



Polarisation and the use of technology in political campaigns and communication

STUDY

Panel for the Future of Science and Technology

EPRS | European Parliamentary Research Service

Scientific Foresight Unit (STOA)

PE 634.414 – March 2019

EN

Polarisation and the use of technology in political campaigns and communication

Digital technology was once heralded as a boon for democracy. However, the contemporary political reality has underscored the ways in which technology can also undermine citizenship, democracy and the international liberal world order. Thriving on new media ecosystems, populist voices in European democracies have established themselves as mainstream political actors.

This report offers a comprehensive overview of the relationship between technology, democracy and the polarisation of public discourse. Technology is inherently political, and the ways in which it is designed and used have ongoing implications for participation, deliberation and democracy. Algorithms, automation, big data analytics and artificial intelligence are becoming increasingly embedded in everyday life in democratic societies; this report provides an in-depth analysis of the technological affordances that enhance and undermine political decision-making, both now and in the future. To conclude, we formulate principles and policy options for fostering a better relationship between digital technology and public life.

AUTHORS

This study has been written by Lisa Maria Neudert and Nahema Marchal of the University of Oxford, at the request of the Panel for the Future of Science and Technology (STOA) and managed by the Scientific Foresight Unit, within the Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament.

ADMINISTRATOR RESPONSIBLE

Philip Boucher, Scientific Foresight Unit (STOA)

To contact the publisher, please e-mail: STOA@europarl.europa.eu

LINGUISTIC VERSION

Original: EN

Manuscript completed in March 2019

DISCLAIMER AND COPYRIGHT

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

Brussels © European Union, 2019.

PE 634.414
ISBN: 978-92-846-3963-2
doi: 10.2861/167110
QA-02-19-127-EN-N

Executive summary

Populist and extreme voices that once existed at the margins of European party systems have now become mainstream. From the rise of Italy's Five Star Movement to the electoral successes of Alternative für Deutschland (AfD) in Germany and Poland's Law and Justice Party (PiS), the populist wave that has swept over Europe in the last decade has baffled many political analysts and observers. Whilst initially understood as a reflection of the economic anxieties brought about by the 2008 financial crisis, it has become apparent that the recent surge of anti-establishment parties was strengthened by growing fears of cultural liberalisation, mass immigration and the perceived abandonment of national sovereignty to the diktats of globalised elites.

As public discourse coarsens and tensions rise, there is a growing sense that European society is more polarised than ever before. Against this backdrop, new digital technologies have taken centre stage in political processes – both as a source of information and a campaigning platform. In critical moments of public life, citizens of Europe increasingly consult social media for news about politics and public affairs. Such new and relatively unregulated platforms create new opportunities for nefarious actors to deliberately push false content and distort information flows for political gain. While the political, social and ideological forces that drive citizens apart are varied and complex, in recent years, scholars and policy-makers have increasingly pointed to digital technology as one of the potential drivers of polarisation.

The following report explores the relationship between these two phenomena in contemporary Europe. It does so by first reviewing two core mechanisms through which social media could be polarising European publics: inadvertently, through design choices and incentives (2.1 Polarisation by design) that potentially narrow the diversity of information accessed by individuals while facilitating the dissemination of divisive and emotionally-charged content; and deliberately (2.2 Polarisation by manipulation), through the exploitation of loopholes in an attention-driven media ecosystem to stoke divisions and manipulate users.

We then outline three counter-trends, whereby technology has and could continue to facilitate a better relationship between European publics and civic life, starting with social media as a catalyst and focal point for political activism, mobilisation and organising (4.1). We then touch on the powers of digital nudging, its effects on giving, civic debate and voting practices, paying special attention to how purposeful design and positive social nudging can help create healthier online environments and incentivise political engagement (4.2). Finally, we survey how advances in artificial intelligence, although still in their infancy, offer new opportunities to bring about better accountability and transparency in online information environments (4.3).

In the last section of this report, we sketch out how these trends may continue in the future. We note that as individuals increasingly retreat to private spaces to access and discuss political news and information, new challenges will emerge for policy-makers to monitor and remedy the spread of misleading, false or polarising information (5.1). Beyond that, many of today's political manipulation tools are likely to sharpen with time as they benefit from technological advances in artificial intelligence and increasingly embedded communication cultures. This could take the form of increasingly sophisticated conversational interfaces and 'deepfakes' for example – an image synthesis technique already used to create hoaxes and falsehoods involving politicians (5.2). Yet as technology becomes more and more politicised and regulatory efforts are ramped up to address these new threats, we caution against short-sighted legal action, which if taken precipitously, could have chilling effects on democracy itself (5.3).

To enhance future policy-making in this area, we conclude by highlighting guiding principles and recommendations for effectively tackling polarisation, which revolve around three axes: accountability and transparency, user activation and contextual literacy, and greater investment in digital infrastructures.

Glossary

Algorithm	A specification of rules a computer must follow in order to perform a task.
Artificial Intelligence	A field of science and engineering dedicated to the design of intelligent agents that act and think like humans and can make autonomous decisions rationally.
Big Data	A cultural and technological phenomenon that describes the maximisation of 'computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets' as well as the process of 'drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims' (Boyd and Crawford, 2014). It also has a mythological component in that it encompasses 'a widespread belief that large data sets offer a higher form of intelligence and knowledge' (ibid.).
Bot	Pieces of software designed to automate repetitive tasks, such as posting content online. On social media, bots often purport to be genuine human agents, with the intent to deceive both humans and algorithms.
Computational propaganda	The use of algorithms, automation, big data analytics and human curation to manipulate public life over social media networks.
Connected devices	Electronic devices connected to the internet and to other devices via wireless.
Deepfake	An artificial intelligence-based image synthesis technique that involves creating fake but highly realistic video content misrepresenting the words or actions of politicians and celebrities.
Digital literacy	A component of media literacy that describes individuals' capacity to locate, navigate and critically evaluate information found in digital environments.
E-governance	Electronic governance, or the application of information and communication technologies for delivering government services.
Fake account	A social media account of a user who purports to be someone other than his or her real self. Fake profiles and accounts can be used for deception and manipulation, but they also fulfil a valuable function for political dissenters who might need to protect their real identity.
ICT	Information and communication technologies. These include telecommunication technologies such as mobile phones, as well as computers and other connected devices.

Machine learning	An application of artificial intelligence that provides computer systems with the ability to 'learn' through observations or data and perform tasks without being explicitly programmed to do so.
Natural language processing	The application of computational techniques to detect and analyse natural language and speech.
Polarisation	The political, social and ideological division of societies into distinctly opposed groups.
Populism	A performative political style that employs a divisive, anti-establishment rhetoric pitting 'the people' against an elusive 'elite.'
Social media	Digital platforms that facilitate the creation and sharing of information in networked communities.
Teclash	A portmanteau of 'technology' and 'backlash' that describes a public backlash against technology companies following high-profile scandals involving data and privacy breaches.

Table of Contents

1. INTRODUCTION	8
2. BACKGROUND.....	10
2.1 Social media	10
2.1. Algorithms	11
2.2. Polarisation	11
3. PUBLIC LIFE IN THE DIGITAL AGE.....	13
4. POLARISED PUBLICS IN THE DIGITAL AGE.....	15
4.1. Polarisation by design?	15
4.2. Polarisation by manipulation	22
5. DIGITAL TECHNOLOGY FOR A BETTER POLITICAL FUTURE.....	30
5.1. Social media as a catalyst for political activism, participation and change.....	31
5.2. The powers of digital nudging: how design can affect civic engagement	32
5.3. Towards greater accountability? Artificial intelligence and the quest for algorithmic fact-checking.....	35
6. FUTURE TRENDS	38
6.1. The changing nature of digital campaigning	38
6.2. Machine-generated political communication & deepfakes	39
6.3. Regulation and techlash.....	40
6.4. The death of the 'digital public sphere'?.....	42
7. POLICY PRINCIPLES AND OPTIONS	44
7.1. Promoting transparency & accountability	44
7.2. Governing the digital.....	45
7.3. Bridging digital divides.....	46
8. CONCLUSION.....	48
REFERENCES	49

1. Introduction

Digital technology has fundamentally altered how we access information, communicate and collaborate with our peers, and engage in political processes. Today, 72 % of Europeans go online on a daily basis and over half of them participate in social networks (Eurostat, 2017). Not only that but social media is now also fully embedded in political communication. Most political parties and candidates have an online presence and digital targeting has become an indispensable tool in any politician's electoral kit, with campaigners developing increasingly sophisticated strategies every election cycle. The massive uptake of information and communication technologies has also enabled new forms of political discourse and engagement. Platforms like Facebook, Twitter or WhatsApp are allowing more people than ever before to connect over shared interests and mobilise. From its role in the Arab Spring to the Maidan protests, research has shown that the internet can be a powerful catalyst for political participation and social activism by encouraging small acts of participation that cascade into larger movements (Gil de Zúñiga, Veenstra, Vraga, & Shah, 2010; Margetts, John, Hale, Yasseri, & John, 2015a).

But while the early days of the internet prefaced an era of great democratic renewal, there is now a growing concern that the very technologies that once promised to set us free are polarising democracy. This is particularly true in the European context, where government actors have expressed fears about the role of the internet and connected devices in the spread of viral online misinformation and increased legal scrutiny around the market power of technology companies and their repeated infringements of data protection and intellectual property law.

More often than not, these threats are manifestations of deeper-seated issues within European societies. While it is generally acknowledged that well-functioning polities require consensus and the capacity for citizens to engage with a range of political viewpoints, these basic positive conditions are increasingly being put under strain in a number of Western democracies. During recent years, waves of nationalism have swept over Europe; as a result, groups and parties that had for many years lived at the fringes of the political spectrum have taken centre stage. This phenomenon is exacerbated by a widespread sentiment that our societies are falling prey to radical voices and becoming more polarised. As these concerns mount in Europe, many point to digital technology as the root cause of growing partisan divides and deepening social inequalities.

Some for instance claim that polarisation can in large part be explained by recent changes to our information infrastructures. Certain critics have discredited the adoption of data-centric technologies in our societies with the emergence of 'echo chambers' and 'ideological silos' online (Sunstein, 2007) – hyperpolarised spaces where users are only shown information they already agree with, thus reinforcing their pre-existing beliefs and driving them towards extremes. But recent events also show that technology may be manipulated with the explicit intent to influence politics. Cases such as the Cambridge Analytica scandal and Russia's highly mediated interference in the 2016 US presidential election have exposed these platforms' loopholes and shown how easily their business models can be exploited with the deliberate goal of polarising communities.

As algorithms and automation become more and more embedded in civic life, the real question is: to what extent does technology play a role in driving citizens apart? And how may it be used creatively to remedy these problems and foster greater social cohesion across Europe? In this report, we seek to answer these questions by providing a comprehensive overview of the mechanisms through which technology can exacerbate social and political polarisation. We are nonetheless careful to avoid pessimistic tropes about the place of social media in our lives. As such, the second part will consider how technology might also contribute to countertrends, such as new possibilities for productive

engagement and debate amongst citizens. Finally, we will end by formulating policy recommendations to establish a positive relationship between digital technology and political life.

2. Background

As the digital revolution marches on, keeping up with ever evolving technological innovations has become more challenging than ever. The public discourse around basic and more complex technologies notably often lacks conceptual and terminological clarity. In this chapter, we set out a basic framework for evaluating the relationship between technology and polarisation. Specifically, we aim to equip readers with the toolkit and vocabulary necessary to participate in a productive discourse on central digital trends and their effects. Rather than providing prescriptive definitions and ground truths, we wish to offer a framework for readers to understand and discuss central technological trends as they relate to society. First, we provide an overview of several technological phenomena and their relationship to European politics and media ecosystems. In order to ensure a precise problem definition and analysis of polarisation, we then define some key terms for the interplay between technology, society and politics.

2.1 Social media

Social media services have been inundated with user groups from all age ranges and social backgrounds across Europe. But what exactly is 'social media', when so many technological platforms fall under the term in popular verbatim? On Facebook, users curate a personal profile where they post personal messages, images and links to online content, all of which eventually flow into a crowd-sourced newsfeed. Billions of messages are sent over WhatsApp, an encrypted peer to peer messaging app for mobile devices, every day. On the ephemeral social media service Snapchat, which is especially popular with teenagers, images and videos disappear within seconds of being opened. Users of TikTok record short videos to pop culture sound clips. Given the sheer number and diversity of platforms, it is arguably unsurprising that a comprehensive definition of social media remains contested.

As social media encompasses more and more functionalities—from messaging, to ephemeral stories, to long-format video—its definition is constantly evolving with technological progression. In 2007, Boyd and Ellison set a framework for future research with a widely-cited definition. They define social media as networks that allow users to '(1) construct a public or semi-public profile within a bounded system; (2) articulate a list of other users with whom they share a connection; and (3) view and traverse their list of connections, and those made by others within the system' (boyd & Ellison, 2007)

As social media matured and its user numbers grew, researchers increasingly placed emphasis on the networked properties of platforms as a constitutive element of social media. It became apparent that social media afforded the emergence of global networks connecting the user not only to other users, but also to information. Howard and Parks observe the role of social media 'to produce and distribute content [...] that takes the digital form of personal messages, news, ideas', as well as 'the people, organizations, and industries that produce and consume' content through social media (Howard & Parks, 2012). Social media companies themselves are rarely content creators. Rather, they are mediators that make content visible to users via search, recommendation and relevance algorithms (Tarleton Gillespie, 2014a). Drawing from these established definitions for the context of this paper, social media platforms thus include (1) the various personal, corporate, media and political actors that produce content on social media, (2) the content produced, which is mediated by algorithms, (3) the networked connections between users.

In Europe over the past decade, social media has become ubiquitous, transforming the ways in which citizens and corporate and political actors access, share and present information. According to a report by the European Commission, 63 % of internet users in the EU aged between 16 to 74 years used social

networks in 2016. Amongst younger users, aged 16 to 24 years, as many as 88 % were social media users (Eurostat, 2017). Europe's most popular social media site is Facebook, according to an analysis by World Economic Forum (Hutt, 2017). Facebook has remained tight-lipped about regional user numbers, but the company reported that it had 2.23 billion monthly active users in 2018 worldwide (Facebook Newsroom, 2017). In the European context, chat applications such as WhatsApp and Messenger—both owned by Facebook—also have very large active user bases. Instagram, Twitter, LinkedIn and Snapchat are further examples of popular social networks. Since the invention of social media only several decades ago, tens of millions of users in Europe have integrated these platforms into their daily lives.

2.1. Algorithms

Algorithms are essential to the functionality of digital technologies. Informally defined, an algorithm is 'a set of rules that precisely defines a sequence of operation' (Stone, 1973). In other words, at a basic level an algorithm is a set of rules a computer follows in order to solve a problem. In a non-technical sense, algorithms are used in everyday tasks, as for instance when we follow a recipe for apple pie or consult an instruction manual. In a technical sense, computers use algorithms for carrying out functions or processing data, typically at a much higher rate, speed and precision than the manual execution of that task would require. Algorithms are often expressed as computer code and, like lines of code, they can be chained together.

Algorithms operate all kinds of tasks. They steer operations from calculating values in an excel file, to sorting information, to displaying the colours of an image. As algorithms are designed to execute a pre-defined set of rules, they are often considered 'neutral'. Recently, however, the idea that algorithms that have been programmed and trained with human input might reflect underlying bias has become more widespread. We will discuss this further in the main part of this research review.

For the purposes of this paper, search, recommendation and relevance algorithms are central. A common feature of almost all social media apps and sites, these algorithms select pieces of information to display to a user from a much larger pool of information. Based on a particular set of rules, the algorithm decides that the selected piece of information is the most important or adequate to a given user. This functionality is present in the aforementioned scenario of social media sites or apps acting as 'mediators', with algorithms making content decisions.

Artificial intelligence, too, relies on algorithms to enable learning and problem solving. Machine learning is a primary application of artificial intelligence, although the two are often conflated with each other in popular discourse. Machine learning is derived from the idea that machines should be able to learn independently from data provided to them, rather than computers being prescriptively taught exactly how to carry out a task. In the practical process of learning from data, computers rely on algorithms that guide machine learning.

2.2. Polarisation

Polarisation has long been a concern for political scientists and opinion scholars alike. The term is employed almost ubiquitously in the press to express divergences in public opinion. Yet polarisation is a multi-faceted concept that can refer simultaneously to 'states of being' and 'processes' (DiMaggio, Evans, & Bryson, 1996). It is most commonly applied in reference to political or ideological polarisation, the former being the process by which a population splits into two distinct groups with contrasting views on political issues (Isenberg, 1986), and the latter the measure of ideological distance between them.

Beyond this traditional definition, however, polarisation might also allude to other divisions within society, as in social and affective polarisation. Social polarisation, for instance, describes the segregation of society into separate groups depending on economic and cultural factors such as income inequality, race, religion or differential access to jobs and real estate markets. Affect-based polarisation, on the other hand, is informed by the social psychological literature on group dynamics and describes a widening emotional gap, lack of trust and growing animosity between partisans (Iyengar, Sood, & Lelkes, 2012). According to this view, party membership also constitutes a form of group identity akin to social class that shapes an individual's behaviour and appraisals towards their in-group (in this case, members of their own political family) and a specific out-group (e.g. the political opposition). This type of polarisation has been shown to have risen over the last decade in a number of Western democracies (Iyengar, Sood & Lelkes, 2012). Negative media coverage, especially during political campaigns when partisan media outlets employ a harsher rhetoric against the political opposition, strongly reinforces this phenomenon (Levendusky, 2013).

