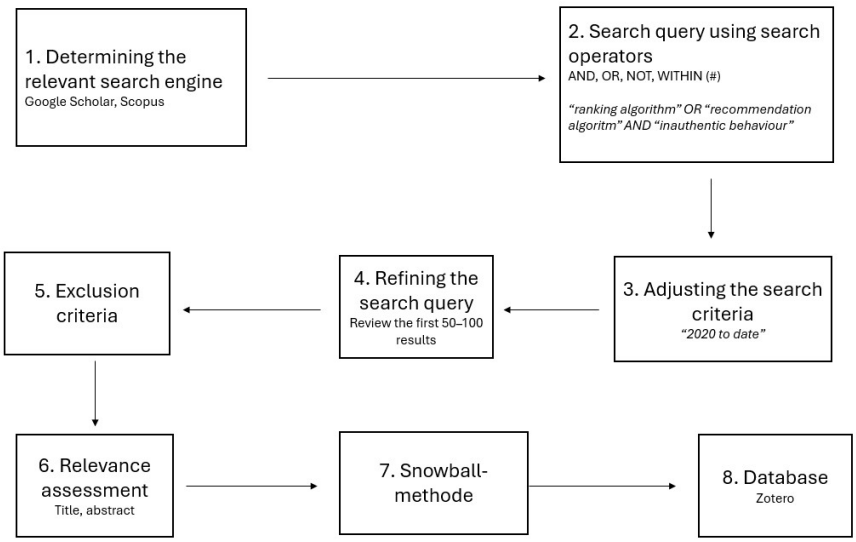
Source: Rathenau Instituut

Data collection
During the desk research, we analysed various sources: academic literature, grey literature (including journalistic sources and reports from NGOs and research institutes) and policy and legal sources. To create a database of relevant literature, we followed the step-by-step plan in Figure 4 (see below).

We began by using two search engines: Google Scholar and Scopus. We used Google Scholar primarily to formulate the search queries (step 2). For each sub-question, we developed one or more search queries, distributed among the various researchers.

Once the search queries for each sub-question had been formulated and the search criteria established (steps 2 and 3), we reviewed the first fifty to one hundred search results in Scopus (step 4) and carried out the following steps using this search engine. For each sub-question, we selected a maximum of twenty sources. This selection formed the database of sources with which we answered the sub-questions.

Figure 5 Schematic representation of the data collection



Source: Rathenau Instituut

Zotero database of sources
We used Zotero to manage the database. Once the database had been populated with relevant literature, the researchers divided the literature by sub-question and read this section of the sources from the database. The researchers labelled the

five to ten most relevant sources as key publications. These publications were read by all researchers.

Expert interviews
The researchers conducted interviews with experts to answer sub-questions 1 to 4. The individual interviews were carried out with the aim of:

- gain an understanding of the subject;
- test preliminary findings with experts;
- identify existing sources and hypotheses within the field;
- identify ongoing research and upcoming publications;
- elaborate on previously published research carried out by experts.

We did not follow a fixed questionnaire during these interviews. However, we did put forward one or more of the sub-questions during the interview.

Expert session
We organised an expert session to validate and refine preliminary findings and identify appropriate policy measures. The session helped to answer sub-questions 1 to 5.

Selection of experts
The session was organised with experts from various disciplines, such as political science, communication studies, media studies and cybersecurity. Investigative journalists and NGOs that publish specifically on inauthentic behaviour on social media were also present.

The experts were selected on the basis of the Rathenau Instituut's existing network. This was supplemented by experts identified from the analysed sources of the desk research and by asking experts if they could recommend people with relevant knowledge of the subject.

Data collection during the expert session and processing
During the expert session, preliminary findings were presented to the group of experts. The various sub-questions were discussed in breakout sessions. Notes were taken by the researchers during the expert session.

