ABSTRACT

Social media is increasingly used to spread misinformation, so-called fake news, potentially threatening democratic processes and national interests. Developments in the area of artificial intelligence (AI) have rendered this trend even more pronounced. This chapter looks at social media and how they moved from being regarded as a positive force in the public sphere to a negative one. AI will be explained, and its potential in the context of social media will be decrypted. Consequently, AI, social media, and fake news will be analysed and the apparent threat to democratic functions explored. Finally, this chapter proposes how AI can also be used to safeguard democracy.

Keywords: Artificial intelligence, fake news, social media
Introduction

Approximately ten years ago, it was said that social media would restore the power to citizens, particularly to consumers (Kaplan & Haenlein, 2010). Information can rapidly be disseminated on platforms such as Twitter (Kaplan, 2012; Kaplan & Haenlein, 2011), Facebook, and the like. Using such platforms, democracy could be experienced more directly and in a more participatory manner. For example, during the Arab spring – a series of anti-government protests, uprisings, and armed rebellions against oppressive regimes that spread across North Africa and the Middle East in late 2010 – social media played a determinant role by facilitating communication and interaction among participants of these protests.

However, within just a decade, social media – newly powered by artificial intelligence and big data – went from being a facilitator of democracy to a serious threat of the same, most recently with Facebook: Through the Cambridge Analytica data scandal, the world understood the power of these tools to undermine democratic mechanisms. The political consultancy Cambridge Analytica used several million Facebook users’ data to successfully influence and manipulate public opinion in such events as the 2016 US presidential election and the 2018 Brexit referendum. This consequently created an outcry and public discussion on ethical standards for social media companies, data protection, and the right to privacy.

Social media are indeed increasingly used to spread targeted misinformation, or so-called fake news, in order to manipulate entire groups of people. The rapid developments in the area of artificial intelligence (Haenlein & Kaplan, 2019; Kaplan & Haenlein, 2019; 2020) in particular, and the digital sphere in general, will render this trend even more pronounced. Instead of fake news via text only, in the future everyone will be able to produce videos where one can insert one’s own words into another’s speech, making the latter appear to say things which s/he never would have said in reality. Actually such deepfakes already exist. Imagine Photoshop for audio and video content. Just about anybody might and will be able to create videos where people seemingly say something that they never actually uttered.

This chapter firstly takes a brief look at social media and how they moved from being a positive force to becoming a negative one. Also, artificial intelligence will be briefly explained, and its potential in the context of social media will be decrypted. In a second section, how artificial intelligence, social media, and fake news represent a danger to democracy will be discussed, as well as the various ways they are applied for the purpose of undermining democratic mechanisms. Thirdly, this chapter shows how artificial intelligence can also be used to safeguard democracy. It also gives insights into how to fight fake news, deepfakes,
or simply targeted misinformation. This chapter concludes with food for thought as to what a future might look like where AI potentially dominates politics.

**Social Media Powered by Artificial Intelligence and Big Data**

Social media are defined as “a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content” (Kaplan & Haenlein, 2010, p. 61). They can be classified into collaborative projects (e.g., Wikipedia; Kaplan & Haenlein, 2014), micro-blogs/blogs (e.g., Twitter), content communities (e.g., YouTube), social networks (e.g., Facebook), virtual game worlds (e.g., *World of Warcraft*), and virtual social worlds (e.g., *Second Life*; Kaplan & Haenlein, 2009). They have doubtlessly begun to play a significant part in all sectors, from business to education, and from public administration to politics (Kaplan, 2017, 2015, 2014b).

Entertainers such as Britney Spears have built their communication strategies entirely around social media (Kaplan & Haenlein, 2012). In academia, social media are increasingly integrated into courses (Kaplan, 2018; Kaplan & Haenlein, 2016; Pucciarelli & Kaplan, 2016). Several public administrations make use of Facebook, Twitter, and the like, for example the European Union actively makes use of social media with the objective of creating a sense of European identity among its approximately half a billion citizens (Kaplan, 2014a). Finally, in politics, social media have been part of the game for more than a decade. Social media communications were a key element in Barack Obama’s presidential campaign, which led to his first election in 2008.