But beyond technological advances and their profound impact on public life, Western European politics has also evolved in ways that could account for current levels of polarisation in the general public. The slew of political and financial scandals that have come to define the turn of the century – from the global financial crisis, to endemic corruption and the unpopular politics of austerity – have all dented public confidence in traditional political institutions. In particular, Italy and the United Kingdom have experienced a marked decline in public trust: in 2012, one in ten Italian citizens said they 'tended to trust' political institutions, however a decade prior that figure was one in three (Nielsen & Kuhn, 2014).

This mounting sense of distrust, coupled with institutional changes at the European Union level that have strained the ties between government and corporate interest groups, and an ongoing trend towards abandonment of traditional party politics (Nielsen & Kuhn, 2014), has only exacerbated the political alienation of electorates, and this situation has been readily exploited by populist forces. Vocal criticism of the relentless pace of globalisation, privatisation and a meritocracy that seemingly only benefits elites has given new impetus to 'anti-establishment' rhetoric and exacerbated schisms in societies. Throughout the report, we adopt a broad approach to the term 'polarisation' as denoting all potential factors that might drive European citizens apart.

3. Public life in the digital age

The ongoing public debate surrounding the relationship between polarisation and digital technologies is inherently connected to the increased integration of the internet and connected devices in everyday life. In the last decade, citizens and corporate and public actors across Europe have overwhelmingly embraced digital technologies as key instruments for political communication. This rapid digital transformation has fundamentally altered the ways in which Europeans access information about public life and participate in public decision-making processes.

Digital technologies, and specifically information and communication technologies (ICTs), have become widely available to the general public in Europe. Accessibility has advanced and costs have decreased, such that in 2007 the majority (55 %) of households in the European Union had internet access. Ten years later, in 2017, 87 % of households in the European Union were connected to the internet, with 72 % of all EU-28 citizens using the internet on a daily basis (Eurostat, 2017). The internet has been assimilated into daily European life and is used for a broad variety of applications: news and information seeking, shopping, professional networking, dating and to assist with everyday life tasks. The role of the internet, particularly as an information highway for political news, is constitutive to the European democratic process – as we will explain at length in this review. According to the Reuters Digital News Report 2018, between 65 % (Germany) and 94 % (Greece) of users in Europe get their news online, whereby in the majority of cases online is the primary source of news, coming even before TV and print (Newman, Fletcher, Kalogeropoulos, Levy, & Nielsen, 2018).

The shift to online sources of news and information has seriously impacted the media and advertising sector. Traditionally, the media monetised content by selling physical copies of their products (such as newspapers) and selling advertising space to commercial and political actors. Online, the majority of news and information is free of charge for users. Publishers thus generate their revenue primarily from advertising and sponsored content on their pages. When publishers distribute their content over social media pages like Facebook and Twitter, the advertising revenue goes to these social media pages instead. Decreasing revenue streams from traditional media channels, combined with the new challenges and technological requirements accompanying digitisation, have left quality journalism in crisis with many legacy publishers and local news outlets struggling to monetise content online (Levy & Nielsen, 2010; Nielsen, 2014). Publishers have also become increasingly reliant on social networks to distribute their content and ultimately drive traffic to their pages. A large body of literature has examined the relationship between traditional and social media. Although the evidence is far from unambiguous, a growing number of scholars fear that social media's tendency to reward virality over veracity may harm information quality and democratic discourse.

Despite these systemic changes, it should be noted that the distribution of news content over social media networks is also strongly intertwined with user behaviour. Social media users disseminate information on social networks by sharing links to online news sites and distributed content, a type of behaviour that is often encouraged in the new digital media ecosystem. Indeed, tools allowing users to share content at the click-of-a-button are now widely available and commonly integrated on news sites (Kümpel, Karnowski, & Keyling, 2015). A substantial amount of research on individual sharing behaviour shows that two main factors motivate information sharing: users share news first and foremost to enhance their social status, but also to interpret and contextualise information in networks (Bright, 2016a). The ability to disseminate content widely, both within one's social network and beyond, has shifted the power to set the public agenda away from the traditional gatekeepers and onto users themselves. In other words, through their engagement with content – in the form of likes and shares –

users can amplify specific media messages and have a strong impact on which news becomes more or less salient. (Singer, 2014) refers to this process as 'secondary gatekeeping.'

But the media sector is not the only one to have witnessed fundamental changes prompted by the advent of digital technologies. The internet has emerged as a central locus of global politics. As political actors connect with electorates directly over social media, governments and institutions are leveraging online information networks to make democratic processes more transparent and accountable. This situation, coupled with the fact that campaigns are increasingly relying on social media as a tool for mobilising voters and garnering electoral support, has arguably made democratic discourse richer (Howard, 2006; Nielsen, 2012). The internet has also undoubtedly empowered the voices of citizens around the world in many ways. User-generated content has democratised the free flow of information. A plethora of platforms specifically designed for the expression of political will – from online petition sites such as Change.org to digital calls to action – now provide new opportunities for participation, grassroots campaigning and movement building, as exemplified by the recent success of the #MeToo movement (Bimber, Flanagin, & Stohl, 2012; Karpf, 2012; Lessig, 2008; Margetts et al., 2015a).

In breaking down barriers between citizens and government actors, some of these advances have also directly impacted statecraft itself. In Europe, Estonia is perhaps the quintessential example of the revolution in e-governance. In less than a few years, the small Baltic nation of 1.3 million has set forth a successful model of digital statecraft by uploading almost its entire governance mechanisms to the cloud. In Estonia, 99% of government services – from citizenship to voting, from health care to education, from justice to taxes – are now available digitally (Ee, 2018). Every citizen is granted a digital identity card, which serves as the key to accessing these services. All of their data, from doctor's notes to tax information, is then securely stored online every time a service is rendered, without the need for office visits or any kind of paperwork. Since its inception, this digitisation process has reportedly saved 2% of Estonia's G.D.P. (Heller, 2017) and has encouraged international start-ups to do business in the country, where the barrier of entry, administrative burden and tax rates for foreign businesses are now extremely low.

4. Polarised publics in the digital age

4.1. Polarisation by design?

- The advent of digital technology has led to more personalised online experiences, potentially narrowing the diversity of information accessed by individuals.
- Information sharing has increasingly taken on a social function, facilitating the dissemination of partisan and emotionally-charged content that reinforces tensions between users.
- Greater customisation and moderation policies that lack legislative oversight run the risk of exaggerating existing inequalities of power and expression.

Digital technology has revolutionised the way information is produced, consumed and shared. Over the last twenty years the spread of the internet has democratised access to knowledge and information for broad sections of the population. In 2017, nearly three quarters of all European adults surveyed said they got their news online – up from 57 % in 2016 – and 65 % said that they turn to the Web to find information about goods and services (Eurostat, 2018). But when faced with the sheer volume of information available in digital markets, human attention has also emerged as a limited resource, and those who succeed in commanding it can leverage attention for economic, political, or social gain (Wu, 2016). Online public attention is held by just a handful of powerful digital companies. With the adoption of technology platforms as purveyors of news content, media companies have had to adapt their publishing strategies, including what type of content they push, as well as when and in what format they do so, in order to meet the demands of an increasingly competitive online advertising market powered by data-driven algorithms.

This move has spawned a number of developments with far-reaching consequences for our information ecosystems. One of those consequences is increased personalisation. From Facebook's News Feed to Google's PageRank, algorithms that tailor content to users' needs based on their previous online activity and preferences are now ubiquitous online. A second, related development is the symbiotic relationship between social media and 'shareability'. In the digital age, information sharing has increasingly taken on a social function (Beer, 2017). Promotional material aside, what people come across on websites like Facebook or Twitter depends almost entirely on what their friends, family, co-workers and acquaintances decide to share. Users also know how many times their posts have been shared or endorsed by others. All of this 'social' or 'meta' information forms the backbone of these platforms (Margetts, John, Hale, Yasseri, & John, 2015) and, over the years, has made 'shareability' a core marketing strategy for content creators everywhere. Content moderation is another development with serious implications for information ecosystems. Where news publishers formerly acted as information gatekeepers, this power has now shifted to users and platforms themselves, who can choose to opt in and out of, and even to report on, the content that they find disagreeable or offensive, thereby filtering out users with whom they would prefer not to interact.

In recent years, these mechanisms have become critical for an understanding of the ways in which social media can influence public opinions and attitudes. In the following section, we review each of them in turn and explore their implications for social and political polarisation. We are careful to avoid suggesting that technology itself wholly determines these social outcomes – a view that has become known as technological determinism – but rather take the view that technology is sociotechnical (Fischer, 1992; MacKenzie & Wajcman, 1999). Or, in other words, technology often impacts society in ways that were not foreseen or intended by its designers. As such, while networks, design and

infrastructure often enable or even catalyse society-wide practices, these cannot be divorced from the specific contexts in which they are situated.

4.1.1. Algorithmic ranking & content diversity

It is widely thought that a healthy democracy requires that people be exposed to a diversity of opinions (Downs, 1957). Yet in recent years there have been growing concerns that social media might be hampering that process by distorting citizens' political realities and sheltering them from dissenting views. Algorithmic decision-making and increasing personalisation are often cited as the main causes of selective exposure and isolation from diverse arguments. Algorithms play a central role in our day-to-day lives, dictating everything from search results to credit score ratings, mortgage loans and work schedules (O'Neil, 2017).

These systems are designed to follow a set of predefined rules in order to execute complex and sometimes repetitive tasks involving millions of data points. As such, algorithms are frequently touted as the best and most impartial means of sifting through massive amounts of information, predicting behaviour, making decisions and even performing medical diagnoses (Tarleton Gillespie, 2014b), and technologists and the general public alike share this attitude. A survey conducted in 2017 for Edelman's Trust Barometer found, for example, that 60 % of people surveyed worldwide were more likely to believe a search engine over a human editor (Edelman, 2017). In the same way, social media algorithms sort through content (e.g. status updates, pictures, videos and links) and rank it to determine which posts users see on their feed when they log onto the platform. While in the early days of social media algorithms ranked content in reverse chronological order, today Facebook's News Feed takes thousands of metrics into account, including the degree of affinity between the different users and the level of engagement their posts generate (Mosseri, 2018), in order to create a personal timeline based on relevancy and previous activity on the platform. These design choices are now the norm across other social media platforms such as Instagram, Twitter and Snapchat and therefore affect millions of users across the EU on a daily basis.

Over the last decade, this has prompted scholars and pundits to suggest that social media feedback loops could be trapping us in 'filter bubbles' or 'echo chambers', where users are only exposed to opinions they already agree with and never come across challenging content (Pariser, 2011). One of the most vocal exponents of this theory is legal theorist Cass Sunstein, who has long warned that the internet would segment society into digital silos where people only hear 'louder echoes of their own voices' (Sunstein, 2007). This opinion hinges on the idea that people naturally self-select into social and media niches that reflect their own ideological preferences, a process referred to as 'homophily.' By allowing users to 'wall themselves off from others' and prioritising content they are already predisposed to 'like,' social media algorithms prevent them from encountering other people's viewpoints and drive attitudes further towards the extremes.

Similar concerns have been raised about the radicalising potential of recommenders' systems (Tufekci, 2018). In a digital economy that monetises attention, recommender algorithms such as Youtube's Watch Time explicitly reward engagement – by pushing further content that will keep users on the platform for longer – over veracity or diversity of information. These videos often already draw a high volume of traffic and, at engineers' own admissions, tend to 'be sensationalist and on the extreme fringe' (Nicas, 2018). In this sense, algorithms are far from neutral, and play directly into human biases at the risk of propagating false and extreme beliefs (Noble, 2018). YouTube has, in fact, come under heavy criticism in recent years for disseminating and amplifying what was widely deemed to be misleading, conspiratorial and extreme content (Nicas, 2018). On the surface, there is thus good reason to believe that personalisation creates dangerous information asymmetries that starkly polarise users.

The reality, however, is slightly more nuanced. While controlled experiments confirm that people (especially the most politically active users) do seek like-minded news when given the choice, this does not mean that they are entirely averse to challenging content, but quite the opposite (Garrett, 2009; Munson & Resnick, 2010). Digital trace and survey data shows that people do in fact have a much more diverse information diet than they explicitly purport, and that selective exposure is more contingent than we think. In a choice-driven media environment, internet users do not exhibit monolithic consumption habits: most individuals source information about politics and current affairs through a wide range of channels and platforms (Ahlers, 2006; Dutton, Reisdorf, Dubois, & Blank, 2017; Newman, Fletcher, Kalogeropoulos, Nielsen, & Levy, 2017) and often share reading preferences with one another, which results in a high degree of audience 'duplication' between different news outlets (Fletcher & Nielsen, 2017; Webster & Ksiazek, 2012).

Finally, even when people do not seek out challenging opinions online, they are still likely to come across them inadvertently (Brundidge, 2010) as online networks tend to be less homogenous than personal networks (Gentzkow & Shapiro, 2011) thus boosting the likelihood of exposure to alternative perspectives (Bakshy, Messing, & Adamic, 2015; Barberá, 2014; Messing & Westwood, 2014). In other words, individual choice still plays the greatest role in driving exposure to diverse content. This does not mean that ideologically-closed spaces or 'echo chambers' do not exist online, but rather that they do not accurately reflect the digital reality of the average user. Unless we take social media platforms to exist in isolation from one another, more often than not, those that do are already populated by strongly partisan individuals who have actively sought them out.

4.1.1. The sharing imperative

Another way in which social media could be driving citizens apart is by encouraging the dissemination of increasingly partisan and emotionally-charged content. Openness and connectedness are fundamental pillars of social media. By virtue of their business models, platforms like Facebook, Instagram, Snapchat or Twitter have a vested interest in having users continually share information about themselves with known and unknown others. The more open and social users are, the more valuable metadata they produce, and this metadata can then be processed to inform targeted advertising strategies. In this way, social media sites might be said to have ushered in a new kind of 'participatory subjectivity' (Bucher, 2012), where engaging socially with others – whether through messaging, sharing content or commenting – is a prerequisite of being seen to exist on the platform. In many ways, every aspect of user experience, from 'like' buttons to push notifications, is geared to reward frequent engagement and active participation. The more friends a user makes, the more social currency he or she will have on these platforms (Ellison, Steinfield, & Lampe, 2007).

This 'sharing imperative' (Dijck, 2013) has also profoundly altered the journalistic and editorial practices of content creators. In a digital advertising economy where marketers fight for attention, monetisation and shareability have become inextricably linked: those videos and articles with the most traction on social media are also the most profitable in terms of advertising revenue. This is especially problematic, as studies have found that negative, sensational, controversial, and other kinds of polarising content frequently elicit the highest levels of user interaction and are therefore economically attractive (Bright, 2016b; Ziegele, Breiner, & Qiring, 2014). It is not surprising, then, that optimising stories' headlines and leads to 'draw readers in' and maximise exposure is now common practice in most newsrooms. Likewise, under the pressure to meet stringent objectives in terms of audience growth, many digital-born outlets actively favour sensationalist, emotional or partisan stances over more balanced reporting (Martens, Aguiar, Gomez-Herrera, & Mueller-Langer, 2018).

This is alarming for a number of reasons. First and foremost, there is already ample evidence that partisan media has the potential to turn relatively moderate readers into extremists (Levendusky, 2013). Numerous studies indeed show that partisan cues in news coverage of politics tend to intensify motivated reasoning, leading readers to more readily accept information that aligns with their pre-existing beliefs while rejecting evidence that might contradict them (Levendusky, 2013). In the same vein, exposure to ideologically-slanted news is shown to be associated with holding altogether false or inaccurate beliefs about political issues, even when readers are presented with counterfactuals (Garrett, Weeks & Neo, 2016).

Beyond making partisan identities more salient, the primacy of 'shareworthiness' (Trilling, Tolochko, & Burscher, 2017) is all the more problematic since partisan and sensationalist content can elicit strong emotional reactions that further exacerbate tensions between different groups. Generally speaking, researchers agree that stronger emotional arousal increases the diffusion of information in social networks (Berger, 2011), as well as audience engagement. Such patterns are evident offline in rumours or urban legends. It is now well-known, for instance, that content eliciting disgust is more likely to be shared than others. Anger and fear perform similar functions: outrage and anger make people less likely to distrust inaccurate information that supports their views, and more likely to distribute it. But while partisan and sensationalist content is explicitly geared towards grabbing readers' attention and evoking strong emotions, it can also easily fan the flames of existing tensions and turn people against one another. Providing evidence for this, a recent study by Hasell and Weeks (2016) has shown that partisan news arouses negative political emotions that boost information sharing, and that readers of ideologically agreeable news report higher levels of anger towards members of the opposite party.