Platform interviews
Platform representatives were interviewed to answer sub-questions 2 and 4. Platforms were asked questions about the functioning of the recommendation algorithms on their platforms and their efforts and results in preventing interference via recommendation algorithms during election periods.

Prior to the interviews, desk research was conducted into platform policies and DSA audit reports. The platforms we invited to participate in an interview were selected based on user numbers in the Netherlands: YouTube (Google), Facebook and Instagram (Meta), LinkedIn (Microsoft), TikTok, Snapchat (Snap) and X.[24] On the advice of the expert session, Telegram was added to this list.

The platforms that participated in the research were Meta, TikTok and Snap. Meta and TikTok responded in writing to pre-prepared questions. An interview was conducted with Snap. Google YouTube and Microsoft (LinkedIn) indicated that they did not wish to comply with our request for an interview. X and Telegram did not respond to a request for an interview.

During the interviews, we followed a semi-structured questionnaire with a few questions directed at individual platforms. For the questionnaire, we drew, amongst other sources, on a publication by various civil society organisations containing questions that the European Commission might ask of major online platforms.[25]

Internal quality control

We carried out a comprehensive internal quality control process by submitting the draft report, in turn, to: the coordinator, an internal reviewer (not involved in the research), the management team contact, and a communications officer.

24 For this purpose, we are using data from the marketing agency NewCom on user numbers per platform: https://www.newcom.nl/nationale-social-media-onderzoek-basis-2025/

25 Fundacja Panoptykon, Irish Council for Civil Liberties, & People vs Bigtech. (2024). Fixing Recommender Systems (Briefing Note). https://peoplevsbig.tech/wp-content/uploads/2024/05/Fixing-Recommender-Systems_Panoptykon_PeopleVsBigTech.pdf

Appendix 3: legislation and policy

In this appendix, we discuss the most relevant legislation. It is important to note that each law has its own scope and objectives and therefore addresses the issue of interference via recommendation algorithms from different angles, both directly and indirectly. These are laws that may apply depending on the type of issue (or tool from the toolbox).

The most relevant legislation includes the Regulation on Transparency and Targeted Political Advertising (VPR), the Digital Services Act, the AI Regulation, legislation on media freedom, the Unfair Commercial Practices Directive, the Audiovisual Media Services Directive and the General Data Protection Regulation (GDPR).

The *government-wide strategy for effectively tackling disinformation* and the European Democracy Shield (EDS) are relevant policy initiatives. We will discuss the government-wide strategy separately, unlike the EDS. A large number of the EDS measures fall under the legislation discussed below and are therefore mentioned there.

We do not discuss the Political Parties Act(Wpp) in this appendix. The Political Parties Act would have imposed requirements on political advertising and microtargeting, but these have been removed from the bill as they are now regulated by the European Regulation on transparency and targeted political advertising (*Parliamentary Papers II, 36 742, no. 8*, 2025). This Regulation is implemented at national level by the Implementation Act on the Regulation on Transparency and Targeted Political Advertising. The legislative process for the Implementation Act has not yet been completed, although the relevant supervisory authorities (Media Commission, ACM and AP) have already been designated (Ministry of the Interior and Kingdom Relations, 2025).

Government-wide strategy for effectively tackling disinformation
The *government-wide strategy for effectively tackling disinformation* is intended to address disinformation which, according to the government, poses a major risk to free and open debate (*Parliamentary Papers II, 30 821, no. 230*, 2024). The strategy consists of three strands: 1) measures to tackle those who spread disinformation and the dissemination of disinformation, 2) measures to strengthen citizens' resilience, and 3) developing knowledge of the issue and effective approaches to tackling disinformation.

Under this strategy, the Ministry of the Interior and Kingdom Relations may, in exceptional cases, utilise its election flagger status (see Chapter 1). The government also works closely with other European Member States and institutions to detect signs of undue influence and take appropriate measures. Examples include the European Rapid Alert System (RAS) and the European Cooperation Network on Elections (also mentioned in the EDS). The Minister is currently investigating how to gain a better understanding of the spread of disinformation and foreign information manipulation and interference (FIMI, see Chapter 1) (*Parliamentary Papers II, 36 552, No. 12*, 2025).