At their advent, social media were considered an opportunity for democratic mechanisms, a booster for democracy, and a source of citizen empowerment (Deighton et al., 2011). They still are, as exemplified by the #MeToo movement against sexual harassment and sexual assault which was begun solely by individuals and rapidly went viral with the help of social media. Public administrations use social media to interact with their citizens, to foster citizen participation and collaboration, to increase transparency and information dissemination, and much more. For example, in 2018 when the UK Royal Navy wanted to increase public awareness its role it created several Instagram stories wherein Lieutenant Matt Raeside responded to various queries concerning work conditions and the recruitment process. Undemocratic regimes have been challenged by their citizens who have resorted to social media to voice their disapproval and to organize demonstrations, even entire revolutions. In authoritarian regimes the regular media is ordinarily supervised by the state, and thus it is usually impossible to disseminate critique thereby. Yet via social media it has become possible
to do so, as was observed when the Arab Spring began in Tunisia, and Facebook, Twitter, and others enabled the organizing of mass protests, finally leading to dictator and president Zine El Abidine Ben Ali being forced into exile.

However, with the advent of artificial intelligence and big data, social media have increasingly evolved toward constituting a potential threat to democracy. Artificial intelligence (AI), defined as “a system’s ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (Kaplan & Haenlein, 2019, p. 17), can be divided into three types: analytical, human-inspired, and humanized (Kaplan & Haenlein, 2019). Analytical AI has characteristics consistent with cognitive intelligence only and such a system could learn the campaign platforms of various parties and respond to questions from citizens with respect to the contents. Human-inspired AI has elements of cognitive and emotional intelligence, i.e., understanding human emotions, in addition to cognitive elements, and uses these in its decision making. Such a system could use facial expression to detect when a citizen appears to have problems understanding a party’s platform, and can providing him or her with more information. Humanized AI exhibits characteristics of all types of competencies (i.e., cognitive, emotional, and social intelligence), and is able to be self-conscious and self-aware in interactions with others. A humanized AI system could actually have a full-fledged discussion with the interested citizen, self-reflect on its own opinions, and form its own ideas about the various parties, including which one would best represent its own interests.

Within just a couple of years, the early promise of the internet and then of the social media revolution to provide a more transparent, democratic, and informed world devolved into an online environment where one cannot be sure of what is true and what is false. Russian interference in the 2016 US presidential election is broadly known. Also in several other elections such as those of Austria, Belarus, Bulgaria, France, Germany, and Italy, there appears to be evidence of Russia manipulating voter outcome via fake news and misinformation posted on social media (Kamarck, 2018). With respect to Brexit, data scientists at the universities of Berkeley and Swansea discovered that more than 156,000 Russian Twitter accounts were used to disrupt the Brexit vote. During the last two days of the referendum alone, more than 45,000 such tweets were posted (Mostrous, Bridge, & Gibbons, 2017).

It is not only Russia which is making unethical use of social media. The far-right candidate of Brazil’s 2018 presidential elections Jair Bolsonaro appears to have won partly due to having benefited from a disinformation campaign on WhatsApp (Tardáguila, Benevenuto, & Ortellado 2018). Creative yet fake pictures showed the opponent’s party members posing
with Fidel Castro with the objective of positioning opposition leaders as radicals. Within a short period – perhaps a decade or so – social media went from being the tool which threw off the tyranny of state-controlled media and facilitated anti-regime demonstrations to being the platforms used by those same regimes to destroy or at least destabilize democracy in the West.

**Threatening Democracy: Supervision, Manipulation, Frustration**

Artificial intelligence represents a threat to democracy which needs to be taken seriously. There are at least three areas in which AI might be a danger to democratic life and mechanisms: supervision, manipulation, and frustration. Firstly, states now have very advanced means of controlling and supervising their citizens’ daily behaviour, which could be abused by governments to limit freedoms. Secondly, citizens can increasingly be manipulated in their voting behaviour by ample use of artificial intelligence and social media. Thirdly and finally, such AI-driven supervision and manipulation can, in short, lead to citizens’ frustration and their deciding to no longer take part in democracy.

**Supervision**

With respect to supervision, we will illustrate AI’s possibilities using the example of China, whose government broadly embraces artificial intelligence technology in order to track, monitor, and control its citizens. The Chinese government collects individuals’ big data from a variety of sources such as finance, tax, and health records. It monitors online purchasing behaviour, social media activity and information resulting from facial recognition via the country’s approximately 200 million surveillance cameras. This data is then used to calculate an individual’s “social credit score” which is supposed to give incentive to lawful behaviour and good citizenship. The score goes up for good behaviour such as donating blood, volunteering in a hospital, or repaying your loan on time. The opposite happens if you get a speeding ticket, if you fail to pay your taxes, or if you drop litter. Consequences of bad scores are non-eligibility for jobs in public administration, being turned down for loans, or even not being allowed to board an airplane. Good scores, in contrast, provide exemplary citizens with discounts on their electricity bills or a privileged access to health care (Marr, 2019).