Another unfortunate consequence of the prominence of emotionally-charged and partisan online media is its likelihood to boost exposure to incivility – in turn a further catalyst of polarisation. Despite initial hopes that the advent of the internet would bolster better and more constructive political debate, online discourse is mostly characterised by incivility (Coe, Kenski, & Rains, 2014; da Silva, 2013; Papacharissi, 2004): one analysis of online newspaper comments found that one out of five comments featured unnecessarily inflammatory remarks and insults. Moreover, the bulk of online comments emphasise opinion over fact, exhibiting a lack of desire to understand perspectives others than one's own (Hagemann, 2002).

Such incivility has been shown to decrease political trust (Borah, 2013, 2014), reduce open-mindedness (Mutz, 2006) and induce opinion polarisation (Asker & Dinas, 2017) and perceptions of bias in news coverage (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014). Finally, because information sharing on social media also carries a social function, it runs the risk of perpetuating confirmation bias. Belief formation is heavily influenced by the community in which information sharing is embedded. As Zollo et al. (Zollo et al., 2015) point out, this is particularly true online where readers process information 'through a shared system of meaning.' Consequently, they tend to place greater trust in content that has been socially endorsed by peers in their network (particularly if it comes from a like-minded other) than in the source of that information itself and are more likely to build collective narrative frames that contribute to consolidating established beliefs.

CASE STUDY I: THE BREXIT CAMPAIGN

If the 2015 UK General Election was widely dubbed the first 'digital election' (Edmonds & Elder, 2015), the 2016 EU membership referendum could, in very much the same way, be seen as the first 'digital referendum' – digital, in that both sides of the campaign eschewed traditional campaigning methods, instead crafting meticulous online strategies to further their cause, but also because the media battle played out largely, although not exclusively, in the digital realm.

Social media platforms are widely used for accessing news in the UK. According to a 2017 report from the Reuters Institute for the Study of Journalism, 41 % of British internet users say they get news via social media, especially Facebook, which is used as a source of news by 29 %, but also Twitter, which is used as a source of news by 12 % in the UK (Newman et al., 2017). All major newspapers covered the referendum extensively: close to 15 000 EU Referendum-related articles were published online across 20 national news outlets during the official campaign, according to a report by the Centre for the Study of Media, Communications and Power (Moore & Ramsay, 2017). But while most media outlets covered a wide range of issues – from the economy to immigration and health – the tone of the coverage was partisan and acrimonious at best.

The unprecedented nature and historical significance of the choice facing the British public gave the campaign an emotional undertone that contributed to shaping how far and how well campaign messages on both sides performed on platforms. In this respect, the Brexit vote provides a perfect case study for examining the role of emotions in information diffusion on social media. As we have discussed in the previous section, media content that arouses emotions such as anger and fear makes them more likely to go viral, especially on social media. Furthermore, the presence of 'moral emotional' language in political messages makes them more likely to be shared on Twitter (Brady, Wills, Jost, Tucker, & Van Bavel, 2017). While Remain initially received wide support from the political establishment, research shows that supporters of the Leave campaign were more effective in their use of social media and more engaged in the issue, aided greatly by the fact that they had an intuitive and more emotionally-charged message (Polonski, 2016).

Notably, the 'StrongerIn' campaign was primarily framed around economic arguments, casting the decision to leave the EU as an irrational choice that would bring financial dismay to the country. But while the economy was the most extensively covered issue overall – encompassing a total of 7 028 articles – Leave campaign leaders successfully took control of the media agenda, steering it away from economic issues proper and towards the impact of immigration on the job market and health sector. Coverage of immigration more than tripled over the course of the 10-week campaign, and was overwhelmingly negative (Ramsay and Moore, 2017).

Aside from publicly-owned broadcasters who were obliged to keep a more neutral stance, the majority of media outlets stoked partisan rivalries in the debate, a fact reflected both in the issues they chose to highlight as newsworthy and the tone of the coverage. Refugees, migrants and asylum seekers, for instance, were regularly turned into scapegoats for the country's woes in the tabloid press, playing on base affective states such as fear and resentment.

Careful analysis of social media data speaks to the efficacy of this strategy. Using 30 weeks of social media data from Instagram, Vyacheslav Polonski, a researcher at the Oxford Internet Institute showed that posts from supporters of the Leave camp received 26 % more likes and 20 % more comments than Remain posts, while the most active users in the dataset were also all campaigning for a Leave vote. In the same vein, data pulled from CrowdTangle, a social media analytics firm, made it clear that nearly half of the Facebook pages with the most engagement around the European Union were explicitly pro-Brexit

(Herman, 2016). According to the same analysis, in the second half of 2016, posts on pro-Leave Facebook pages produced more than 11 million interactions (defined as likes, shares and comments) – three times more than comparable pro-Remain pages.

Emotional appeals are nothing new and are far from unique to digital campaigning. The way the campaign played out on social media is, in some ways, neither an accurate representation of the deeper social and political dynamics within British society, nor the product of some kind of politically-driven 'digital gerrymandering' (Zittrain, 2014) on the part of Facebook or Twitter. But the surprise outcome of the vote should nonetheless give policy-makers pause and calls for greater attention to the mechanisms through which social platforms might inadvertently amplify certain messages, with real and, at times, irreversible consequences.

4.1.2. Shaping the digital public sphere

The last mechanism through which digital technology could be reinforcing polarisation within society is by shaping 'user participation into a deliverable experience' (Gillespie, 2018) at the risk of entrenching existing inequalities of power and expression. Major technology companies are playing an increasingly important role in controlling the global flow of information and setting the boundaries of the digital public sphere. As corporate entities striving to attract and retain users, they must strike a fine balance between upholding freedom of speech and maintaining a brand-safe environment for advertisers and customers.

Over the years, this has led technologists not only to grant users more freedom to customise their experiences online, but also to increasingly curate the content users are shown. Unlike system-driven customisation – the set of algorithmic tools that automatically dictates content exposure – user-driven customisation lets people specify their personal interest and preferences, opt out of content they find disagreeable or would rather not be exposed to, or block other users altogether. Research suggests that unfriending or unfollowing others is common practice among the most politically savvy users, especially when they perceive a high level of disagreement in their social networks, such as during election cycles (Bode & Vraga, 2018; John & Dvir-Gvirsman, 2015). Although affording greater flexibility to users, this practice has the potential to reinforce some of those insular information consumption patterns outlined above: the more people are given the choice to cull opposing voices from their social media feeds, the more likely they are to insulate themselves from dissenting views.

Over the last decade, as technology companies have scaled and grown their user bases, they have also had to manage ever-increasing amounts of data. According to Domo's latest *Data Never Sleeps* report, over 2.5 quintillion bytes of data are generated every single day around the world (DOMO, n.d.). With its 2 billion active users around the world and more than 307 million in Europe, Facebook is still the largest social media platform and one of the biggest data generators: 510,000 comments and 293,000 statuses are posted on Facebook every minute, and more than 300 million photos get uploaded per day (DOMO, n.d.). Photo sharing platform Instagram, a subsidiary of Facebook, follows closely, with 95 million photos and videos shared on the network every day.

This deluge of data in its multifarious forms – from photos and videos to text-based comments and updates – have created an urgent need for 'content moderation.' Content moderation is the practice of regulating or censoring user-generated posts in an attempt to keep online spaces free of harassment and nefarious content such as violent images or porn. As Tarleton Gillespie rightly highlights, moderation, recommendation and curation are baked into the DNA of social media companies who flip between these different levers to create the 'right' social exchanges and the 'right' kind of community

for their users (Gillespie, 2018). This endows them with enormous power to decide what type of content does and does not belong on their platform and more broadly, what kind of speech is appropriate online. For that reason, over the last five years, social media companies have taken centre stage in contemporary debates about free speech.

One of the core challenges of this practice is the secrecy and 'murkiness' of decision-making processes (LaRue, 2011) surrounding content take-downs, which are often imposed without public input and enforced 'with little clarity', as David Kane, UN Special Rapporteur on freedom of opinion and expression, notes in a report for the Human Rights Council. Given their quasi-monopolistic position, technology giants like Facebook – which has inadvertently become the world's largest media organisation – must create and enforce rules that can be upheld in multiple linguistic, societal and cultural contexts. This is no easy task. Less than a year ago, the Guardian published an investigation highlighting parts of Facebook's moderation guidelines. These threw into sharp relief the rationale behind the company's editorial judgements, as well as their limitations, which were chiefly criticised by journalists and scholars alike. Noteworthy examples include guidelines covering hate speech. Following the wave of terrorism that has struck the UK, France and other European countries in the last five years, social media giants have come under increasing pressure from governments to remove dangerous speech that is thought to promote terrorist acts or incite violence. In 2016, Facebook, Youtube, Twitter and Microsoft agreed to abide by a code of conduct on countering illegal hate speech online while 'defending the right to freedom of expression' and 'identifying and promoting independent counter-narratives' (European Union, 2016). Yet actual company policies on hate speech, abuse and harassment often fail to clearly define what constitutes a punishable offense. Both Twitter's prohibition against 'behavior that harasses, intimidates, or uses fear to silence another user's voice' and Facebook's two-pronged distinction between direct attack on a protected class and merely humorous or distasteful commentary that might be 'self-referential' (Facebook, 2018) are excessively subjective and vague, and have been criticised as insufficient bases for good content moderation.

In recent years, the vagueness of these harassment and hate speech policies have led to complaints that technology companies are reinforcing the status of powerful groups while unjustly penalising minorities. Over the years, reports have emerged of systematic forms of online abuse and threats against women, minority classes, disenfranchised groups, immigrants, refugees and asylum seekers going unchallenged. In parallel, platforms have been accused of unfairly suppressing the voices of political dissidents advocating against authoritarian regimes, but also social justice activists denouncing racist or gender-based discrimination and oppressive structures of power. This is compounded by the fact that social networks tend to follow a power law distribution, where a few powerful voices can have a disproportionate amount of influence on the information flow (Adamic & Glance, 2005). As part of their effort to police content and decide what type of information should be restricted or suppressed altogether – the bulk of which relies heavily on automation – technology companies thus play a primary role in shaping the diversity and character of public discourse, creating 'systems of governance' (Klonick, 2017) that dictate regimes of visibility and invisibility for others.

Needless to say, the murkier the rules, the less accountability users feel social media companies hold towards them and their grievances, which can foment a sense of distrust. This year, both Twitter and Facebook were accused of secretly 'shadow banning' and stifling conservative voices through overzealous content moderation. One of the major risks of this perception of unequal treatment is that when contributors feel they are being shunned or unfairly discriminated against on one platform, they will go one to create their own alternative social media spaces. This is already the case in the United States where alternative social media sites such as Gab.ai – the alt-right's answer to Twitter and the first self-proclaimed 'free speech' social network – have emerged as credible, if circumscribed, alternatives

to the mainstream (Marwick & Lewis, 2017). Once again, this is one of the ways in which these policy choices can reinforce intellectual retrenchment and seclusion, giving credence to Sunstein's idea that the digital public sphere is becoming increasingly Balkanised.

4.2. Polarisation by manipulation

- As user data becomes more abundant, campaigns and advertisers use it for micro-targeting, psychographic profiling and personalisation.
- Populist and extremist political actors exploit attention-driven media ecosystems for political and economic gain.
- Social media ecosystems and traditional media have proven susceptible to amplifying the reach of polarising and conspiratorial content and spreading it into the public mainstream.
- Influence campaigns designed to sow division and manipulate the public thrive on social media and use bots, junk news and propaganda to disrupt public discourse.
- These tactics have become entrenched in political discourse where foreign and domestic actors rely on them to influence political life.

In the previous section of this review, we surveyed how the design of modern technical infrastructures, algorithms and networks can become a root cause of polarisation. In the following section we discuss how technology in its social embeddedness is connected to polarisation processes. We dismiss a technological-determinist stance as reductive and assume that technological properties influence, enable and incentivise how specific technologies are used, appropriated and, most relevant to our inquiry, manipulated by human users. Technology shapes social processes and vice versa; society shapes the uses and applications of technology. This section focuses on how politically and economically motivated actors use technology to manipulate political discourse and public life in countries across Europe.

While digital ecosystems have certainly liberated global information flows, it has recently become clear that they are vulnerable to manipulation. In tandem with digital tools, human actors are developing strategies to harness these vulnerabilities for their own gain; they interrupt, distort, and suppress information flows in ways that may ultimately undermine democratic discourse and perpetuate polarisation, both unwittingly and intentionally (Tucker et al., 2018; S. C. Woolley & Howard, 2016). The 2016 Brexit referendum and the 2016 US Presidential election shed light on how foreign and domestic actors deploy technological systems for social, economic and political gain. In this section we will show that these techniques and strategies have become embedded in European political life.

Several manipulation techniques and mechanisms are well established in the political mainstream. The first of these is the misuse of personal user data harvested over social media for micro-targeting and psychographic profiling. The second one is the manipulation of attention-driven social media algorithms for political, economic or social gain in new and traditional media ecosystems. A third and related technique is the strategic use of automated bots and social media algorithms to spread fake news, hate speech and conspiracies tasked with the distortion of political processes.

4.2.1. Microtargeting, psychometrics & the weaponisation of personal data

Concerns over the misuse of personal user data for political purposes first appeared on the public agenda with the scandal surrounding the data analytics company Cambridge Analytica. During its heyday, the UK-based political consultancy specialised in data-driven campaigning and political marketing, attracting prestigious clients such as former presidential candidate Ted Cruz and current President of the United States Donald Trump. The company's data analytics methods were largely based on the academic work of Michal Kosinski at the University of Cambridge, who had developed a model to cluster people into personality types by combining seemingly unobtrusive data points, including demographics, survey responses, internet activity, social media data such as 'likes', and other publicly available data points (Grassegger & Krogerus, 2016).

Specifically, Cambridge Analytica described its services as 'data enhancement and audience segmentation techniques' offering a 'psychographic analysis [...] [and] behavioural microtargeting' that might be used for tailor-made marketing and granular voter mobilisation (Cambridge Analytica). The company even boasted that its psychographic techniques are at least partially to credit for Trump's success, a view that has also been iterated in the press (Kanter, 2018). When the news broke that Cambridge Analytica had illegally obtained personal data on millions of social media users, regulators and media across Europe and the United States sharply criticised the industry's inability to protect and secure personal user data and raised fundamental questions over the possible misuse of big data for political gain. In response, several countries such as the United Kingdom and the United States launched parliamentary inquiries into the Cambridge Analytica Scandal. Eventually, the company shut down in 2018, but questions about the efficacy and replicability of its techniques still persist today.

Some data can be harvested because users actively reveal information about themselves, for instance when they participate in a survey, add interests to their public social media profile, or geotag images they upload to a photography platform. Yet data is also being mined in more obtrusive ways: information on web traffic, browsing history, location, search and communication are all collated, as well as time and location logs on a plethora of connected devices, from mobile phones to IoT kitchen devices. The European General Data Protection Regulation (GDPR) seeks to mitigate some of the risks and infringements on user privacy that can accrue from such data practices. GDPR is an unprecedented regulation in its impact for protecting data privacy and security, and enforcing data processing standards throughout Europe.

Nevertheless, it is still possible—and common practice amongst marketers, advertisers, publishers, and political campaigners alike—to leverage user data to gain insights about users' behaviour, preferences, attitudes and lifestyles, which then ultimately informs decision-making in areas as diverse as product design to formulating policy and campaign slogans (Ghosh & Scott, 2018). The result are precise microtargeting and profiling algorithms that involve a more granular market segmentation. Segmentation is the process by which large groups of customers, users or voters are combined into smaller subgroups (referred to as segments) based on shared characteristics and behaviour. It allows for predictive modelling of customer or voter wants and needs, in a much more flexible manner than had been possible prior to the advent of big data analytics. Big data thus supercharges digital advertising, content marketing, and political campaigning by appealing to an audience in a more directed and intricate way.

But can the technologies that help us to activate and appeal to users foster polarisation? As early as 2006, when data driven technologies were nowhere near as refined as they are today, Howard weighed in on this question with cautionary words about the near future. He warned that politicians would be able to mine personal data to draw political inferences and 'redline' particular communities from political information, and that data could be used to manufacture attitudes resulting in what he termed

'managed citizenship' (Howard, 2006). Nieslen (2012) similarly acknowledges that data-driven targeting increases the control that campaigners have over political information, resulting in more personalised communication with messages designed to appeal to specific voter groups. Some fear that data-enabled tailored political messaging might result in the fragmentation of the public sphere and the creation of political bubbles, yet scientific evidence for this possibility remains thin.

As big data analytics techniques become mainstream, evidence about their misuse is accumulating. During the 2016 Brexit campaign in the United Kingdom, campaigners on both sides of the debate used so-called dark ads to target voters with conflicting information, according to the Bureau of Investigative Journalism (McClenaghan, 2017). Dark ads are pieces of advertising that are displayed only to certain user groups, and inaccessible to others. The Bureau of Investigative Journalism found evidence that pro-Remain groups targeted voters in at least one Labour seat with ads reassuring them that, if elected, Jeremy Corbyn would not become prime minister. Elsewhere, their advertising focused on the dangers of 'extreme Brexit.' Kreiss and McGregor (2018) have also communicated directly with employees of technology platforms like Google, Facebook, Twitter and Microsoft, to show that they actively helped to facilitate new ways of profiling, identifying, tracking and messaging individuals during the 2016 US election.