Regulation on Transparency and Targeted Political Advertising (VPR)
The aim of this legislation is to help citizens recognise political advertising online more easily. The Regulation sets out transparency requirements for political advertising and rules on the use of targeting and optimisation techniques for political advertising. Interference and online manipulation were among the reasons for drafting the Regulation on transparency and targeted political advertising.

Optimisation techniques are methods used, for example, to increase the reach of an advertising message and thus target a specific group of people. Both targeting and optimisation techniques involve the processing of personal data. Targeting and optimisation are, in principle, permitted, but require, amongst other things, the explicit consent of the individual who is the subject of the technique. Furthermore, the personal data of the data subject (such as the social media user) may only be used by the data controller (such as the intervening party) if the latter party has collected the data itself. Indirect collection via data brokers is therefore prohibited. Political targeting of 17-year-olds or younger children, and targeting using special categories of personal data (such as information about political preferences), is not permitted.

It is unlikely that parties engaging in interference would seek consent or that such a party would acknowledge any involvement in an interference activity. Furthermore, campaigns of this kind are often conducted via data brokers, and indirect data collection through such parties is not permitted (Council of State, 2025). Furthermore, in many cases, data brokers possess data for which users have not given explicit consent. Consequently, in practice, this type of influence campaign is rarely lawful, and parties other than the interfering party may also be held responsible. For example, because an influencer is not transparent, or because, in a particular case, a social media platform itself acts as a provider of political advertising services.

This law also imposes transparency requirements on anyone producing political advertising, including influencers, bloggers, vloggers, etc. The law does not concern

the content of political advertisements. According to the VPR, when engaging in political advertising, influencers must also verify that the client does not originate from a prohibited third country (see below), submit information about this assignment to a European register, and include a transparency statement in the post.

Political advertising refers to a message (including its promotion, dissemination, etc.) that is conveyed, in return for payment, in one's own name or as part of a political advertising campaign, by, for or on behalf of a political actor, and which may influence or is intended to influence the outcome of an election, voting behaviour or a regulatory process. According to the European Commission's guidelines, discouraging people from voting is also a form of political advertising. Political actors are not only political parties, but may also be individuals or organisations that promote political objectives.

An example of political advertising might be a political figure paying a network of influencers to post similar messages, thereby indirectly promoting them. One example of this is the use of influencers in the run-up to the Romanian elections, who did not make explicit political statements but did share content that indirectly referred to the values espoused by a particular political candidate.

Platforms may also have a responsibility under the VPR. This applies as soon as they receive payment for the dissemination of political advertising. For example, if a platform, in return for payment, promotes a political 'advertising post' by an influencer by displaying it prominently, and the influencer is paid for that post by a political party. In such a case, the platform is providing a political advertising service and is deemed to be a publisher of political advertising within the meaning of the VPR, as set out in the European Commission's guidelines. Very large platforms and very large search engines have a number of specific obligations, primarily focused on public transparency regarding advertisements and responding swiftly to requests from, for example, complainants and regulators. For instance, to provide information or to take down unlawful advertisements.

The VPR states that, three months prior to an election, political advertising services may only be provided to clients who are EU citizens or have a specific connection to the EU. The Netherlands may impose stricter requirements in this regard. As part of the European Democracy Shield, the European Commission will bring these and other relevant EU rules to the attention of a voluntary network of influencers and encourage the development of standards.

The Digital Services Act (DSA) and the Code of Practice on Disinformation
The DSA aims to protect users of digital platforms from the dissemination of illegal content and to ensure that users' fundamental rights are respected. To this end, the law sets out responsibilities and accountability requirements for providers of intermediary services, such as Instagram, Snapchat and Telegram.