Some might say that such a system is specific to China, and it is certainly correct that the Chinese credit score system is regime specific. However, other countries make use of artificial intelligence for surveillance purposes. The US police apply AI in order to predict where to send police to patrol as it can predict potential violent crime occurring. The algorithm even provides names of individuals who most likely will become either victims or perpetrators.
Consequently, the police might be asked to visit these identified individuals to warn them against engaging in illegal acts. The problem with such algorithms is that they are only as good as the data used to train them. If such data is biased, then so will the algorithm be in its suggestions and conclusions. Unfortunately, this was the case in the aforementioned application of AI, leading to undesirable results, such as identifying African-Americans as more prone to engage in illegal behaviour than other ethnic groups (Portilho, 2019). These examples show how much AI potentially threatens freedom in any democracy, even more so than in other regime types.

**Manipulation**

In addition to surveillance possibilities, AI-powered social media can be heavily abused and can become the tool which manipulates voting behaviour and beyond. Social media function on the basis of algorithms that provide content to each individual according to her behaviour on these social media platforms. Ultimately, we only see what we most likely enjoy seeing, i.e., if you're more into cats than dogs, you will most likely “like” more cat pictures than those of dogs on Facebook. The algorithm will learn this and consequently expose you to cat pictures only. What works with animals obviously also works for political views. Moreover, at some point you will get the impression that everybody else thinks as you do, as you will not be exposed to other opinions.

This in itself is obviously not conscious and targeted manipulation, at least not yet. But social media allow for hypertargeting, i.e., they feed very detailed information to one specific group of people. A political party could, for example, stress one of their campaign issues without giving the user the full picture. A party in favour of gun control and LGBTQ rights could, for example, provide a group of Facebook users who are in favour of stricter gun laws but not necessarily supporters of more LGBTQ rights, with information on its plans for gun restrictions only. As such, you can feed more or less everybody what they want to hear and read. As many people no longer watch TV news or read a newspaper, the only thing they would know about this specific party is its plans on gun control – which might be enough for them to vote for it.

The aforementioned case is a matter of unbalanced information and disclosing only part of the picture. However, there is also the aforementioned fake news, where candidates or their campaigns spread false information on their opponents on social media. If you see such misleading and false information on your Facebook feed and nothing else, at some point you will believe it, as you never or rarely see and read a counter-argument or a post citing it as
fake news. Thus in the three months prior to the 2016 US presidential election, the top 20 fake news stories on Facebook resulted in more comments, shares, and likes than did the top 20 news stories from 19 major news sources combined (including *The Huffington Post*, *The New York Times*, and *The Washington Post*). Three quarters of those who read such fake news believed it to be true (Silverman, 2016).

This already shows the manipulative potential of fake news in text format. Now imagine so-called deepfakes, i.e., AI-based technology used to produce or alter audio or video content so that it presents something that did not, in fact, occur. With this technology it is possible to seemingly have Donald Trump say that he wants to ban all firearms in the US; or the Pope say that he is in favour of same-sex marriage. An AI-driven system could thus learn during a phone call the respective voter’s preferences and adapt in real-time the conversation accordingly, using the voice of a politician, a celebrity, even a relative of yours pretending to be them. This could clearly lead to extreme manipulation in the future.

**Frustration**

A third effect of the abuse of AI on democracy is the frustration of citizens who at some point become non-voters and will no longer participate in political or democratic life. Due to no longer knowing what the truth is and what is fake, people might abandon participation completely. The more deceptions that occur and the more difficult the verification of content becomes, the more likely that people’s trust in their institutions will continue to decrease.