Putting voter privacy concerns aside, Kreiss (2017) argues that the pessimistic narratives surrounding targeted data-driven campaigning practices are shaped by fear: He argues that scholars overstate the effects of microtargeting because it falls short of democratic ideals and emphasises differences between competing social groups, interests and values. Kreiss contends that microtargeting is 'reflective of the fact that democracy is often partisan, identity-based, and agonistic.' What is more, the effectiveness of campaigning to persuade voters remains widely debated among scholars, who question the potential impacts of microtargeting altogether (Henderson & Theodoridis, 2017; Kalla & Broockman, 2018).

On the widely debated issue of psychographics specifically, scholars have disputed whether it is even possible to build the extensive psychological profiles on which companies such as Cambridge Analytica have supposedly built their success. Dobber and Helberger (2017) argue in an interview with Internet Policy Review that in Europe data-driven campaigning is still 'at an experimental stage': while parties dispose of the means to employ sophisticated, data-driven tools, political actors in Europe generally still refrain from using them. In large European democracies like Germany or France, campaigns still widely rely on more traditional media channels, such as broadcasting, billboards, or even door-to-door canvassing. Moreover, micro-targeting practices might also increase the cost-efficiency of political campaigns, aid the coordination of campaigning activities, and help to activate and mobilise voters with more appealing messages (Nielsen, 2012).

4.2.2. Media manipulation & attention-hacking

'It may not be good for America, but it's damn good for CBS'. This statement, made by TV CEO Leslie Moonves during the 2016 US campaign, is indicative of what Whitney Phillips describes as the 'oxygen of amplification', a phenomenon she claims to be unique to digital media landscapes (Phillips, 2018). The UK pro-Leave Brexit campaign, the right-wing Alternative for Germany party, AfD, the Front National in France and, most recently, the Sweden Democrats in Sweden all demonstrate how the mainstream digital media ecosystem has rewarded polarising, controversial and even radical voices with excessive attention and 'oxygen' to inflate their ideas. As a result, populist and ideologically extreme political actors have received billions of euros worth of free media.

Admittedly, some reporting has fiercely critiqued extreme positions in defence of democratic values. And scholarly research about whether this phenomenon is directly related to the populist Zeitgeist,

which appears to have captured ballots all over Europe, remains widely inconclusive. What is clear, however, is that populist rhetoric, deliberate provocation and conflict, and persona-focused politics are now well-established in the political mainstream of European democracies such as Sweden, Germany, France or Italy.

These strategies make the current media ecosystem precariously vulnerable to 'being played like a fiddle' (Danah, 2017) by manipulative actors. Marwick and Lewis (2017) have shed light on how a variety of fringe political actors, like nationalists, the US 'alt-right', and conspiracy theorists, successfully leveraged networked information dynamics to manipulate news frames, set agendas and propagate ideas. They infiltrated traditional media ecosystems by pushing outrageous, sensationalist and conspiratorial content into the public consciousness through social media. Once a piece of information has surpassed a certain threshold of visibility online, journalists and other traditional gatekeepers typically report on the news item and eventually fall prey to media manipulation. Marwick and Lewis presume that the mainstream media's dependence on popularity metrics, sensationalism, clickbait and novelty over newsworthiness is what makes them susceptible to these attacks.

These strategies have also trickled into the international political mainstream, where politicians and parties use populist and controversial rhetoric to hack user attention. Schroeder (2018) argues that populist and ideologically extreme actors leverage direct communication channels over social media in order to bypass traditional gatekeepers and garner disproportionate attention with voters, a tactic which he links to electoral success and polarised media landscapes. In a self-energising process traditional media outlets fall for this exploitative political rhetoric, and reward it with an even larger platform through their (albeit often unfavourable) coverage (Phillips, 2018; Schroeder, 2018).

The driving factors of media manipulation are not exclusively political but also economic, and thrive on the basic mechanisms of the digital media ecosystem. This has resulted in a political economy of manipulation designed to prey on flawed information systems. On Facebook 'engagement bait' pages became so prolific that the social media giant eventually banned them in 2017 (Silverman & Huang, 2017). In Macedonia, a group of teenagers published hundreds of fake news stories during the US election to turn a profit, making thousands of dollars of advertising revenue (Subramanian, 2017). Click-farms and follower factories have been discovered in the dozens on the dark and open web. They offer social media engagement as a service, selling fake accounts, bots, likes, and shares, to boost engagement and manipulate advertising systems (Confessore, Dance, Harris, & Hansen, 2018). On the internet, monetisation strategies are closely intertwined with user attention; economic profit provides a strong incentive to attract attention at all costs.

CASE STUDY II: FROM 4CHAN TO PIZZAGATE

In the run up to the 2016 US Presidential election a series of highly politicised conspiracy theories and 'fake news' items gained prominence in the mainstream media after infiltrating social media ecosystems. One of the most widely shared theories, 'Pizzagate', claimed that Democratic candidate Hillary Clinton was involved in running a child sex ring that promoted satanic rituals from the basement of a Washington D.C. Pizza restaurant. The rumours were traced back to 4chan and extremist circles on Twitter and Facebook, where proponents of the Pizzagate conspiracy falsely claimed that the leaked emails of John Podesta, Hillary Clinton's campaign manager, contained coded messages detailing human tracking operations that were supposedly run by several high-ranking officials of the Democratic Party from the restaurant Comet Ping Pong (Marwick & Lewis, 2017).

Eventually, the allegations were taken up by a number of extremist alternative media sites that further propagated the theory amongst alt-right supporters and conspiracy theorists. Several junk news sources

jumped on the story and published it on social media, where their stories generated hundreds of thousands of shares and interactions. Several mainstream alternative media outlets with large followings like True Pundit or Alex Jones' Infowars reported repeatedly on Pizzagate and bolstered the theory with even more conspiratorial claims and falsehoods (Silverman, 2016).

After months of sustained conspiratorial reporting, a US native entered the Comet Ping Pong restaurant to investigate the Pizzagate theory, which he claims to have come across in various articles and media coverage from known conspiracist Alex Jones. The individual fired an assault rifle inside the restaurant, yet no one was harmed. This incident in turn attracted prominent coverage in mainstream media, thereby amplifying the original conspiracy theory to an even wider audience. Prominent mainstream political actors also commented. Despite mainstream efforts to debunk Pizzagate, many of its proponents remained undeterred, demanding a more thorough investigation and discrediting the media coverage as dismissive (Marwick & Lewis, 2017; Miller, 2017).

4.2.3. Bots, fake news, and foreign influence campaigns

After reports of meddling in the US Presidential election and in the Brexit referendum in 2016, nefarious interference campaigns have taken centre stage on the global public agenda. Groups at the fringes of the political spectrum successfully hijacked the social media ecosystem to spread falsehoods, conspiracy theories and hate speech that fuelled cynicism and polarisation between voters. Politically and economically motivated actors deployed automation to amplify divisive content, weaponised an amalgam of facts and fiction to generate attention for their messages, and exploited the vulnerabilities of social networks to distort the flow of information.

In light of reports about election interference in Western democracies, security experts have expressed concerns about ongoing Russian information operations that have reached millions of social media users during critical moments of public life. With an unsophisticated and cost-efficient technical toolset and an expertly organised army of employees, Russian influence campaigns have disseminated disinformation, undermined the public discourse with bots, inflated controversial viewpoints at the fringes of the political spectrum, and hacked political actors across the globe. While attacks on the 2016 US election are to date certainly among the best-researched instances of interference, this tactic is not new. Former Soviet states like Ukraine, Latvia and Estonia, as well as countries like Hungary, France, and Germany, have repeatedly been exposed to Russian information warfare. And more worryingly, there is growing evidence that governments and political actors across the political spectrum, in both democracies and regimes, are embracing these techniques to manipulate political conversations.

Computational propaganda is the use of automation, algorithms, and big-data analytics to manipulate public life (Woolley & Howard, 2016). The umbrella term spans diverse phenomena in relation to attacks on the digital public sphere, which include the spread of disinformation, automated amplification with bots and fake accounts, the suppression of opposition with hate speech and trolling, and the infiltration of political groups and events. Computational propaganda exploits the very same vulnerabilities of the digital media ecosystem on which attention-hacking thrives, although it is oriented towards undermining political processes entirely, rather than using them for political or economic gain within the boundaries of democratic values. Typically, social media information operations are low in cost, readily accessible—to a large extent legally so—and can reach hundreds of thousands of users.

Influence campaigns take many forms and often span across various social networks. Recent campaigns have used three key techniques. Firstly, automated social media accounts, so-called bots, have been used to computationally create large quantities of messages in an effort to amplify controversial viewpoints, to flood genuine discourse with spam, or to suggest widespread consensus on issues.

Second, misinformation and conspiratorial 'fake news' pages purporting to be real news have been deployed to sow division and confusion, and to provoke controversy and resistance to political leadership. Third, armies of trolls and automated fake accounts feigning to be genuine citizens have been harnessed to infiltrate political discourse and then disrupt it with hate speech, junk content and vitriolic rhetoric.

Yet information warfare on social media is merely a symptom of a bigger disease. Influence campaigns do not operate in a political vacuum, but prey on existing public bias, fatigue with democratic processes, and scepticism vis-à-vis government leadership and politicians. To date, foreign information operations have rarely been tasked with supporting a particular political agenda. Instead, they have sown confusion by planting multiple and often conflicting misinformation stories in an effort to deny the public an objective narrative to debate. This technique seeks to stir division and is often employed strategically to inflate issues that polarise public opinion (Neudert, 2018; Woolley & Howard, 2018).

CASE STUDY III: THE CATALAN REFERENDUM

After years of separatism, in October 2017 the semi-autonomous community of Catalonia scheduled a referendum to decide if it should take its independence from Spain without the consent of the government in Madrid. The move was fiercely opposed by the Spanish officials, who declared the vote illegal and urged voters not to participate.

But despite Spain's best effort to thwart the referendum, Catalonians headed to the polls en masse on October 1st in defiance of what many saw as an obstruction of their democratic right to vote. An overwhelming majority of votes cast that day (92 %) supported Catalonia becoming an independent republic. But the results were immediately declared null by the then Prime Minister of Spain, Mariano Rajoy, further exacerbating tensions between the Spanish federal Government and separatist leaders and culminating in one of the biggest political crises in Spain's recent history. Citizens clashed with police forces; leaders of the secessionist movement were arrested; and in an unprecedented spin, Spain imposed unilateral rule over the rebellious region, suspending its autonomy for seven consecutive months.

As with many other polarising political moments, social media played a seminal role in Catalonia's push for independence and its fallout. First, in boosting the independence movement itself: hundreds of photos and videos circulated online and on social media in the lead up to the vote to reach out to potential supporters, and mobilise people to take part.

But beyond rallying voters, social media also contributed to the spread of unverified rumours and false information, which sowed confusion about what was happening on the ground on the day of the referendum. Faced with the Catalan's police refusal to intervene to halt the vote, Madrid sent out police officers from all over the country to shut down polling stations. Violent raids took place in multiple locations, including schools, where both civilians and police officers were injured. Using the #10Oct and #referendumcat hashtags, Catalan citizens took to Twitter to livestream and share images of the Spanish confiscating ballot boxes and firing rubber bullets at voters. Catalan emergency officials and several media outlets reported that nearly 900 people were injured in the police crackdown on the referendum – a figure that was widely picked up in the international press.

Yet, the referendum also stoked fears about possible foreign intervention. Several unverified claims and pictures circulated on Twitter and Facebook aimed at magnifying the perception of state brutality levelled against secessionists. According to reports by Politico and by the daily newspaper El Pais, several English-speaking pro-Russia Twitter accounts hijacked the Twitter conversation around the referendum, boosting

the #catalanreferendum hashtag by 7.500 % that day (Alandete, 2017). These accounts also successfully amplified multiple stories by Kremlin-backed Russia Today reporters about police violence at the polls.

The independent Spanish fact-checking organisation Maldito Buló also spotted a number of viral hoaxes about the police presence in Barcelona. Most of them included pictures depicting police officers in riot gear hitting defenceless voters and bloodied protesters. These, the organisations found, were in fact pulled from a 2012 miner strike in Madrid and doctored. Other suspicious stories that made the rounds on social media that day included allegations that a 6-year old had been paralysed as a result of police brutality, and reports that a police officer sent to Catalonia to block the vote had died of a heart attack there (Funke, 2017).

This prompted Mariano Rajoy to express concern about the spread of misinformation about Spain in the wake of the referendum – concerns which he later brought to the European Council. If the referendum itself and the political gridlock that ensued were driven by deep-seated political, economic and social divisions within the country, the case of Catalonia's botched independence votes exemplifies how effective foreign operators can exploit already-existing tensions and reinforce them through careful manipulation.

Social media companies like Facebook and Twitter have shared unique evidence with the US government, which documented in detail how Russian meddling targeted issues on race, specifically the Black Lives Matter community, gender and sexuality, and US veterans (Justice, 2018). More recently, evidence detailing how domestic actors in Myanmar, Brasil and Mexico exploited social media in an attempt to influence political processes has emerged. In Europe, evidence is more fragmentary. Pivotal information is often not accessible to the public, but locked away in the ivory towers of social media platforms. In their study of elections in France, Germany, the UK, and Sweden, the Oxford Internet Institute's Project on Computational Propaganda has found that disinformation campaigns in Europe predominantly focus on issues surrounding immigration, Islam, and xenophobia, with the aim of activating and reinforcing widely-held public fears (Hedman, Sivnert, Kollanyi, Neudert, & Howard, 2018; L.-M. Neudert, Kollanyi, & Howard, 2017).

As we are dissecting how impactful computational propaganda is for effectively promoting public polarisation, it is worth noting that scientific evidence on the effects of such campaigns remains scarce and ambitious. While propaganda content has been shared millions of times on social media, Allcott and Gentzkow (2017) claim that the reach of political misinformation is marginal, with many users not believing it. Similarly, the Reuters Digital News Report (Nic Newman et al., 2018) finds that the time users in France and Italy spent on quality sources of information drastically outweighs the time they spend with fake news.

Conversely, Vosoughi, Roy & Aral. (2018) find that falsehoods spread farther, faster and more dramatically than genuine information. Del Vicario et al. (2016) show that misinformation on Facebook spreads in communities of homogenous users that exhibit polarised viewpoints on specific issues. Research on the impact of bots is controversial, with some scholars postulating minimal effects for the spread of propaganda, and others finding that bots have strategically distorted civic discourse in moments such as the Brexit referendum (Howard, Woolley, & Calo, 2018; Vosoughi et al., 2018)). As mounting evidence of foreign and domestic influence campaigns amasses, vigilance and caution seem an appropriate response.

CASE STUDY IV: THE ALTERNATIVE FOR GERMANY

Angela Merkel sits before a glass of freshly poured German beer. A German stands on the dining table right in front of the chancellor. Her hands are folded into her signature 'Merkel diamond', a gesture that has become symbolic of the chancellor's dependability. 'It's fine', she says smilingly, gazing into the distance. On first sight, this would appear to be a moment of peace and quiet in a tumultuous political year. But the chancellor is not alone: five hooded figures are hiding in the background, pointing rifles at her. They have set the room into a sea of flames.

This image was one of many xenophobic, Islamophobic and nationalist messages that the ultra-right-wing Reconquista Germanica disseminated over social media in the run-up to the 2017 German election. The group had publicly declared a 'war of memes against the half-breeds in parliament' in an effort to excoriate the political establishment and to support the right-wing party Alternative for Germany, the AfD. Mobilising followers over the chat application Discord – a notorious breeding ground for extremism which is also used by the US 'alt-right' – the Reconquista Germanica crafted an Armada of manipulative propaganda messages in an effort to infiltrate the political discourse. With hashtags like #remigration and #duelofthetraitors2 their posts were shared hundreds and thousands of times and provoked a resounding echo in the mainstream media (Der Spiegel, 2017).

These tactics were not deployed exclusively by extremist fringe groups: they were deeply-rooted in the political mainstream. The right-wing AfD also took matters into their own hands, deploying hateful scare tactics, appealing to bias, and toying with the truth. They banked on populist slogans, captivating imagery and controversial political leaders to spread their right-wing agenda (Fuchs, 2017). One of their most viral campaigns depicted a pet pig with the slogan 'Islam? It doesn't work with our cuisine. Take the plunge, Germany!'. The party dominated the political conversation over social media, generating more traffic than any other party and outweighing even general traffic on the election itself. In a content analysis of German media Schärdel (2016) finds that AfD receives dramatically disproportionate amounts of media attention although reporting is predominantly negative in tone. Irrespective of effects on actual voting behaviour, these provocative communication tactics are successful in setting the political agenda, spreading like wildfire online, and reverberating within the media and political leadership.

5. Digital technology for a better political future

- The internet has emerged as a focal point of collective action, grassroots campaigning, fundraising, and petitions in digital political systems, both democratic and authoritarian.
- The internet and specifically social media offer users the tools to create and distribute content which has democratised the production of media and cultural products.
- Purposeful design and positive nudges can help create healthier online environments that are conducive to better and more civic engagement.