The DSA requires platforms to do their utmost to limit the dissemination of illegal content. Companies moderate content in accordance with European and national legislation. Various types of content (whether online or offline) are illegal in European Member States, including the Netherlands, such as discriminatory content, threatening content and terrorist content. In addition, the DSA contains provisions for complaints procedures and due diligence obligations for platforms, for example by responding appropriately after illegal content has been posted.

Stricter rules apply to very large platforms and search engines. These platforms and search engines include YouTube, Facebook, Instagram, LinkedIn, TikTok, Twitch, Google Search, Snapchat, X, WhatsApp and Bing. Telegram is not currently included, but is likely close to meeting the required user threshold (*EU in Touch with Telegram as It Nears Criterion for EU Tech Rules*, 2024).

Key provisions for our research are Articles 34 and 35, which require platforms to adopt a proactive approach. Under the DSA, very large platforms and search engines must identify systemic risks and take reasonable, proportionate and effective measures, for example to prevent negative impacts on electoral processes (such as the circulation of false information about candidates, voting data or fabricated expressions of support for candidates). The platforms are free to determine for themselves what these measures should be.

The DSA sets out a range of risk-mitigation measures, such as adjusting algorithms and recommendation systems to temporarily limit the spread of misleading or harmful content, monitoring viral trends on other platforms, deploying moderation teams familiar with the local language and culture, or temporarily banning AI-generated content in political advertisements. Platforms must report on their efforts to the European Commission.

The European Commission is currently investigating the functioning of X's recommendation systems (European Commission, 2025a). The European Commission is also investigating TikTok's algorithms (in relation to 'coordinated inauthentic manipulation or automated exploitation of the service') and its policies regarding paid political content and political advertisements (European Commission, 2025e).

There are also Commission guidelines for platforms specifically relating to electoral processes, such as removing financial incentives for platforms and influencers to spread disinformation, hate speech and extremist content via advertisements (demonetisation) (European Commission, 2024a). The guidelines also prescribe measures against botnets, inauthentic accounts or other misleading use of the service (anti-manipulation). Together with regulators and the European Election Cooperation Network, the European Commission is updating these guidelines under the EDS.

The European Commission may impose temporary measures on a platform under the crisis response mechanism in the event of a 'crisis', such as restricting the dissemination of certain information. The European Commission may also establish voluntary crisis protocols. Under the European Democracy Shield (EDS) policy initiative, the European Commission is working with DSA supervisory authorities to draw up such protocols.

In addition, major platforms and search engines must maintain registers containing information about the advertising (such as the intended target audience and the advertiser) placed on their platforms. These registers must be publicly accessible. Furthermore, providers of online platforms must take measures against unauthorised misleading practices (*dark patterns* and *deceptive design*).

Platform providers must specify, in clear and understandable language, the key parameters used in their recommendation systems. Users should have the option to modify or influence these parameters. Under the DSA, users must be offered, as standard, a feed containing content only from accounts they follow (a chronological timeline) in addition to an algorithmic feed. The NGO Bits of Freedom has won a court case against Meta in which it demanded that users be able to set the chronological timeline as the default, rather than the algorithmic timeline (*Meta must (for the time being) offer the chronological timeline as the permanent default option* until 2026).

Platforms and search engines must provide designated supervisory authorities and researchers with insight into how these platforms operate. Supervisory authorities may request data to assess compliance with the DSA, and platforms must, at the request of supervisory authorities, explain how their algorithms work. Designated researchers must also be granted access to data to investigate systemic risks. Platforms may only refuse a request from researchers if they do not possess the data or if access would lead to serious security or confidentiality risks.[26]

26 X has recently been fined for breaching transparency obligations. Among other things, it restricted access to data for researchers and provided insufficient insight into advertisements (European Commission, 2025d). The

In addition, there is the Code of Practice on Disinformation, which, according to the DSA, may give rise to an investigation by the European Commission in the event of non-systematic compliance. The Code of Practice can be seen as a further elaboration of the rules set out in the DSA. For example, signatories to the Code of Practice must endeavour to limit unauthorised manipulative behaviour and practices. This includes the use of bots, fake accounts and coordinated inauthentic behaviour, as well as 'user behaviour' aimed at artificially increasing the reach or perceived public support for disinformation. Reference is made to a regularly updated database of 'tactics, techniques and procedures' (DISARM Foundation, n.d.).