An illustration of this can be given again in the case of the 2016 US presidential election. Some analyses show that Donald Trump won the election not because of more non-college-educated Caucasians voting than in preceding elections (as was often declared the reason for his victory), but due to a decrease in African-American voter turnout (Kamarck, 2018). The reason for some of this decrease can certainly be explained by Hillary Clinton not being Barack Obama. Nonetheless, one also could argue that African-American voter turnout should have stayed at least the same as in previous elections, as Trump’s campaign should have made them vote for Clinton. A plausible explanation for this decline is provided in Robert Mueller’s report, where one can read, “The Russians allegedly masqueraded as African-American and Muslim activists to urge minority voters to abstain from voting in the 2016 election or to vote for a third-party candidate” (Mosk, Turner, & Faulders, 2018). Fake African-Americans’ social media accounts persuaded actual African-Americans to “Choose peace and vote for Jill Stein. Trust me: It’s not a wasted vote” or pointed out, “We’d surely be better off not voting AT ALL” (Mosk, Turner, & Faulders, 2018).
Protecting Democracy: Technology, Regulation, Education

The previous section clearly shows that artificial intelligence definitively represents a threat to democracy. Next, three areas will be discussed that could help to avoid misuse of AI and to limit its danger to democracy: technology, regulation, and education. Technology and AI itself can be applied to detect unethical and illegal behaviour. Regulation will be necessary and can define voter manipulation and the dissemination of fake news. Finally, it will also be a question of how to educate (future) citizens and make them more conscious of the various manipulation techniques.

Technology

Just as artificial intelligence can be used to hyper-target individuals with fake news and the production of deepfakes, it can also detect them. AI can, for example, reverse-engineer the potentially manipulated data and restore video, audio, and pictures to their original states. AI can also be useful in detecting where the fake news originated by searching the cybersphere for similar content and thus drawing conclusions of who/where the author(s) might be located.

Much of such technology can already be found and is already being implemented: Facebook has hired third-party, independent fact checkers searching for misinformation. WhatsApp limits forwarding text to five recipients at a time in order to deliberately slow the potential dissemination of false information. Microsoft applies AI to develop trustworthy algorithms potentially applied for the detection of fake news. Services such as Storyful help to identify the truthfulness of trending stories across various social media, using journalists’ and professional investigators’ toolsets. TinEye, another online service, enables the verification of whether pictures are real or manipulated by applying reverse engineering and crawling through search engines to potentially identify full or partial matches of such pictures.

As the likely quantity of fake news will be impossible for humans to identify and analyse rapidly enough, AI technology will certainly become key in detecting and identifying it. This is also the opinion of the European Commission, as can be read in their communiqué, “Tackling online disinformation: A European approach” (European Commission, 2018). However, the Commission also warns that such an automated approach could also lead to falsely labelling truthful content as disinformation. Moreover, it would provide potential for undermining freedom of expression and, as such, democracy and democratic mechanisms.

Regulation

The use of artificial intelligence in combination with social media in particular, but also more broadly, must certainly be regulated. However, this is not an easy endeavour for several
reasons: Current regulation is not adapted to the challenges that AI poses, and new regulation often represents a trade-off for innovation. Regulation in some cases hinders and slows down creativity and invention. Moreover, regulation of fake news and misinformation can lead to censorship and restricted freedom of speech. In the words of political scientist Darrell West (2017), “Overly restrictive regulation of internet platforms in open societies sets a dangerous precedent and can encourage authoritarian regimes to continue and/or expand censorship”, which would again obviously constitute a danger to democracy. Finally, the expertise of designing good regulation is simply missing with lawmakers, who are often older and not familiar with the latest digital advances.

Nevertheless, several countries have developed the early stages of regulation. France passed a new law in 2018 called ‘Fight Against Information Manipulation” that compels online platforms to be transparent with respect to sponsored informational content, and furthermore with respect to the sponsor’s name in cases where remuneration is above the threshold of €100. Additionally, the law provides the possibility of an emergency fast-track judicial procedure that fights the conscious dissemination of (false) content to potentially harm the fair process of an election.

However, for the moment at least, several countries count on self-regulation of social media actors. Several of the major platforms have indeed adopted various processes and mechanisms aimed at combating fake news, including the use of fact checkers (human or machine), the identification and flagging of such false information, the closing down of fake accounts or accounts applying bots, or the provision of facts and real information to counter fake ones. Facebook, for example, began flagging, but soon switched to demoting, i.e., the posting of potentially fake news is simply not prominently shown in users’ newsfeeds. The flagging approach was abandoned as reactions from the community were vociferous against placing red warning labels next to potential misinformation. Also, Facebook deliberately decided against a strategy of removing fake news, as doing so could have been viewed as an attack on free speech. Therefore they defaulted to demoting.