Political polarisation is widely considered a problem for democracy. The more elites and electorates diverge ideologically, the harder it is for them to overcome long-standing political conflicts. For Binder (2003), this explains why polarisation is often equated with fears of legislative gridlock, where parties struggle to reach democratic consensus on salient policy issues. A related fear is that of political apathy, whereby too large an ideological gulf between political elites can create frustration in the electorate, leading people to lose faith in political institutions or to become disillusioned with the political process altogether (Jones, 2015). In the previous section, we surveyed the various ways in which digital technology can further propagate polarisation within society.

However, history shows that polarisation need not always be a bad thing for democracy. Ideologically divided parties provide voters with a clear choice in terms of policy-based preferences. In the absence of such clear divisions, citizens tend to fall back on old habits and vote on the basis of geographical location for instance (Hetherington, 2008). Furthermore, eras of polarisation have the potential to drive political engagement and energise disgruntled voters, even those seemingly turned off by the extreme political choices with which they are confronted (Hetherington, 2008). Polarisation can be a real catalyst of political engagement. As Abramowitz notes, in the case of the United States, 'polarisation has actually contributed to increased engagement in politics, because people do perceive important differences and they think that there are big stakes in elections' (Abramowitz, 2010). This has been particularly true in the aftermath of Donald Trump's inauguration. Just a few months after the 2016 presidential bid that put the former real estate magnate in the White House, women began running for office at an historic rate, emboldened by the show of solidarity of the Women's March and the misogynistic tone of Trump's campaign (Chira, 2018). Similarly, the first month of his presidency saw the resurgence of spontaneous acts of participation, such as thousands of people showing up at airports to protest the Trump Administration's newly instated 'Travel Ban', which barred people from six Muslim-majority countries from entering the United States.

Similar patterns can be observed in Europe. According to observers, the European Union is now more polarised than ever before (Pisani-Ferry, 2015). But the wave of populism and far-right nationalist parties that have swept a great part of the continent in recent years has led to a broader ideological shift and given new momentum to youth-led progressive movements powered by social media. From widespread protests against tax cuts and increases in university tuitions fees in the UK to demonstrations against unemployment in Spain and Greece, European youths are far from politically apathetic. On the contrary, as we will see, digital technology has brought about new opportunities for wide-scale social activism – whether in the form of demonstrations, community involvement or charity work. The nature of social polarisation is complex and digital technology need not be a catalyst for social tensions and can instead foster better and more constructive engagements with politics. Here, we explore these countertrends, and provide a snapshot of how the same digital tools may support a more productive relationship between citizens and public life.

5.1. Social media as a catalyst for political activism, participation and change

The very technologies that may lock users in also have tremendous potential to set democratisation processes in motion. The internet, and social media in particular, have become a central breeding ground for social movements, political activism and grassroots campaigns in democracies and regimes alike (Howard & Hussain, 2013; Karpf, 2012; Shirky, 2008). During the #MeToo movement, women across the globe posted experiences of prevalent sexual abuse to social media that spilled over into a new wave of feminist politics. The movement sustained an ongoing conversation about women's rights and equalisation. During the Arab Spring internet users in Tunisia, Libya, Egypt, Yemen, Syria, Bahrain and other countries protested oppressive regimes and dictatorships in a vocal push for democracy, using social media to mobilise, organise and share their opinions within strictly controlled media systems. In 2014, tens of thousands of social media users participated in the ALS ice bucket challenge to raise awareness for Lou Gehrig's disease, including many prominent supporters. Movements have supported many different causes, ranging from large-scale global movements to local neighbourhood petitions.

So, how does the internet sustain such movements and activism for political change? Rotman et al. (2011) argue that the internet, and social media specifically, offer participatory, easy-to-use infrastructures that empower users to share content widely. Online information travels in real time. It bridges physical and social divides and circumvents traditional information gatekeepers such as journalists and political leaders. This trait is especially relevant in countries that experience oppressive communication cultures and censorship, often with tightly controlled media systems. In such countries, social media becomes a vehicle for communication that is free from government scrutiny. The possibility to anonymously create content is of tremendous importance in such political contexts. In Iran, for instance, blogging became an important source of news and information and a platform for expressing opposition views, often anonymously so, during the 2000s.

Margetts et al. (2015) argue that political activism online also takes the form of 'micro-donations of time and effort' in the form of liking, sharing, endorsing or downloading to support a myriad of political causes. When these tiny acts of participation prompt chain reactions, they amount to collective action that spans issues like university campaigns or global political movements. Frequently, the means of political participation are built into platforms designated for fundraising and petitions, such as change.org. In a detailed analysis of the advocacy group MoveOn, Karpf (2012) argues that online advocacy groups and political movements have transformed and democratised political action, and that technological advances have given rise to 'organizing through different organizations' rather than 'organizing without organizations'.

In addition, political activism and participation are inherently connected with user-generated content. The internet, and social media in particular, have transferred the formerly knowledge- and resource-intensive potential of professional content creation towards the user. This shift has democratised the production of media content, and has made content creation accessible to the mainstream public. Lessig (2004, 2008) argues that technological possibilities for generating content and 'remixing' media products are entrenched in internet culture to such an extent that anybody can now produce meaningful content.

Yet Lessig also claims that copyright issues and established structures of media ownership and power have in the past hindered adoption of digital technology. Similarly, Castells (2000, 2005) notes that with networked technologies the production of cultural goods is increasingly taking place in a networked (rather than hierarchical) manner, such that users have the technological means to reach masses of

people with user-generated content. Castells refers to this as mass self-communication. As a handful of social media platforms have risen to become primary information gateways on the internet, it is clear that content production has become far more user-oriented and participatory. Yet while power structures may have shifted, they have not dissolved entirely.

CASE STUDY V: DIGITAL ACTIVISM AND THE 'ICE BUCKET' CHALLENGE

In the summer of 2015, a social media campaign to raise awareness of amyotrophic lateral sclerosis (ALS) and encourage donations took the internet by storm. Dubbed the 'Ice Bucket Challenge' the viral campaign that saw millions of people all over the world (including a roster of A-list celebrities) challenge their friends and family to pour a bucket of ice over their heads to shed light on and raise money for research against the rare motor neurone disease.

Today, whether we like it or not, social media channels occupy a central place in the digital strategies of advertisers, politicians and campaigners, many of whom have made social media their favourite marketing playground.

From a communication perspective, social networks present significant advantages over other mediums when it comes to message diffusion. Their interactive nature, and the tendency for users to engage in copycat behaviour once enough people in their circles share a certain type of information, both make online publics particularly responsive and social media a powerful tool for incentivising 'digital natives' to take action in short spaces of time.

Social networks facilitate the exchange of information, especially when that content is emotionally charged or motivational, evoking emotions such as outrage and anger. By many standards, this is the secret ingredient to successful and viral campaigning: prompting social identification with a cause and nudging users to act. We have seen this strategy work successfully in the case of the ALS 'Ice Bucket Challenge', the viral social media campaign that saw thousands of people all over the world (including a roster of A-list celebrities) challenge their friends and family to pour a bucket of ice over their heads to raise money and awareness about a rare motor neurone disease.

In many ways, the challenge skilfully played off the possibilities and social norms of a platform such as Facebook: broadcasting one's accomplishment publicly and engaging in a socially desirable activity by supporting a good cause. The playfulness and competitive aspect of the task, which encouraged each participant to make their videos more amusing or entertaining than the last, combined with a low entry-barrier and a willingness to share, made the Ice Bucket Challenge a resounding success: a total of 2.4 million videos of people doing the challenge were posted on Facebook, raising an estimated \$220 million in donations (Erbenraut, 2015).

5.2. The powers of digital nudging: how design can affect civic engagement

Design can shape and be used to influence social behaviour. Persuasive technology – systems designed to steer people to adopt positive habits, for example using mobile apps and interactive websites – have long been deployed in domains such as sales, management, diplomacy and public health to incite behaviour change (Halpern & Sanders, 2016) Over the last decade, technology designers have increasingly developed motivational functions in the health and wellness sector, in the form of wearable fitness and tracking devices to incentivise users to meet the daily recommended amount of exercise and providing feedback on their performance. Behavioural economists use the term 'nudging' to

describe such design features (e.g. push notifications, messages, gamification) aimed at getting users to perform desired actions. Research shows that individuals are greatly influenced by those around them when they need to make decisions or decide on a course of action (Thaler & Sunstein, 2009). These 'social influences' can take the form of social and emotional cues and appeals, individual and group biases such as accepted norms within one's social group (such as one's family). Social psychologists and behavioural economists have shown that, special interests aside, people tend to default to the status or majority behaviour. But when prompted to engage in a certain way, they can be made to alter their behaviour. This is where digital nudging comes in.

Digital nudging is a simple and cost-effective way to incentivise individuals within networks and organisations to think and act in different ways by playing off social norms and acceptable standards of behaviour. In this sense, digital nudging can be a powerful way to spur people to engage civically: to pay their bills on time, to save money towards retirement or to donate organs. In recent years, policy-makers interested in 'nudge units' within the UK and US governments have applied insights from behavioural science and successfully piloted certain projects aimed at increasing charitable donations and encouraging people to save towards retirement (Halpern & Sanders, 2016). Specifically, a growing number of organisations in both countries are requiring their employees to 'opt out' of retirement schemes, rather than opt in. In the UK, the move to the 'opt out' approach has led 5 million extra workers since 2012 to start putting money aside (Halpern & Sanders, 2016).

CASE STUDY IV: THE FACEBOOK VOTING EXPERIMENT

Nudging can even help to battle political fatigue among citizens and actively engage them in democratic processes. Facebook enthusiastically uses nudging features across its different platforms: the ever prominent 'What's on your mind?' entry banner at the top of the Facebook Newsfeed; push notifications that ask you to tell your friends whether you can make it to their party; and an 'I voted' button that has become the subject of much controversy.

Out of the blue, on the day of the 2010 congressional elections, Facebook sent all of its 61 million US users – a quarter of the voting population – a reminder to vote. Every user in the USA over the age of 18 who logged onto Facebook was advised to 'tell your friends you voted'. Diligent voters were awarded with an 'I voted' button displayed on their profiles together with a counter showing the number of Facebook users who voted. Leveraging social information and desirability, the button was designed to boost voter turnout in the USA (Grassegger, 2018).

But can a simple 'I voted' button really increase voter turnout? A team of researchers from Facebook and the University of San Diego set out to probe this question. The team compared datasets from what they called a '61-million-person experiment in social influence and political mobilization' against elections returns (Bond et al., 2012). Their study found that 'more of the 0.6 % growth in turnout between 2006 and 2010 might have been caused by a single message on Facebook' – a small percentage but given the number of voters and the fact that elections in the USA have been decided by margins in the past, the result was outstanding.

The study that was published in the highly-renowned journal Nature, however, was heavily contested and even rejected by members of the research community. Several academics criticised it on the basis that an experiment performed on unknowing participants without prior approval was unethical, especially given the tremendous implications it may potentially have had for the outcome of the election. Others criticised methodological flaws and the very small effect size, which may have been subject to statistical error. In 2018, Facebook came under scrutiny once again for a lack of transparency and consistency in rolling out

voter button features in a handful of seemingly arbitrary countries over the years, and allegedly failing to include some voters (Grassegger, 2018).

Another way in which digital technology can help mitigate polarisation is by facilitating better civic engagement. Each social media platform is unique in that it possesses distinct affordances – or, as Ian Hutchby puts it, 'inherent properties' (Hutchby, 2001) – that structure the norms and behaviours of its users. Through a platform's interface and the presence or absence of features such as a 'like' or 'share' button, users are constrained and subtly encouraged to act in specific ways, which taken in their entirety form the social fabric of an online community. In this sense, the affordances of social media – the way these spaces are designed and the rules that govern them – can play a critical role in shaping users' social behaviour and can even be steered so as to boost people's abilities to engage in better political conversations. A case in point is a discussion forum on the social news platform Reddit, called ChangeMyView (CMV).

Reddit is a social media site where users can post and discuss content in bulletin-board style communities called subreddits. Users are pseudonymous and empowered to join and contribute to whichever community interests them. Over the years, this has contributed to a strong culture of free speech on the platform. ChangeMyView (CMV) is a subreddit that encourages its users to change their opinion by inviting reasoned counter-arguments on specific topics of interest from other members. CMV has been the subject of numerous studies and academic research papers and widely heralded as an example of how design mechanisms and social norms can be deployed successfully to reduce partisan antipathy.

CMV subreddits are powered by gamification. Each user submits a question to the forum with the explicit aim of having their view challenged in an intelligent and reasoned manner. Submitters who have successfully had their view changed are expected to award other users a 'delta', a form of digital reward. Deltas allow the community to keep track as well as rank its members based on the quality of their contributions. It has been shown to be a powerful driver and justification for users to engage in sustained and polite discussion with their peers online. In addition, the rules of engagement on the forum are clearly stated: ad hominem attacks and incivility are prohibited and comments that do not follow these rules will be swiftly taken down. For Jhaver, Vora and Bruckman (2017), this represents the perfect combination of descriptive, injunctive and personal norms: the strict enforcement of rules by ChangeMyView moderators sets clear expectations, while gamification encourages new users to be polite with others, all contributing to maintaining civil discussions on the platform.

In recent years, numerous online tools have similarly been developed, such as efforts to spur political engagement and participation through gamifying voting. In 2012, an Argentinian coalition of coders and political scientist launched their own political party – the Net Party – and ran for a seat in parliament, with the aim of electing a representative who would represent the wishes that constituents had expressed online. To achieve this goal, the group built an open software platform, DemocracyOS, to allow citizens to directly participate in the political process by voting on and editing legislation under consideration in parliament and proposing new laws themselves. The federal government in Mexico and the municipal government of Paris have since been using the tool to get feedback on proposed policies and give citizens a stronger voice. The platform has also enabled more than 1 200 citizens around the world from 40 different countries to contribute to the debate on a simplification of the text of the COP21 negotiation, made available on the OpenCop 21 (DemocracyOS, n.d.).

Similar efforts also exist in the United States, where the Open Gov Foundation's Project Madison, also an open source software platform, has made available online every single bill currently being debated in the US House of Representatives for people to view, annotate and comment on. Government agencies and civil society groups can also use the tool to crowdsource policy ideas directly from members of the public. While these initiatives are no panacea for the challenges of contemporary democracy, they offer a glimpse into the potentiality of technology and purposive design to bring about a more open, civil and participatory politics.

5.3. Towards greater accountability? Artificial intelligence and the quest for algorithmic fact-checking

Another significant way in which digital technology can mitigate the negative effects of some of the developments we highlighted in the first part of this report is by fostering healthier online communities and greater accountability. Whether on social media platforms, private messaging groups or newspaper comments, people are increasingly going online to voice their opinions and discuss current affairs with known and unknown others. But while the digital era carried with it the promise of democratic renewal and greater public deliberation (Papacharissi, 2004), evidence suggests that online discussion spaces are most often characterised by incivility and vitriol. These forms of aggression are manifold and rarely go unnoticed by users: according to an IPSOS MORI poll conducted for Amnesty International, roughly one in four women (23 percent) in eight countries said they have personally experienced online abuse or harassment – many because of their gender, sexuality or political views (Amnesty International, 2018).¹ According to another recent report by Pew Research Centre, the majority of American adults describe talking about politics on social media with people they disagree with as 'stressful and frustrating,' and agree that this frustration often leaves them feeling like they have less in common with others (Smith & Duggan, 2016). The negative impact that these experiences have on the well-being of users are an obvious source of concern.

As we saw in earlier sections, some platforms already crowdsource content moderation to users, by asking them to report inappropriate behaviour, and analyse meta information such as the controversy of a specific post or comment (based on how many times it has been upvoted or downvoted) to flag and quickly take down inappropriate content. These approaches, however, pose a number of ethical issues and tend to be inefficient or too slow. As a result, many technology and media companies have now turned to algorithmic and AI-powered solutions to combat online toxicity. Today Facebook, Twitter and most major newspapers, include the *New York Times*, task automated systems with identifying hate speech in comments as well as offensive or racist content, and instead prioritising constructive arguments.

Recently, Google developed a Perspective API, a program which uses machine learning to automatically detect toxicity in comments in real time. Once installed on a news site, Perspective 'scores the perceived impact a comment might have on a conversation' on a toxicity scale, which is created by feeding millions of comments from Wikipedia editorial and *New York Times* article comments, ranked by their degree of toxicity, into the algorithm (Google, 2018). These programs allow content moderation to work to scale in the ocean of data that is produced by netizens every day. They also help circumvent ethical issues surrounding content moderation performed by humans, which has been shown to have long-lasting negative effects on workers who are exposed to disturbing content day in day out. But while

¹ The poll was conducted on a total of 4,000 women aged 18-55 in the United States, Britain, Spain, Poland, Sweden, Denmark and New Zealand.

offering hope for the future, these systems are still in their infancy and can be easily fooled. A recent review study of the robustness of Google's Perspective API for example showed that the program was especially vulnerable to typos and misspellings (Hosseini, Kannan, Zhang, & Poovendran, 2017) – a loophole that a number of internet trolls have already exploited.

