INTRODUCTION

Technological progress in the last few decades have made monitoring, tracking and profiling techniques easier, cheaper and more accurate. As a result, surveillance has increased in both the public sector (for law enforcement purposes and public security for example) and in the private sector (for targeted advertising for example). (…) Any form of surveillance is an intrusion on the fundamental rights to the protection of personal data and to the right to privacy. It must be provided for by law and be necessary and proportionate. (EDPR, n.d.)

We do just about everything on and through the internet, including planning our next holidays, buying tickets to a museum, finding the shortest path to a destination, applying for a new job, making a doctor’s appointment, purchasing a book, checking our bank accounts, reading the news and blogs, and of course connecting with others through social media platforms. In this day and age, very few domains of our personal and professional lives are not supported by digital technologies. Digital technologies reveal themselves to be extremely helpful for performing many tasks: we can do most of them from a smartphone anywhere anytime. In other words, digital technologies have contributed to making our lives easier and more comfortable. Haggerty and Ericson (2000) describe the increasing visibility of a large number of people as a “levelling of the hierarchy of surveillance” (p.606).

Our use of digital technologies generates unprecedented amounts of data. Today, data is collected from multiple sources, from web-browsing devices (e.g. laptops, smartphones) to non-browsing devices (e.g. house, car). According to an IDC report, the total amount of data generated across the globe will grow from 33 zettabytes in 2018 to 175 zettabytes by 2025 (Mohit, 2019). Data is also collected by a large array of actors, ranging from social media platforms to supermarket food chains. Data is collected in different countries and at the local, national, and international levels. Data is collected from professional and personal applications. Data is collected day and night, everywhere, all the
This thirst for data is well captured by the sentence: “data is the new oil.” This refers to the value of data in the information age: data is the source of innovation, wealth, and even political power (The Economist, 2017). Artificial intelligence (AI) and machine learning (ML) enable private and public actors to make sense out of all the data collected.

The data collection tactics and tools examined in this book are developed by a relatively small number of actors, who benefit from the captive audiences of large social media platforms. Their limited visibility, and accordingly accountability, highlight the asymmetry of power between on one side citizens (whose data is collected) and on the other side big tech companies and data brokers (those who collect, process, and commercialize citizen data) and their clients (e.g. governments).

Big data and AI present also numerous challenges, including technical or adversarial vulnerabilities (Mitchell, 2019). Vulnerability consists of weaknesses or flaws whether in the hardware, software or data security, which can enable an attacker to compromise its integrity (i.e. trustworthiness of a resource), availability (i.e. appropriate user is denied access to a resource), or confidentiality (somebody gains access to information that she should not have had access to) (see Bowen, Hash, & Wilson, 2006).

Data protection, or the security of personal data, has become a growing concern for European citizens over the last few years. Data leaks, data breaches, and intent to manipulate their behavior has led to a low trust in social media platforms, even though they are used on a regular basis. According to a Eurobarometer survey, eight out of 10 citizens feel they do not have full control over their personal data; six out of 10 say they do not trust online businesses; and more than 90% of them wish for the same data protection rights across all EU Member States (EU Commission, 2018).

The capacity for processing personal data can affect privacy in numerous ways: “Understanding the full picture that without data, a big part of modern AI cannot exist, puts data privacy and democracy at the epicenter of concern” (Manheim & Kaplan, 2019, p.123). To express their concerns about the growing role of digital technology and AI in society, some scholars and experts went as far as to label our era as “digital authoritarianism” (Wright, 2018) and “algocracy” (Danaher, 2014). The lack of accountability and transparency of these tactics and tools, coupled with their capacity to reach out to large audiences, creates an invisible layer of influence between the citizen and his political representatives and authorities.

According to this English philosopher and social theorist, the Panopticon was suitable for addressing potentially violent behavior. This idea motivated the wide deployment in the world of CCTV camera networks and more recently of AI-powered CCTV cameras. Today, it’s the internet, social media platforms, smartphones, Internet of Things, and other AI-powered recognition
AI in public and private forms of surveillance

Technologies that enable surveillance agencies and tech companies to observe citizens 24/7. The importance of social media platforms is well illustrated by the fact that “[t]he Facebook profile has arguably overtaken the CCTV camera as the primary imagery for surveillance studies” (Trottier, 2011, p.66). Surveillance is carried out by social media platforms and other third-party companies, governments, and intelligence agencies, among others. Trottier (2020) considers three kinds of surveillance on social media platforms:

- users watching over one another,
- states and intelligence agencies watching over a target population,
- companies watching over their markets.

This chapter explores the new surveillance paradigm adopted by governments of liberal democracies (and others) and questions whether these practices are compatible with the intention to strengthening relations with citizens “it contributes to building public trust in government, raising the quality of democracy and strengthening civic capacity” (OECD, 2001, p.11). The OECD (2001) recommends indeed that governments enable citizens to express their views on projects and societal issues that affect them in consultations and engage them in decision-making processes. But the surveillance practices adopted by governments “can profoundly affect how individuals think and act, as well as other personal rights (such as freedom of expression or association)” (EDPR, n.d.). This is indeed crucial for the participation of civil society in policy making.