**Education**

Education will definitely help citizens be more conscious of how social media can be misused in manipulation. Education also helps individuals to be more aware of data security and the possibilities and challenges triggered by big data. Ultimately, it is the individual who makes an evaluation and decision as to which content s/he believes. We doubtlessly can apply technology to potentially detect fake news and deepfakes, and we can also design more or less
effective regulation. Regardless, ultimately it is the citizen who is the final judge over what is reality or what is fake. Therefore, AI and big data, as well as the ethics involved in them, should be broadly integrated into curricula in order to foster empowered individuals ready for the Information Age.

Yet also those who have passed the age of school and university will need to be trained for our new world. A good example of this can be found in the Netherlands. In February 2019, the Dutch government launched a four-month online campaign right before the European Parliament elections, informing its citizens of the possibility of fake news. On the campaign’s website one could read (Dutch government, 2018):

Disinformation and fake news are a big problem in other countries. There are many different media in the Netherlands. And they show different sides of the news. As a result, fake news and disinformation in the Netherlands do not yet have much influence. But the government wants to prevent that from happening. Everyone in the Netherlands must remain critical and curious about where the news comes from. Check the tips to recognize the difference between fake and real news.

An example of a university having developed a news literacy program is the one at the Stony Brook University School of Journalism, which helps students to distinguish between fact and rumor, advertising and news, and so forth. Also, ESCP Europe Business School is progressively introducing courses on AI and its potential impact into its study programs. As an example of an industry player, Google recently developed a news initiative to enhance and teach digital literacy (Schindler, 2018). That education is definitely necessary is shown in a study by the German Max Planck Institute for Informatics wherein real videos and deepfakes of Barack Obama, Theresa May, and Vladimir Putin were shown to a group of respondents. More than 50% believed the fake videos to be authentic. In the case of Putin, 65% believed them to be real; and only 80% identified the real videos as actually being real (Kim et al., 2018).

Food for Thought:
Instant Democracy and Machines as Democratic Leaders?

In this chapter we surveyed the change in the nature of social media which has been triggered and enabled by artificial intelligence as well as big data. We illustrated how such AI-enabled social media applications can mean a threat to democracy and democratic mechanisms by controlling and supervising citizens, manipulating voters, or simply frustrating them until they drop out of political life. Furthermore, we mentioned what has been shown to help protect
democracy, namely technological possibilities, (potential) regulation, and finally education. Ultimately, artificial intelligence and social media can constitute both an opportunity and a danger to democracy. It all depends on how they are used and by whom.

Artificial intelligence and big data provide an additional, even bigger, possibility, i.e., to potentially create a system wherein (human) political representation of the people is no longer needed. An AI-driven system could constantly collect and gather big data on the current opinions, preferences, and desires of a nation’s people and citizens. Policy and decisions could reflect in-the-moment public interests and potentially represent these more accurately than within a system where political parties are elected for several years, drifting more or less away from public approval over time. The technology to do this already exists. You would not even need to issue precise queries to the citizenry on the relevant topics. A study by Wu, Kosinski, and Stillwell (2015) showed that computer- based judgments of one’s personality are far superior to those of human beings. With only 10 of your likes, Facebook’s algorithm will better predict your opinions than one of your colleagues. With 150 likes, it will do so more accurately than your own family. And with 300 likes, it will do so better than your spouse. This shows the future potential for AI-driven instant democracy.

The question is whether we would trust such a system. Or, might it be better or more trustworthy to be governed by machines than by humans? We already trust machines when we ask apps which political party best represents our interests given their platforms. One could argue that machines, in comparison to human politicians, might make more rational, fair, and evidence-based decisions. They appear to be less biased by personal interests, ideological extremism, narcissism, and the like, than are their human counterparts. This appears to be the opinion of a quarter of the respondents to a 2019 survey on Europeans’ attitudes toward technology undertaken by the Center for the Governance of Change at IE University. Analyzing responses from 2,500 adults from France, Germany, Ireland, Italy, the Netherlands, Spain, and the UK, the study concluded that around 25% would actually prefer decisions to be made by artificial intelligence instead of politicians.

Obviously there are risks involved in such a system, such as manipulation or hacking. Moreover, we could definitely face a transparency issue, as we would not necessarily know or even understand exactly on what such a system would base its decisions, just to mention a few limitations. However, there is certainly room for listing similar risks for any human politician or political system.
References