Shortly after the revelation in the summer of 2016 that Macedonian teenagers had been flooding the web with sensationalist and counterfeited stories to make money from Google ads, the existence of state-sponsored disinformation campaigns aimed at influencing electorates and destabilising Western societies quickly rose to wide public attention. Since then, there have been increasing efforts on the part of politicians, journalists and technology companies to effectively tackle the problem of online misinformation. As part of this attempt to craft effective, large-scale policy solutions, a great deal of attention has been brought to the potential of automated fact-checking tools. These systems function in very much the same way as toxicity detection algorithms: the task at hand is to be able to automatically and reliably detect false information coming from a news article or the transcript of a speech. That means deciphering language using natural language processing tools and collecting sufficient evidence to assess the veracity of facts. Academic research is quickly expanding in this area, with a number of initiatives having come about in recent years. For instance, the British charity Full Fact has launched two tools this year, Live and Trend, that automatically scan parliamentary sources and content from major newspapers such as the BBC and transcribe them in real time, as well as automatically displaying existing statements repeated throughout the news and government, making it easier for journalists to access them and fact check them (Funke, 2017).

These real-world applications are developing at a rapid pace and showing promising results. But if separating false claims from legitimate opinions can already prove difficult for humans, entrusting a machine to be the arbiter of truth is not without its own controversy or challenges (Fletcher, Cornia, Graves, & Nielsen, 2018). For this reason, both developers and practitioners point out that while progress is being made in this area, with detection algorithms now able to track false claims repeated in multiple places at once online, automated fact checking cannot be a substitute for the informed judgment of journalists. This is all the more critical, given that even reliable sources of information sometimes make mistakes: if fact checking algorithms determine credibility based on the language of a news report or the position of a news outlet in relation to other sources, as for instance some already do, they might as time goes on unintentionally learn to privilege legacy media organisations and consequently to gloss over their inaccuracies, whilst systematically downgrading smaller publications (ibid.).

CASE STUDY V: FACT-CHECKING AND AMPLIFICATION

In 2016, the Oxford Dictionary declared 'post-truth' as international word of the year, prompting a deluge of think pieces about the rise of a fact-free world order. Behind this claim? The idea that we are so polarised and blinded by ideology that most people reject any factual information that contradicts their pre-existing beliefs. That same year, the spread of online disinformation and falsehoods emerged as a major threat to public life, spurring technologists, educators and news organisations all over the world to find robust and scalable solutions.

This, notably, led to broader interest and support for fact-checking initiatives, including automated fact-checking tools – the use of artificial intelligence and natural language processing techniques, to automatically verify the accuracy of claims made by journalists or politicians against various databases. Today, there are at least 149 fact-checking projects currently active around the world – up from 44 in the spring of 2014, according to the Reporter's Lab (Wang, 2018).

Yet the academic debate still rages on as to the power of fact-checking to effectively debunk false information and therefore preventing further polarisation of the population along ideological lines. In a much-cited study, Brendan Nyhan and Jason Reifler popularised the idea of a 'backfire effect' (2010), according to which presenting someone with information that contradicts their original belief could have the opposite effect than the one intended by fact-checkers – entrenching falsehoods in people's memory and increasing misbelief.

Evidence of this is, in fact, more tenuous and nuanced than previously thought (Mantzaris, 2016). Recent studies show that while people still engage in motivated reasoning and tend to be resistant to corrections – especially when emotional issues are at stake – by and large, it seems people take note of factual information and debunking efforts do go a long way towards correcting false beliefs.

However, trust is key. There is growing evidence that if individuals do not trust a source of information, there will be more likely to flatly reject what that source says without engaging with the underlying argument. At a time of low trust in media and politics, this does raise the question of how journalists, fact-checkers and civil society should address with problematic or misinformation while avoiding spreading it further.

*A number of scholars have indeed shown that even critically reporting on misinformation and conspiratorial content in an attempt to debunk it can end up inadvertently boosting the kind of rhetoric and tactics that bad actors aim to spread. That is because online propagandists deliberately leverage some people's lack of trust in contemporary news media to seed doubts about the honesty of journalists' reporting. Once that trust is eroded, differences in worldview are easily weaponised to pit citizens against each other. As Alice Marwick and Rebecca Lewis noted in their 2017 report, *Media Manipulation and Disinformation Online*, '[F]or manipulators, it doesn't matter if the media is reporting on a story in order to debunk or dismiss it; the important thing is getting it covered in the first place.'*

For that reason, some argue that media companies should engage in 'strategic silence' (Boyd, 2018) – carefully understand the ways in which propagandists can exploit the vulnerabilities of the contemporary media ecosystem to polarise society and responsibly report on certain issues to avoid amplifying hoaxes and falsehoods that they deliberately push to social media in hopes that they will be picked up by the mainstream media.

Thus, as we have seen, there is no single solution to the multifarious challenges brought about by the digital transition, but a more value-oriented and effective deployment of AI systems, coupled with purposive design aimed at creating environments that are conducive to better political engagement, could set us on the right path and ensure that the internet does not go down in history as the tool that destroyed democracy for good.

6. Future trends

Concerns over social and political discord and populist currents are widespread across Europe. Looking back at the ways that technology has become embedded in society, it is clear that sociotechnical relationships are as versatile as they are volatile. It is often at the very beginning of the widespread adoption of new technologies that cautiousness and fears about their impact persists. Frequently, technologies become entrenched in ways that are very different from those envisioned by their inventors. The meanings and uses of technology are deeply intertwined with society. But society, too, develops and evolves with the implementation of technologies.

Will political polarisation emerge and manifest in democracies across Europe? Will divides widen to such an extent that they can no longer be bridged? Is technology a root cause, or rather a symptom of, much deeper-seated systemic issues? In an attempt to answer these questions and identify future trends, we believe we must reflect on both society and technology, and their interplay. In the following section we point to key trends situated at the intersection of politics and digital technologies. It is not the goal of this exercise to make prescriptive predictions about the future or to provide an exhaustive account of future technological innovations. Rather, we wish to point to emerging trends that are beginning to shape public life.

- Political campaigning is quickly evolving in Europe, with new forms of political engagement constantly surfacing.
- Nefarious actors will continue to generate innovation in artificial intelligence for machine-generated influence campaigns. Conversational interfaces and 'deepfakes' have emerged as key mechanisms.
- As technology becomes politicised, short-sighted regulation can have chilling effects on democracy.
- Individuals are increasingly moving to private spaces to access and discuss political information and news, creating new challenges for policy-makers.

6.1. The changing nature of digital campaigning

As we have previously discussed, the nature of political campaigning in Europe is fast evolving. For one, advertising spending is now chiefly directed towards the digital realm, with more traditional forms of outreach losing ground. In the UK alone, national political parties spent a total of £3.2 million on Facebook during the 2017 general election campaign, a figure up from £1.3 million two years prior (Sabbagh, 2018). Initially brought over from the USA after the success of Barack Obama's 2012 presidential run, digital marketing for political campaign is now a common practice which is quickly spreading across Europe.

However the precise means by which such digital ad spending changes hands and the messages it funds remain shrouded in mystery. For this reason, this year the UK Electoral Commission called for urgent legal changes to increase transparency and ask questions about how digital tools are being used for political campaigning. We are at the dawn of a new era that will see important changes to the ethical and legal landscapes of electoral campaigning. Furthermore, digital technologies, and specifically social media, are set to continue to alter the power relations between citizens and political parties (Gibson, 2015). They could generate issues in terms of campaign coherence, narrative control and resources, creating a rift between more digitally savvy campaigners and activists capable of mastering the tools required to create engaging social media content and those who are not.

At the same time, new practices are also emerging. As traditional campaigning techniques dwindle, actors situated outside of official party structures – such as civil society organisations and individuals – are acquiring growing visibility and digital capacity. In recent years there has notably been an increase in what Rachel Gibson refers to as 'citizen-initiated campaigns', i.e. those which encourage citizens to come together, participate in the political process and mobilise others. Exemplified by Momentum and other platforms such as the Progressive Alliance and CrowdPac, these 'satellite campaigns' (Dommett & Temple, 2018) harness the power of social media to recruit swathes of activists on specific issues and messages that transcend traditional ideals. As Temple and Dommett (2018) explain: 'These democratic intermediaries are more open and easily accessible than parties [...] By allowing people to engage without signing up to a programme of party ideals, satellite campaigns access and mobilise a wealth of new activists, and in the process bolster traditional party campaigns. To borrow Andrew Chadwick's term, the adoption of the internet and social media has precipitated a degree of 'organisational hybridity' for political parties who rely on the decentralised structures and fluid support bases typical of such social movements (Chadwick, 2013).

6.2. Machine-generated political communication & deepfakes

Personal computer and networked infrastructures have become prolific amongst users in Europe. Technologies such as social media, smartphones and connected devices have become staples for European citizens of all social, regional and age groups. Technological progress is often accompanied by promises of impending change and prosperity. In the recent wave of digital innovations, political actors across the globe have pointed specifically to artificial intelligence as a driver of democracy. Indeed, artificial intelligence will most likely emerge as a crosscutting agent for positive change. Yet the technology may also have stifling effects on political discourse. We will now identify several contexts in which artificial intelligence might, if misused, have questionable effects.

With recent innovations in artificial intelligence, technological systems are becoming increasingly capable of emulating human behaviours, including speech. Largely, these advances are driven by machine learning, an application of artificial intelligence, in which machines learn from real-life data such as conversation transcripts, video or social media posts. Users in Europe have already embraced and widely adopted the innovations that drive these developments. Voice- and language-operated interfaces like Amazon Alexa, Apple's Siri or Google Assistant can generate humanoid conversation. Furthermore, they can understand natural language and reply accordingly, rather than merely responding in a pre-programmed mode. Natural Language Processing is already fundamentally changing the means by which humans interact with machine interfaces. The customer service, marketing and healthcare industries have all been quick to adopt systems that automate client communication.

But when innovations such as these are used by nefarious actors they can be misused to automate propaganda and information campaigns. Equipped with conversational abilities and natural language processing, bots and fake accounts on social media are become more eloquent when distributing misinformation, conspiracy, and falsehoods. Today, most bots are blindly flooding the ecosystem with propaganda or reacting with stiff, pre-programmed replies to keywords. With advances in artificial intelligence, propaganda is becoming ever more conversational, issue-based and capable of understanding user intent and context (Neudert, 2018).

Similarly, so-called deepfakes have benefited from advances in artificial intelligence, which are used for image synthesis. Systems that produce deepfakes use existing images and video material, which are combined and superimposed onto source images and videos. For example, the Desktop app FakeApp

allows users to create videos easily, swapping in faces from elsewhere and doctoring them onto others' bodies in order to generate new, fake footage. When combined with voice software, these developments can create convincing deepfake videos of, for instance, politicians, public figures or fake witnesses. Researchers at the University of Washington produced a highly convincing deepfake video of President Obama delivering a speech, using 17 hours of video material to train their AI (Suwajanakorn, Seitz, & Kemelmacher-Shlizerman, 2017). There are currently no known cases of deepfake videos that have been used for influence campaigns. Yet as these technologies become more entrenched and accessible, it is highly likely that they will be deployed for propaganda.

Indeed, the capabilities and know-how required to develop such applications are already becoming more readily accessible. Almost every major technology company, from Facebook to Microsoft, has made AI toolkits and algorithms publicly available. Open source software libraries like Google's TensorFlow or Facebook's wit.ai are designed to attract developers in an effort to support third-party innovation and promote new features and applications. Unfortunately, these platforms are also widely accessible to bad actors who might repurpose the knowledge and capabilities made available to them for hacking, propaganda and cybercrime. Consequently, tech companies have begun to require developers wishing to use their APIs and platforms to pass an accreditation or approval process first. Critics have admonished that approval processes are often slow, lack transparency and may ultimately stifle innovation (Gorwa, 2018).

6.3. Regulation and techlash

In the aftermath of the 2016 US Presidential Election, the market dominance, algorithms, and content moderation systems of technology companies have come under increased regulatory scrutiny. A growing number of voices in government, industry and the broader public, have expressed concerns over the general vulnerability of information ecosystems to attention-hacking and influence campaigns, and their ability to promote content that may be harmful to democratic processes. In response, social media companies have stepped up their efforts to combat such threats, deactivating millions of fake accounts and bots (Timberg & Dwoskin, 2018) and establishing new standards for content moderation (Newton, 2018). They are also revising central features of their algorithms to promote quality content (Mosseri, 2018). Nevertheless, our analysis in this paper thus far reveals that fundamental harms, design flaws, and weaknesses persist, and pose an ever-present risk to the freedom of our democracies.

In Europe, policy-makers have begun to tackle the shortfalls of technology companies with legal and regulatory mechanisms. Measures like the General Data Protection Regulation, upload filters and link taxes, as well as national information controls on hate speech and misinformation such as the German Network Enforcement Act are tasked with the protection of citizens' rights and democratic values. The ongoing calls for regulation among policy-makers in Europe, along with the rise in tech-sceptical voices in the public domain have been referred to as 'techlash' (Evans, 2018). As social media companies continue to return record profits and grow in size, claims of a techlash seem a precipitant for the time being.

With policy-makers becoming involved in information controls and content regulation, vocal concerns about overregulation, content blocking and inadvertent side effects have flared up. Regulating emerging technologies is practically highly challenging, especially when they are as complex as social media and other digital technologies. Indeed, to date, calls for increased regulation of social media are generally flawed, short-sighted, and demonstrate a lack of expertise of system intricacies and basic technological functionalities, thus posing a hindrance to digitisation. Experts have strongly criticised

the tendency of EU regulation to shift copyright enforcement and intellectual property responsibilities onto technology firms, and to privatise the prosecution of illegal content. At the same time, countermeasures currently lack comprehensive guidelines, public oversight or transparency.

In a comparative review of emerging regulation efforts to counter malicious usage of social media, Neudert and Bradshaw (NATO StratCom, 2018) find that regimes around the globe are copying regulation aimed at controlling online information from European countries in order to legitimise censorship and content criminalisation. In countries where the rule of law is weak, governments appropriate these frameworks to tighten their grip on information flows. For example, Reporters without Borders have pointed out that a recently implemented Russian bill against hate speech is a 'copy paste' of the German Network Enforcement Act (Reporters Without Borders, 2017). In Germany the Act is already highly controversial, but in a repressive regime with a questionable rule of law, such bans on 'unlawful content' are likely to be used for suppression (Reporters Without Borders, 2018).

CASE STUDY VI: THE GERMAN NETWORK ENFORCEMENT ACT

After months of ongoing public debate, the controversial Network Enforcement Act (German: Netzwerkdurchsetzungsgesetz, or abbreviated: NetzDG) came into force in Germany on January 1st 2018. The Act pioneers regulation of online public discourses on social media in Europe, but does little more than strictly enforcing already existing law in the online sphere. The Act was introduced in response to pressing concerns about the freedom of online expression, hate speech, and unlawful content, as well as concerns regarding influence campaigns and fake news. Under the Network Enforcement Act, social media platforms are obliged to take down illegal content that does not conform to German law. If they fail to do so, they face heavy corporate fines up to 50 million euros (Reporters Without Borders, 2018).

On the very first day the NetzDG came into effect, a prominent public figure became subject to the new regulation. An Alternative for Germany politician had tweeted about 'barbaric, Muslim, group-raping hordes of men' (Hamann, 2018). Twitter responded by promptly deleting the tweet under the Network Enforcement Act and suspending their account. A day later the German satire magazine, Titanic, humorously reported that their Twitter account had been taken over by the politician. The satire magazine again tweeted about 'group-raping hordes of men' to satirise the original tweet. Twitter promptly deleted and suspended the Titanic account.

The incident fanned the flames of the already fiery debate surrounding the Network Enforcement Act. Public figures across the political spectrum, from Bild editor-in-chief Julian Reichelt, to internet activist Sascha Lobo, to UN rapporteur for Freedom of Speech David Kaye. AfD politician Alexander Gauland and the German digital association Bitkom, issued concerned statements about the restrictive impact the Network Enforcement Act may have on the freedom of expression, and the role of social networks as arbiters of truth (Neudert, 2018; Weber, 2017).

As one of the first pieces of regulation to explicitly address online social networks, the Act seeks to improve the enforcement of existing German law on social media and digital communication platforms. The NetzDG obliges social media platforms to take down unlawful content after receiving a user complaint. According to the German Ministry of Justice, the Act seeks to create compliance obligations for social networks to deal with criminal hate speech and other forms of illegal content. Specifically, the act strives for 'speedy and comprehensive processing for user complaints' (Bundesministerium der Justiz und fuer Verbraucherschutz, 2017). The NetzDG is aimed at commercial social media platforms with more than 2 million registered users in Germany (§1 NetzDG). At the moment it applies to Facebook, Twitter, YouTube, LinkedIn, Xing, Instagram and Reddit in Germany.

Proponents of the NetzDG argue that the Act promotes a healthy public discourse culture, referencing the German constitution and its principles of human dignity and freedom of expression. The official draft law argues that the online sphere is 'aggressive, abusive and often full of hate' and that 'anyone can fall victim to defamation because of his or her opinion, skin colour, origin, religion, or sexual orientation' (Bundesministerium der Justiz und fuer Verbraucherschutz, 2017). As political discourse increasingly takes place over social media, the Act could improve the enforcement of laws that have guarded democratic values throughout the history of the German federal republic.