This chapter first focuses on surveillance performed by private actors. The second part discusses surveillance led by states. This chapter also discusses how AI is used by states to listen to what citizens express online with the objective to better understand their needs and views around various policy problems and issues.

SURVEILLANCE AS A BUSINESS MODEL

On social media platforms, only one type of surveillance is visible to users: users watching over one another. It is the counterpart of the interaction among users. The MLAs that enable this form of surveillance, as well as the other two forms, are hidden from view (Trottier, 2011). Indeed, “[c]itizens only see one interface and other users. They cannot see the mechanisms that enable these companies to make a profit. They have limited knowledge about how their personal information is controlled, who controls it, and how it is used” (Andrejevic, 2007, p.27).

These three forms of surveillance contribute to triggering a sense of suspicion and distrust among users and citizens, and even more so because two of them are hidden from users. Covert surveillance grants the watchers power
Artificial intelligence and democracy

over the citizens and entities they observe: “It might sound trite to say that ‘information is power,’ but the power of personal information lies at the heart of surveillance. The power effects of surveillance illustrate three additional dangers of surveillance: blackmail, discrimination, and persuasion” (Richards, 2013, p.1953). Power applies to both overt and covert surveillance. However, when information is collected covertly, it provides a different form of power (Bernal, 2016).

**Surveillance Capitalism**

Surveillance is an integral feature of social media platforms (Trottier, 2016), and can be defined as the “focused, systematic and routine attention to personal details for the purposes of influence, management, protection or directions” (Lyon, 2007a, p.14). Surveillance on social media platforms is intentional. As Zuboff (2015) contends, “‘big data’ is above all the foundational component in a deeply intentional and highly consequential new logic of accumulation that I call surveillance capitalism. This new form of information capitalism aims to predict and modify human behavior as a means to produce revenue and market control” (Zuboff, 2015, p.75). The global data and business analytics marketplace is growing fast: it was estimated to represent USD 171 billion in 2018 and is projected to reach a staggering USD 512 billion in 2026 (Bloomberg, 2020).

Surveillance is embedded in social media platforms through their machine learning algorithms, which reflect their economic interests (Kruse, Norris, & Flinchum, 2018): their algorithms are designed to keep users engaged and online so that the platforms and others can collect more data and use their attention time to display ads. The big data collected is analyzed by MLAs. In other words, surveillance is dedicated to increasing profits. Facebook has in fact succeeded in commodifying the communication between its users (Fuchs, 2012). Their business model is to identify trends and patterns in netizens’ behavior and sell this information to brands and political parties. Since it is not done for public safety but for profit, Zuboff (2019) called it “surveillance capitalism.” As Zuboff (2016) further contends, this data and business analytics industry does not produce any good, but extracts and processes data to sell.

A fairly small number of companies have developed the audience and the capacity to analyze big data, which results “in a situation where a relatively small number of corporations now wield a substantial degree of power over the social and economic behaviours of consumers and populations around the world” (Aho & Duffield, 2020, p.188). For the top five publicly owned tech companies, their value illustrates well how profitable this industry is and is expected to remain in the future. For instance, Apple had a market cap of USD 1.4 trillion as of July 2020, Samsung Electronics with a market cap of USD
AI in public and private forms of surveillance

325.4 billion, Hon Hai Precision (also known as Foxconn, the Taiwan-based producer of Apple products) with a market cap of USD 36.0 billion, Microsoft USD 1.4 trillion, and finally Dell USD 31.2 billion (Investopedia, 2019a). Added together, these five companies have a net worth of USD 3.192 trillion (Investopedia, 2019a), which is higher than France’s nominal GDP (USD 2.78 trillion), the second largest economy of the Eurozone (Investopedia, 2019b).

Mayer-Schönberger and Ramge (2018) contend that power will increasingly be concentrated in the hands of those who have developed the capacity to collect and control valuable data. Tim Wu (2010) predicts the growth of cartels and monopolies. Harari (2018) argues that regulating data ownership is crucial to avoiding power concentration, cartels, and monopolies. Without control over data accumulation, users are deprived of their agency over personal information, which can then become an open door to unfair data management practices, such as discrimination (Cinnamon, 2017; Lyon, 2007b, 2003).

If only a small number of tech companies concentrate most of the data collection, analysis, and monetization capacity (i.e. the Facebook and Google duopoly), and consequently most of the benefits associated with these data, a wide array of smaller third-party private companies benefits either from the data collected (e.g. they purchase this data from the social media platforms) or by collecting data from users’ online behavior. Nonetheless, their data collection and analysis capacity are not at the same level as the largest tech companies often referred to as GAFAM (Google, Amazon, Facebook, Apple, and Microsoft). Indeed, the MLAs of GAFAM benefit not only from their extensive innovation and financial capacities, but also from all the data they have available. Consequently, surveillance tools developed by GAFAM are more precise thanks to their more advanced MLAs.