Finally, the DSA requires very large platforms and search engines to mitigate the risks of addictive design, such as a highly personalised recommendation system. The European Commission's preliminary finding is that TikTok has not taken sufficient measures to mitigate addictive aspects of its design (European Commission, 2026). The European Commission has also launched proceedings against Meta's Instagram and Facebook on the grounds that their design is too addictive for minors (European Commission, 2024c).

AI Regulation

The aim of the AI Regulation is to ensure that AI systems placed on the market or used by organisations within the EU are safe, do not harm health and respect fundamental rights. For example, there is a ban on AI systems that use manipulative techniques. It is not required that the party placing the system on the market also intends to cause harm, according to the (draft) guidelines of the European Commission (European Commission, 2025b).

Whether social media platforms, and TikTok in particular, constitute manipulative systems is a matter of debate. At the time of writing, a legal case is underway brought by the SOMI Foundation against TikTok, alleging that the platform deliberately manipulates young people and misuses sensitive personal data, which the platform uses to drive its own recommendation algorithm (*Dutch claims foundation launches legal action against TikTok and X (formerly Twitter)*, 2025). It will be interesting to see whether the court will indeed deem TikTok's recommendation systems to be manipulative, and whether this platform or parts of it might therefore already be prohibited under the AI Regulation.

European Commission has also stated in preliminary findings that Meta and TikTok did not provide sufficient access to research data. The investigation is ongoing (European Commission, 2025c). Another investigation is also underway against Meta due to poor access to research data and the taking offline of CrowdTangle, which was intended to monitor elections (European Commission, 2024b).

Under the AI Regulation, addictive design in an AI system may also be considered a prohibited practice if it exploits users' age, disability or socio-economic background.

It is striking that AI systems used to influence the outcome of an election or referendum, or to influence voting behaviour, are not prohibited, but are classified as high-risk. The developer must then, for example, explain in the technical documentation and user manual how fundamental rights risks are adequately mitigated. It is unclear where AI chatbots integrated into social media platforms that spread disinformation fall, as well as personalised, AI-generated news feeds.

Another relevant provision is the requirement that AI-generated content which is virtually indistinguishable from the real thing (deepfakes) must, under the AI Regulation, be labelled by the user as 'AI-generated' or words to that effect. Within the framework of the EDS, additional guidelines are being drawn up for the use of AI in electoral processes by political parties and other relevant actors.

European Media Freedom Act (EMFA)
This law aims to protect media freedom and pluralism in the EU and to promote the free movement of services. Among other things, the EMFA defines what constitutes a media service, which may include professional influencers who post content in return for payment.

According to the EMFA, third countries (non-EU countries) may have a negative impact on services. Media services must therefore disclose information regarding ownership and any advertising revenue they may have received from third countries.

Access to media services from outside the EU is also regulated in the event of a threat to or serious risk to public safety. The European Council for Media Services, established by this EMFA, advises on this matter. The Council comprises the Media Commission and its equivalents in other European Member States.

The very large online platforms also have a responsibility. They must, in accordance with the EMFA, provide a feature that allows users to easily ascertain that they are dealing with a media service and that the service can declare itself to be editorially independent of political parties or third countries. Media services must also be able to use this feature to declare that they do not provide AI-generated content without it being subject to human review.

According to the EMFA, the Council will organise regular structured dialogue between the very large online platforms, media service providers and

representatives of civil society. One of the aims of this is to monitor compliance with self-regulatory initiatives, including the aforementioned Code of Practice on Disinformation. The European Commission must also ensure that the internal market for media services is independently and continuously monitored for media concentration and for risks of foreign information manipulation and interference.