Yet the Network Enforcement Law is also subject to vigorous criticism. Opponents argue that the Act elevates social networks to the position of judges (Reuter, 2017). Indeed, social networks are incentivised to delete extensively in order to avoid legal and financial consequences. Legal experts fear possible 'overblocking' that may threaten the freedom of online conversation (Holznagel, 2017).

Nevertheless, technology firms have failed to efficiently protect civil rights in the digital space. Self-regulation remains fragmentary, yet systemic risks persist. Policy-makers are obliged to address these shortfalls. However, as technology becomes politicised and the need for feasible solutions becomes more urgent, emerging regulation must not sacrifice rigour and far-sightedness in favour of short-term fixes. If it does, the openness and free flow of information on the internet are at substantial risk, which will inevitably propagate polarisation and reduce Europe's appeal as a hub for innovation and development.

6.4. The death of the 'digital public sphere'?

The rise of social bots and trolls, alongside greater moderation and surveillance, has increasingly pushed European publics away from traditional media platforms and caused them to retreat to alternative spaces such as Instagram, Snapchat and Tumblr (Newman et al., 2018). Following a year of scandalous data breaches and increased suspicion about the power of public social media companies to share individuals' data with third party companies with little regard for privacy, private and one-to-one messaging platforms have become increasingly popular venues for discussing or disseminating information about politics.

In recent years, Instagram for example has emerged as an alternative discussion platform amongst young people, notably in the United States. In the aftermath of this year's Florida school shooting in the USA, hordes of teenagers led sophisticated social media campaigns to bring attention to gun control and gun-related issues. In part disillusioned by the lack of adequate platforms for teenagers to express themselves and be taken seriously on big issues, many young people are now turning to 'flop accounts' on Instagram for their daily dose of political news (Lorenz, 2018).

Another salient example of this is the unlikely use of dating apps for political recruitment. Recently, in the run up to the US midterm elections, and last year during the UK's snap elections, digital canvassers and activists took to Tinder to encourage people to vote and engaged in personal conversations in an effort to sway them to their cause (Gorwa & Guilbeault, 2018). In Europe, an increasing number of people are also reporting using WhatsApp – the end-to-end encrypted messaging service – to casually engage in political conversations with friends and family.

As individuals increasingly retreat to private and alternative spaces to discuss or seek out political news, they are also more vulnerable to misinformation. Platforms like WhatsApp, WeChat in China and Telegram in Iran have become a breeding ground for falsehoods, hyperpartisan content and misinformation in recent years. That is because it is becoming easier and easier for politically-motivated

agents and political messaging operations to infiltrate these private services to spread false rumours about political opponents or rival groups.

People are more inclined to believe an article or piece of information if it is shared by a peer, a fellow member of one's own political party, or a respected person within one's social network (Moussaïd, Kämmer, Analytis, & Neth, 2013). Because information on private messaging apps are more likely to come from a known or trusted source, recent research shows that falsehoods and rumours spread much faster on these types of platforms.

If the use of public digital spaces continues to dwindle, this could well spell the end of the digital public sphere as we currently know it. Such a scenario would pose a new set of challenges for policy-makers attempting to remedy the viral spread of misinformation and rumours that continue to erode trust in traditional media and political institutions, whilst also sowing the seeds of radicalisation.

7. Policy principles and options

Freedom of speech and the integrity of political processes lie at the very heart of European democratic values. As politics moves into the digital sphere, emerging technologies have become established in ways that undeniably pose a challenge to core democratic values. Over the course of our inquiry, it has become apparent that technology, and particularly the internet and social media, pose both a tremendous opportunity and risk to political systems in Europe. In a digital age, the implications of technology are far-reaching in society. They span issues related to privacy, information literacy and access, cybersecurity, intellectual property, copyright, election integrity, digital divides and polarisation. Ensuring that its uses and applications are aligned with the public good is pivotal for protecting our democracies, and this challenging aim must therefore concern regulators and policy-makers in Europe.

As catalysts for bridging, expanding and creating divides, technology companies and social media platforms are central to discussions of the polarisation of European publics. As relatively new industries, barely older than a couple of decades, they often remain widely unregulated. Digital innovations and their implications are highly complex and, frequently, existing laws fall short in addressing the intricacies of technological systems. Without specific regulations and principles, the digital sphere is a regulatory 'Wild West'.

The internet's power as an instant source of mass information and communication has wide-reaching implications for political discourse. Our analysis has drawn attention to how system flaws and shortfalls can distort public communication. In addition, we highlight how a myriad of actors are manipulating online ecosystems for political, economic and social gain, placing the integrity of political processes at risk. As reports of interference in democratic systems continue to surface and public concerns apply further pressure, policy-makers encounter an increasingly politicised environment. Against this backdrop, the risks of rushed and short-sighted regulation that may stifle innovation and hinder progress are great.

In this last section of the report, we outline basic mechanisms for the regulation of emerging digital technologies. Rather than suggesting prescriptive solutions, we seek to provide an advisory framework. We highlight overarching ideas and provide direction for regulators to deliberate and expand upon. We believe that sustainable regulation must address the underlying root causes of polarisation in order to effectively address public concerns and political challenges. Rather than providing short-term fixes and patching up symptomatic excoriations, the proposed countermeasures aim to attend to the systemic vulnerabilities of technological infrastructures, and to cure informational maladies on the internet.

7.1. Promoting transparency & accountability

Political communication, public discourse and even governance, have increasingly moved into the digital sphere. Often, online public life unfolds on proprietary platforms principally driven by economic incentives. While modern information and communication technologies offer a forum for the exchange of ideas, our review has raised crucial questions about the freedom and integrity of information on these platforms. Fundamental algorithms remain 'blackboxed', the collection and use of personal user data is opaque, and content moderation frequently lacks public oversight. Furthermore, during moments of crisis – such as interference attempts in elections and referenda, viral conspiracy theories relating to terror attacks and shootings, or data breaches impacting millions of users – policy-makers have criticised the lack of accountability and direct points of contact in technology companies.

Shedding further light on online information ecosystems, their functionalities and core mechanisms is not just a concern for the general public. It is also a concern for political actors and producers of journalistic content who are the traditional gatekeepers and agenda-setters of political information in Europe. As they rely more heavily on social media platforms in order to inform the public, connect with citizens and mobilise voters, they also become subject to algorithmic promotion and filtering processes that lack transparency and oversight.

After reports of electoral interference, social media companies have ramped up efforts to self-regulate. Heightened industry awareness has resulted in laudable and crucial self-correction measures. Yet technology companies are also tasked with the difficulties of content moderation, attribution of cyberattacks and data hacks, tweaking of algorithms for the public good, and the implications of these responsibilities are huge. These tasks generally elude public scrutiny or oversight. Standards and guidelines are rare and thus are frequently self-prescribed. We will now point out three policy options that policy-makers could consider in order to counter these shortfalls.

Policy option 1 – Accountable representation: Governments could oblige large technology companies to be accountable. As digital platforms hold great power over the flow of information, industry actors in countries with significant user bases lack accountable representation. Such representation could extend beyond lobbying and interest groups, and include points of contact devoted to technological mechanisms and activities on the platforms themselves. These could include crisis hotlines for regulators, the wider public, and staff, physically located in those countries with substantial user bases.

Policy option 2 – Transparency reports: Policy-makers could require digital platforms to submit comprehensive annual transparency reports. These could include information about activity on the platform, content moderation and take-downs, financial reports, data collection and brokerage, campaign spending and political advertising. Furthermore, social media platforms could publish general newsfeed statistics, including information about the volume of sponsored and advertising content, the types of content that users are exposed to, and the volume of automated content.

Policy option 3 – Ethics code: Governments could encourage and incentivise tech companies to develop an industry-wide ethical code. The development process could include a diverse cast of industry, government and academic actors. An effective code could include mechanisms for enforcement, oversight and sanctions. This ethical code could inform all product development and platform decisions.

7.2. Governing the digital

Digital communication and information technologies have been adopted by billions of users across the globe. Their networks are at the heart of everyday political life in Europe. These potent platforms are generally no older than a couple of decades. As the digital transformation has overhauled established practices and cultures of political communication and information-seeking, regulation has lagged far behind. Cutting-edge technological advances – such as big data systems, artificial intelligence, and algorithmic decision-making – are so complex that non-professionals are barely able to understand them. The relevant knowledge is also highly specialist and intricate.

It thus comes as no surprise that the digital sphere is still very widely unregulated. The internet was once celebrated as a self-governing marketplace of ideas that would ultimately advance the public good, but it has become evident that careful regulatory intervention and policies are necessary. Some regulatory efforts have sought to enforce existing laws within the digital sphere. Yet faced with the sheer mass of

online information, social media companies have often been left to enforce regulation with little oversight. Existing frameworks are generally ill-suited to digital contexts: campaign blackout laws are for instance difficult, if not impossible, to enforce online.

But the challenges of applying existing regulation, governance measures and diplomacy do not end here. As binding standards and codes of conduct for the digital sphere have on the whole yet to emerge, policy-makers are struggling to effectively scrutinise digital platforms and their activities. The complexity of online ecosystems aggravates these issues, and frequently complicates the applicability of existing laws. Many governments also lack the knowledge and capabilities to regulate political life in the digital sphere. Expert inquiries and government task forces into matters such as digital election interference, artificial intelligence and other implications of technology have become critical instruments of governance. Given the centrality of the digital realm to public life, governments should consider incorporating such knowledge and capabilities.

Policy option 4 – Digital electoral law: Governments must update electoral laws to accommodate digital public life. Standards and regulations relating to campaign spending and information, such as those already in place for print and broadcasting, could be developed. New regulation could also address new tactics and techniques for voter mobilisation, including big data-driven methods, dark advertising and peer to peer messaging. Where violations have occurred, fines and legal consequences must be suitably adapted.

Policy option 5 – Audits: Governments could audit and scrutinise digital platforms to ensure they abide by the law and operate ethically and responsibly. Audits could address cybersecurity, data protection and security, copyright laws, and electoral law. There must be auditing of technological systems (including algorithms, artificial intelligence systems, and data processing mechanisms) as well as of the kind of data that is collected and processed. Auditing processes must go hand in hand with the creation and enforcement of industry standards and ethical codes. In this respect, financial reporting is secondary, and the emphasis is on the organisation and their products.

Policy option 6 – Digital capabilities in governance: Governments could commit to developing their own digital capabilities, and recognising the digital sector. This requires governments to promote digital expertise amongst political leaders, for example by implementing new departments and offices that are tasked with the governance of emerging digital technologies. Policy-making could strive to include voices from technologists and experts.

7.3. Bridging digital divides

Technological accessibility is not always equated with economic prosperity. Physical, functional and capability factors can all undermine or impede the adoption of digital technologies. Users of all economic, educational and age groups frequently experience difficulties navigating the digital public sphere. Often, citizens lack the skills to operate new devices and apps, which come with complex terms and conditions. Faced with the mass of online information, users struggle to select relevant and high-quality information, and younger users especially struggle with managing their screen time efficiently.

Political actors in Europe, too, are facing these challenges. Compared to the United States, politicians in Europe still largely neglect the internet's potential for mobilising voters and connecting with the democratic base over social media. Emerging regulation – like the Network Enforcement Act we have discussed at length – frequently falls short of addressing system flaws and oversimplifies technological complexities. What is more, while populist actors and click-bait publishers solicit public attention with questionable messages and low-quality content, the traditional gatekeepers of information in the

public interest are struggling to find their place online. Quality journalism in particular is struggling with copyright and content monetisation issues.

To combat digital divides, policy-makers must support the accessibility of modern technologies and foster inclusive adoption across European society as a whole. An informed and critical use of digital technologies such as social media must lie at the heart of digitisation processes in Europe. We propose three policy options that are relevant to policy-makers across Europe who wish to promote the use of technology to bridge digital divides for the public good. These options are at the core of policy measures seeking to advance our understanding of polarisation insofar as it relates to digital technologies.

Policy option 7 – Infrastructures: Policy-makers could consider investing in internet accessibility. In Europe, the accessibility of the internet varies substantially by country and region. This includes basic internet access, but also the bandwidth and speed of internet services. We urge governments to support investments in broadband internet across Europe. We also suggest that that public wireless internet hotspots could be installed in cities across Europe. Lastly, policy-makers could consider measures that increase internet accessibility for lower-income citizens.

Policy option 8 – Literacy: We encourage digital literacy campaigns across Europe, which could be aimed at all age groups. Digital literacy must become a fundamental pillar of education. Literacy measures could provide knowledge and practical capabilities for critical thinking in relation to information on the internet, working with digital devices and big data. Curricula could also focus on artificial intelligence and big data and provide a framework for understanding code language. Digital literacy could be taught at schools and universities. In addition, literacy campaigns and classes aimed at reaching the wider public should urgently be considered.

Policy option 9 – Research: Governments could support research into the mechanisms and processes by which polarisation emerges, and its relationship to digital technologies. A core part of this work could be concerned with research into online information dynamics, algorithms and political discourse over social media. We suggest that governments could work with groups of academic and industry experts, and foster the rigorous exchange of relevant data that is often made unobtainable by technology companies. The results could be shared with the public in an accessible and engaging manner.

8. Conclusion

The forces driving polarisation in society are multifarious and complex. Deep-seated political, religious and social cleavages as well as feelings of resentment, inequality, distrust and injustice have worked together in recent years to erode social cohesion in European publics. In parallel, we have seen the meteoric rise of communication technologies that have lowered the barrier of access to information and accelerated connectivity between international actors.

This report has provided an overview of how such digital technologies can be mobilised by political actors, campaigns and movements to motivate action or influence political opinion. As we have seen, complex information infrastructures are vulnerable and prone to design failures, which can be readily exploited by bad actors – both domestic and foreign – for political and economic gain. Recent events have also highlighted the levels of sophistication with which news technologies and analytical methods can be deployed to increase anti-immigration sentiment or amplify nationalist rhetoric among some communities.

Technology today has emerged as a public interest issue. But if the initial idealism surrounding the democratic possibilities that social media would usher has waned – replaced by concerns that these technologies are threatening the health of both advanced and unstable democracies – as we have shown throughout this report, by going beyond simple dichotomies between cyber optimism and cyber pessimism, it is possible to envisage the new possibilities that digital technologies offer for productive engagement and debate amongst citizens, including new forms political participation.

In the future, we expect these mechanisms, from targeted messaging to digital political strategies and computational tools, to evolve, reflecting technological progress, and changes in communication cultures. The rapid pace of these advances will undoubtedly pose new challenges for governments and policy-makers. But as recent efforts to fashion healthier communities, influence positive political behaviour through user-centred design, implement contextual and strategic regulations, and provide the public and policy-makers with enough context and resources to understand contemporary threats to democracy and curb the reach of extremism are beginning to pay off, these prospects also offer hope about our collective potential to realise the internet's original vision: engage, inform, and empower citizens across the globe.

References

- Abramowitz, A. (2010). *The disappearing center: engaged citizens, polarization, and American democracy*. Yale University Press.
- Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election. In *Proceedings of the 3rd international workshop on Link discovery - LinkKDD '05* (pp. 36–43). <https://doi.org/10.1145/1134271.1134277>
- Ahlers, D. (2006). News consumption and the new electronic media. *Harvard International Journal of Press/Politics*, 11(1), 29–52.
- Alandete, D. (2017). Catalan referendum: Pro-Russian networks see 2,000 % increase in activity in favor of Catalan referendum, *EL PAÍS*.
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236.
- Amnesty International. (2018). Amnesty Reveals Alarming Impact of Online Abuse Against Women. *Amnesty International Ltd*.
- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The 'Nasty Effect:' Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication*, 19(3), 373–387.
- Asker, D., & Dinas, E. (2017). Do Online Media Polarize ? Evidence from the Comments' Section. *SSRN Electronic Journal*, (August).
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Barberá, P. (2014). How Social Media Reduces Mass Political Polarization . Evidence from Germany, Spain, and the U.S. *LXXIII Congress of the Midwest Political Science Association*, 44.
- Beer, D. (2017). The social power of algorithms. *Information Communication and Society*, 20(1), 1–13.
- Bimber, B., Flanagin, A., & Stohl, C. (2012). *Collective Action in Organizations : Interaction and Engagement in an Era of Technological Change*. Cambridge University Press.
- Binder, S. (2003). *Stalemate*. Brookings Institution Press
- Bode, L., & Vraga, E. K. (2018). Studying Politics Across Media. *Political Communication*, 35(1), 1–7.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415).
- Borah, P. (2013). Interactions of News Frames and Incivility in the Political Blogosphere: Examining Perceptual Outcomes. *Political Communication*, 30(3), 456–473.
- Borah, P. (2014). Does It Matter Where You Read the News Story? Interaction of Incivility and News Frames in the Political Blogosphere. *Communication Research*, 41(6), 809–827.
- Boyd, D. (2018). Media Manipulation, Strategic Amplification, and Responsible Journalism – Data & Society: Points. *Medium*.
- Boyd, D. (2017). Hacking the Attention Economy – Data & Society: Points. *Medium*.
- Boyd, D., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(28), 7313–7318.
- Bright, J. (2016). The Social News Gap: How News Reading and News Sharing Diverge. *Journal of Communication*, 66(3), 343–365.

- Brundidge, J. (2010). Encountering 'Difference' in the contemporary public sphere: The contribution of the internet to the heterogeneity of political discussion networks. *Journal of Communication*, 60(4).
- Bundesministerium der Justiz und fuer Verbraucherschutz. (2017). Aktuelle Gesetzgebungsverfahren | Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG).
- Cambridge Analytica. (n.d.). 'About Us' [webpage]
- Castells, M. (2000). *The rise of the network society*. Blackwell Publishers.
- Castells, M. (2005). *Communication power*. Oxford: Oxford University Press.
- Chadwick, A. (2013). *The hybrid media system : politics and power*. Oxford University Press.
- Chira, S. (2018). The Women's March Became a Movement. What's Next? *The New York Times*.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, 64(4), 658–679.
- Confessore, N., Dance, G. J. X., Harris, R., & Hansen, M. (2018). The Follower Factory. *The New York Times*.
- da Silva, M. T. (2013). Online forums, audience participation and modes of political discussion: readers' comments on the Brazilian presidential election as a case study/Foros en línea, participación del público y modos de la discusión política: comentarios de los lectores sobre. *Comunicación y Sociedad*, 26(4), 175–193.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3), 554–9. h
- DemocracyOS. (n.d.). OpenCOP21 : the online citizens' COP report. *Medium*.
- Der Spiegel (2017) Trolls in Germany: Right-Wing Extremists Stir internet Hate. *SPIEGEL ONLINE*
- Dijck, J. van. (2013). *The culture of connectivity : a critical history of social media*. Oxford University Press.
- DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's Social Attitudes Become More Polarized? *American Journal of Sociology*, 102(3), 690–755.
- Dobber, T., & Helberger, N. (2017). Is political micro-targeting hijacking European democracy? *Internet Policy Review*.
- Dommett, K., & Temple, L. (2018). Digital Campaigning: The Rise of Facebook and Satellite Campaigns. *Parliamentary Affairs*, 71(1)
- DOMO. (n.d.). *Data Never Sleeps 5.0* [report]
- Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*. The University of Chicago Press.
- Dutton, W. H., Reisdorf, B. C., Dubois, E., & Blank, G. (2017). Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States. *SSRN Electronic Journal*.
- Edelman. (2017). *Trust Barometer* [report]
- Edmonds, T., & Elder, C. (2015). 2015 really was the first digital general election: here are 7 lessons you should know. *Telegraph*.
- Ee, S. (2018). Updating Government: Estonia's Mastery of Digital Statecraft. *Atlantic Council*.
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, 12(4), 1143–1168.
- Erbentraut, J. (2015). Here's The Actual Impact All Those Buckets Of Ice Had On ALS Research. *Huffington Post*.
- European Union. (2016). Code of conduct on countering illegal hate speech online. *European Commission*.
- Eurostat. (2017). *Digital economy & society in the EU* [report]
- Eurostat. (2018). *Digital economy & society in the EU* [report]

- Evans, J. (2018). The techlash. *TechCrunch*.
- Facebook Newsroom. (2017). 'Company Info' [webpage]
- Fischer, C. S. (1992). *America calling : a social history of the telephone to 1940*. University of California Press.
- Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of ' fake news ' and online disinformation in Europe. *Reuters Institute*.
- Fletcher, R., & Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, 67(4), 476–498.
- Funke, D. (2017a). In a step toward automation, Full Fact has built a live fact-checking prototype. *Poynter*.
- Funke, D. (2017b). Online hoaxes about the Catalan referendum concentrated on police action in Barcelona. *Poynter*.
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, 126(4), 1799–1839.
- Ghosh, D., & Scott, B. (2018). Digital Deceit: The Technologies Behind Precision Propaganda on the internet. *New America*.
- Gibson, R. K. (2015). Party change, social media and the rise of 'citizen-initiated' campaigning. *Party Politics*, 21(2), 183–197.
- Gil de Zúñiga, H., Veenstra, A., Vraga, E., & Shah, D. (2010). Digital Democracy: Reimagining Pathways to Political Participation. *Journal of Information Technology & Politics*, 7(1), 36–51.
- Gillespie, T. (2014a). Media Technologies: Essays on Communication, Materiality, and Society. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press.
- Gillespie, T. (2014b). The Relevance of Algorithms. In *Media Technologies* (pp. 167–194). The MIT Press.
- Gillespie, T. (2018). Moderation Is The Commodity. *Techdirt*.
- Google. (2018). 'Perspective' [webpage]
- Gorwa, R. (2018). Facebook's moderation task is impossibly large. It needs help. *WIRED*.
- Gorwa, R., & Guilbeault, D. (2018). Understanding Bots for Policy and Research: Challenges, Methods, and Solutions. *Policy & Internet*.
- Grassegger, H. (2018). Facebook says its 'voter button' is good for turnout. But should the tech giant be nudging us at all? *The Guardian*.
- Grassegger, H., & Krogerus, M. (2016). Ich habe nur gezeigt, dass es die Bombe gibt. *Das Magazin*.
- Hagemann, C. (2002). Participation in and contents of two Dutch political party discussion lists on the internet. *Javnost*, 9(2), 61–76.
- Halpern, D., & Sanders, M. (2016). *Nudging by government: Progress, impact and lessons learnt*. Behavioral Policy (Vol. 2).
- Hamann, G. (2018). NetzDG: Der Storch-Effekt. *Die Zeit*.
- Hasell, A., & Weeks, B. E. (2016). Partisan Provocation: The Role of Partisan News Use and Emotional Responses in Political Information Sharing in Social Media. *Human Communication Research*, 42(4), 641–661.
- Hedman, F., Sivnert, F., Kollanyi, B., Neudert, L.-M., & Howard, P. N. (2018). *News and Political Information Consumption in Sweden: Mapping the 2018 Swedish General Election on Twitter*.
- Heller, N. (2017). Estonia, the Digital Republic. *The New Yorker*.
- Henderson, J. A., & Theodoridis, A. G. (2017). Seeing Spots: Partisanship, Negativity and the Conditional Receipt of Campaign Advertisements. *Political Behavior*, 1–23.
- Herman, J. (2016). 'Brexit' Talk on Social Media Favored the 'Leave' Side. *The New York Times*.

- Hetherington, M. J. (2008). Turned Off or Turned On? How Polarization Affects Political Engagement. In *Red and Blue Nation? Consequences and Correction of America's Polarized Politics* (pp. 1–51).
- Holznagel, B. (2017). Stellungnahme zum Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [webpage]
- Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving Google's Perspective API Built for Detecting Toxic Comments. *arXiv: 1702.08128v1*
- Howard, P. N. (2006). *New media campaigns and the managed citizen*. Cambridge University Press.
- Howard, P. N., & Hussain, M. M. (2013). *Democracy's fourth wave? Digital media and the Arab Spring*. Oxford University Press.
- Howard, P. N., & Parks, M. R. (2012). Social Media and Political Change: Capacity, Constraint, and Consequence. *Journal of Communication*, 62(2), 359–362.
- Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81–93.
- Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*, 35(2), 441–456.
- Hutt, R. (2017). The world's most popular social networks, mapped. *World Economic Forum*.
- Isenberg, D. J. (1986). Group Polarization. A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology*, 50(6), 1141–1151.
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3): 405–431.
- Jhaver, S., Vora, P., & Bruckman, A. (2017). *GVU Technical Report Designing for Civil Conversations: Lessons Learned from ChangeMyView*.
- John, N. A., & Dvir-Gvirsman, S. (2015). I Don't Like You Any More: Facebook Unfriending by Israelis During the Israel-Gaza Conflict of 2014. *Journal of Communication*, 65(6), 953–974.
- Jones, D. R. (2015). Declining trust in congress: Effects of polarization and consequences for democracy. *Forum (Germany)*, 13(3), 375–394.
- Justice, U. D. of (2018). *USA v Internet Research Agency*. Washington, D.C.
- Kalla, J. L., & Broockman, D. E. (2018). The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments. *American Political Science Review*, 112(01), 148–166.
- Kanter, J. (2018). Cambridge Analytica boasts it helped Trump win the US election. *Business Insider Deutschland*.
- Karpf, D. (2012). *The MoveOn effect : the unexpected transformation of American political advocacy*. Oxford University Press.
- Klonick, K. (2017). The New Governors: The People, Rules, and Processes Governing Online Speech. *SSRN: 131 Harv.L.Rev. 1598*
- Kreiss, D. (2017). Micro-targeting, the quantified persuasion. *Internet Policy Review*, 6(4).
- Kreiss, D., & McGregor, S. C. (2018). Technology Firms Shape Political Communication: The Work of Microsoft, Facebook, Twitter, and Google With Campaigns During the 2016 U.S. Presidential Cycle. *Political Communication*, 35(2), 155–177.
- Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News Sharing in Social Media: A Review of Current Research on News Sharing Users, Content, and Networks. *Social Media + Society*, 1(2).
- LaRue, F. (2011). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. *Human Rights Council Report* (Vol. 3).
- Lessig, L. (2004). *Free Culture: The Nature and Future of Creativity*. Penguin Press
- Lessig, L. (2008). *Remix : making art and commerce thrive in the hybrid economy*. Penguin Press.

- Levendusky, M. (2013). Partisan Media Exposure and Attitudes Toward the Opposition. *Political Communication*, 30(4), 565–581.
- Levy, D. A., & Nielsen, R. (2010). *The Changing Business of Journalism and its Implications for Democracy*. Reuters Institute
- Lorenz, T. (2018). Teens Debate Big Issues on Instagram Flop Accounts. *The Atlantic*.
- MacKenzie, D. A., & Wajcman, J. (1999). *The social shaping of technology*. Open University Press.
- Mantzaris, A. (2016). Fact-checking doesn't 'backfire', new study suggests. *Poynter*
- Margetts, H., John, P., Hale, S., Yasseri, T., & John, P. (2015). *Political turbulence: how social media shape collective action*. Princeton University Press.
- Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). The digital transformation of news media and the rise of disinformation and fake news. *European Commission*.
- Marwick, A., & Lewis, R. (2017). Media Manipulation and Disinformation Online. *Data & Society Research Institute*.
- McClenaghan, M. (2017). The 'dark ads' election: How are political parties targeting you on Facebook? – *The Bureau of Investigative Journalism*.
- Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media. *Communication Research*, 41(8), 1042–1063.
- Miller, M. (2017). Protesters outside White House demand 'Pizzagate' investigation. *The Washington Post*.
- Moore, M., & Ramsay, G. (2017). UK media coverage of the 2016 EU Referendum campaign About the Centre for the Study of Media, Communication and Power. *King's College London*.
- Mosseri, A. (2018). 'Bringing People Closer Together | Facebook Newsroom' [webpage]
- Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLoS ONE*, 8(11), e78433.
- Mutz, D. C. (2006). *Hearing the other side: Deliberation versus participatory democracy*. Cambridge University Press.
- Neudert, L.-M. (2018). Future elections may be swayed by intelligent, weaponized chatbots. *MIT Technology Review*.
- Neudert, L.-M., Kollanyi, B., & Howard, P. N. (2017). Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter? *Project on Computational Propaganda, Oxford University*.
- Neudert, L.-M. N. (2018). Computational Propaganda in Germany - A Cautionary Tale. In S. Woolley & P. N. Howard (Eds.), *Computational propaganda: political parties, politicians, and political manipulation on social media*. Oxford University Press.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A., & Nielsen, R. (2018). Reuters Institute Digital News Report 2018. *Reuters Institute*.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Nielsen, R. K., & Levy, D. (2017). *Reuters Institute Digital News Report 2017*. Reuters Institute.
- Newton, C. (2018). Facebook makes its community guidelines public and introduces an appeals process - *The Verge*.
- Nicas, J. (2018). How YouTube Drives People to the Internet's Darkest Corners. *The Wall Street Journal*.
- Nielsen, R. K. (2012). *Ground Wars: Personalized Communication in Political Campaigns*. Princeton
- Nielsen, R. K., & Kuhn, R. (2014). The Many Crises of Western Journalism: A Comparative Analysis of Economic Crises, Professional Crises, and Crises of Confidence.
- Noble, S. U. (2018). *Algorithms of oppression: how search engines reinforce racism*. NYU Press.

- Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32(2), 303–330
- O'Neil, C. (2017). The Ivory Tower Can't Keep Ignoring Tech. *The New York Times*.
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media and Society*.
- Pariser, E. (2011). *The filter bubble : what the internet is hiding from you*. Penguin Press.
- Phillips, W. (2018). The Oxygen of Amplification. *Data & Society*.
- Pisani-Ferry, J. (2015). Responding to Europe's Political Polarization. *Project Syndicate*.
- Polonski, V. (2016). Impact of social media on the outcome of the EU referendum. *EU Referendum Analysis*
- Reporters Without Borders. (2017). Russian bill is copy-and-paste of Germany's hate speech law [webpage]
- Reporters Without Borders. (2018). The Network Enforcement Act apparently leads to excessive blocking of content [webpage]
- Reuter, M. (2017) Reporter ohne Grenzen: Netzwerkdurchsetzungsgesetz ist gefährlicher Schnellschuss, *Netzpolitik*
- Rotman, D., Vieweg, S., Yardi, S., Chi, E., ... J. P.-C. E. (2011). From slacktivism to activism: participatory culture in the age of social media. *DI.Acm.Org*.
- Sabbagh, D. (2018). Rise of digital politics: why UK parties spend big on Facebook. *The Guardian*
- Schärdel, J. (2016). „Ein Schelm, wer Böses dabei denkt'. *Zeitschrift Für Vergleichende Politikwissenschaft*, 10(2), 131–163.
- Schroeder, R. (2018). Rethinking digital media and political change. *Convergence: The International Journal of Research into New Media Technologies*, 24(2), 168–183.
- Shirky, C. (2008). *Here comes everybody : the power of organizing without organizations*. Penguin Press.
- Silverman, C. (2016). How The Bizarre Conspiracy Theory Behind 'Pizzagate' Was Spread. *BuzzFeed*.
- Singer, J. B. (2014). User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society*, 16(1), 55–73.
- Smith, A., & Duggan, M. (2016). Americans, Politics and Social Media. *Pew Research Center*.
- Stone, H. S. (1973). An Efficient Parallel Algorithm for the Solution of a Tridiagonal Linear System of Equations. *Journal of the ACM*, 20(1), 27–38.
- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press.
- Suwajanakorn, S., Seitz, S., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: learning lip sync from audio. *DI.Acm.Org*.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge : improving decisions about health, wealth and happiness*. Penguin Books.
- Timberg, C., & Dwoskin, E. (2018). Exclusive: Twitter is suspending millions of bots and fake accounts every day to fight disinformation. *The Washington Post*.
- Trilling, D., Tolochko, P., & Burscher, B. (2017). From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics. *Journalism & Mass Communication Quarterly*, 94(1), 38–60.
- Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., ... Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature.
- Tufekci, Z. (2018). YouTube, the Great Radicalizer. *The New York Times*.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Wang, S. (2018). A 2018 survey of fact-checking projects around the world finds a booming field, with at least 149 dedicated initiatives. *Nieman Journalism Lab*.

- Weber, M.-T. (2017). Bitkom Stellungnahme Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken der Fraktionen von CDU/CSU und SPD (BT DS 18/12356). *Bitkom*
- Webster, J. G., & Ksiazek, T. B. (2012). The Dynamics of Audience Fragmentation: Public Attention in an Age of Digital Media. *Journal of Communication*, 62(1), 39–56.
- Woolley, S. C., & Howard, P. N. (2016). Political communication, computational propaganda, and autonomous agents: Introduction. *International Journal of Communication*, 10, 4882–4890.
- Woolley, S., & Howard, P. N. (2018). *Computational propaganda: political parties, politicians, and political manipulation on social media*. Oxford University Press
- Wu, T. (2016). *The attention merchants: the epic scramble to get inside our heads*. Atlantic Books.
- Ziegele, M., Breiner, T., & Qiring, O. (2014). What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. *Journal of Communication*, 64(6), 1111–1138
- Zittrain, J. (2014). Engineering an Election. *Harvard Law Review*.
- Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., ... Preis, T. (2015). Emotional dynamics in the age of misinformation. *PLoS ONE*, 10(9).

Digital technology was once heralded as a boon for democracy. However, the contemporary political reality has underscored the ways in which technology can also undermine citizenship, democracy and the international liberal world order. Thriving on new media ecosystems, populist voices in European democracies have established themselves as mainstream political actors.

This report offers a comprehensive overview of the relationship between technology, democracy and the polarisation of public discourse. Technology is inherently political, and the ways in which it is designed and used have ongoing implications for participation, deliberation and democracy. Algorithms, automation, big data analytics and artificial intelligence are becoming increasingly embedded in everyday life in democratic societies; this report provides an in-depth analysis of the technological affordances that enhance and undermine political decision-making, both now and in the future. To conclude, we formulate principles and policy options for fostering a better relationship between digital technology and public life.

This is a publication of the Scientific Foresight Unit (STOA)
EPRS | European Parliamentary Research Service

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.



ISBN 978-92-846-3614-3 | doi: 10.2861/03717 | QA-03-19-071-EN-N