Unfair Commercial Practices Directive
Among other things, the Directive aims to combat unfair commercial practices, such as aggressive marketing or misleading commercial communications. This concerns content with a commercial purpose, as its aim is to sell something (a service, product, etc.) to consumers. Influencers often engage in advertising. It is important that the advertising is clearly identifiable as such. The influencer must therefore be transparent about financial flows and/or benefits they receive. The use of fake likes and fake followers is not permitted. In the Netherlands, the *Unfair Commercial Practices Directive* (UCPD) has been implemented in the Civil Code.

Audiovisual Media Services Directive
The Media Act implements the Directive and covers media services. So-called video uploaders (influencers, vloggers, content creators, streamers) are also covered by this. Video uploaders must also be transparent about advertising. Furthermore, they must join the Advertising Code Foundation. The code stipulates that misleading and aggressive practices, as well as the use of subliminal techniques, are prohibited. Furthermore, an advertisement must not appeal to feelings of fear or superstition.

Video platform services such as YouTube, Snapchat and TikTok must take measures to disclose advertising and the parties behind it. In addition, Snap, which has not signed the Disinformation Code of Conduct, has signed a separate code of conduct with the Media Authority. Among other things, the code of conduct sets out the measures Snap must take to combat harmful content and ensure protection.

General Data Protection Regulation (GDPR)
The GDPR is designed to protect the personal data of everyone within the EU/EEA. Personal data that is publicly available online is also covered by the GDPR. The same applies to derived data, which is data based on a person’s like or share behaviour. Data relating to political opinions enjoys additional protection as special categories of personal data; explicit consent is required from the party wishing to process such data.

All parties involved in the processing of personal data must comply with the law as data controllers. Such a party does not need to collect data itself. There may be multiple data controllers. An actor attempting to target Dutch citizens online must

therefore comply with the GDPR, as must all parties that facilitate this campaign and act as data controllers.

All data controllers, which may include platforms, must be transparent about how and why they process data and must provide access to that data upon request by the data subject.

The GDPR provides protection against practices involving the collection of personal data, such as profiling and targeting on social media and in chatbots and other Large Language Models, insofar as personal data is collected in the context of this investigation. This is because data may not be collected if it is disproportionately detrimental, unexpected or misleading to the individuals concerned (principle of fairness). The European Parliament's research service, the EPRS, expects that the recommendation system of, for example, TikTok – which, according to the research service, uses people's behavioural data to manipulate users – is incompatible with the principle of fairness under the GDPR (European Parliamentary Research Service, 2025).

According to a recent ruling by the Court of Justice of the European Union, platforms must also proactively scan information for, for example, the use of sensitive data in an advertisement and take immediate action in the event of the unlawful publication of such data. Normally, platforms are under no obligation to act proactively, and they may or may not follow up on a complaint (notice-and-action mechanism), but the implications of this ruling appear to be similar (*X v Russmedia Digital SRL and Inform Media Press SRL*, 2025).

Furthermore, personal data may not be transferred outside the EU to specific countries such as Russia and China without certain legal conditions being met. However, it is not easy to ensure that non-EU parties comply with these requirements. The US company Clearview, which has been fined and ordered to take corrective measures by various European supervisory authorities, refuses to pay or to comply with the measures. The AP is currently investigating whether it can take action against Clearview's directors individually (Autoriteit Persoonsgegevens, 2024).

Contact details
Anna van Saksenlaan 51
PO Box 95366
2509 CJ The Hague
070-342 15 42
info@rathenau.nl
www.rathenau.nl

The Rathenau Institute promotes public and political discourse on the societal aspects of science and technology. We conduct research and organise debates on science, innovation and new technologies.

40 JAAR Rathenau Instituut
Onderzoek & dialoog | Wetenschap, technologie en innovatie