With the mass of data produced, social media networks began allowing third-parties to access this data and develop new services or apps for their users: “[t]his suggests a kind of meta-surveillance, with Facebook watching over other watchers” (Trottier, 2011, p.66). By allowing third-parties to create what Diakopoulos (2016, p.178) calls “new forms of value,” social media networks have become platforms. For instance, in 2011 Twitter already had over 1 million third-party apps (Twitter, 2011) and Facebook by 2015 reported having over 30 million third-party apps (Novet, 2015). Social media platforms allow a variety of actors to monitor the location of users (in addition to the platform itself), including journalists (Thurman, 2018), corporate security personnel (Lecher & Brandom, 2016), and public safety organizations (Wieczner, 2015): “[t]his means that personal information that has been uploaded for any single purpose will potentially be used for several kinds of surveillance” (Trottier, 2011, p.61).

Surveillance on social media platforms is not only done by the social media platforms themselves and their MLAs. It is also carried out by multiple other
private actors (e.g. digital marketing companies), brands, data brokers, political parties, and many other users, for the most part legally but also sometimes illegally: “[d]ifferent surveillance models are manifest through Facebook. This suggests a complexity of social media surveillance” (Trottier, 2011, p.66). Illegal data collection is well known in the context of economic espionage, but it also occurs on a larger scale, as massive data breaches have shown (e.g. a Yahoo data breach in 2016 affected over 500 million user accounts (Tsukayama, Timberg, & Fung, 2016) and data was exposed from over 412 million accounts in Friend Finder Network in 2016). Social media platforms are not exempt from data breach, including due to poor security levels: over 540 million records from Facebook users were publicly exposed in 2019 (Silverstein, 2019). Data is accessed thanks to cyberattacks, with tools including viruses, trojan horses, and keystroke logging among others (Brisco, 2021).

A part of the surveillance and data analytics industry is visible to the general public, such as telecommunications companies, smartphone and computer brands, internet service providers (ISPs), and social media platforms. However, another part remains in the shadows: “smaller, quieter firms that specialize in gathering people’s personal information from public and private sources, and making it available to other companies for marketing, employment, financial and other purposes” (Lazarus, 2019).

**Data Brokers and Data-Driven Marketing**

Digital marketing, since its early developments in the 1990s, was based on the continuous data collection of citizens’ behavior and the monitoring of their online activities (Montgomery, 2011). Hence, data collection has become one of the main preoccupations and motivations for developing new tech innovations. In the DEMOS Report, Bartlett, Smith, and Acton (2018) estimate that the world produces about 2.5 quintillion bytes of data every day, ranging from the content of a tweet and its metadata, to the type of music one likes or the car one drives. The internet, and the numerous applications it allows, including social media platforms, enabled digital marketers to monitor individual behavior in real time. In other words, digital marketing entered the house not only through radio and TV, but it progressively accessed the pocket of the individual with the generalization of smartphones in many Western democracies, thereby enabling data collectors to track citizens’ online behavior wherever there is a mobile or internet connection (IAB, n.d.-b). Progressively, digital marketers could collect data from every moment of a citizen’s daily life (Smith, 2014). With the growing use of social media platforms by citizens, digital marketing can only expand: In 2019, worldwide digital ad spending was expected to rise by 17.6% to reach USD 333.25 billion, which corresponds to roughly half of the global ad market (Enberg, 2019).
Political marketing stems from commercial marketing: how, where, and what type of data are collected was originally developed for brands to better know their customers, to reach out to them, and to influence their purchase behavior. The same data collected can be used to target a citizen either to influence his purchasing decisions or his political opinion: “[w]hat we are doing is no different from what the advertising industry at large is doing across the commercial space” (Digital, Culture, Media and Sport Committee, 2019). Hence, social media platforms and their MLAs offer the opportunity to other actors (commercial and political) to both communicate with their captive audience and to garner data from this audience. In other words, social media platforms and their MLAs simultaneously offer personal data (i.e. the citizen is the product) and advertising services (i.e. the citizen’s attention and political agency is the target).

Several forms of data are collected, and from several sources. First, social media platforms are one of the most prominent data collectors today (and online platforms in general) due to the captive audiences they have acquired over the last few years and the dual roles they have: on one side an advertising agency for politicians and brands, and on the other a provider of free entertainment and communication services. On social media platforms, citizens deliberately provide a large array of personal information, such as email addresses, first names, last names, phone numbers, and country of residence (Google, n.d.-a).

Since a large majority of European citizens use social media platforms, data has become the new “currency” that citizens use in exchange for free or low-cost services such as WhatsApp and Facebook Messenger (Gibbs, 2016). This exchange of value has enabled the platforms to develop a highly profitable business model based on tracking and monitoring users, and then monetizing both the data collected and their users’ attention. The business model of social media platforms is indeed based on the principle that “the value of the service increases with the number of users” and this technology advance allows them to “organise new forms of participation” and to conduct “business based on collecting, processing, and editing large amounts of data” (EU, 2016).

In addition to the data collected directly by social media platforms, digital marketing companies also collect data from users on the very same social media platforms, in particular related to their conversations. Moreover, data brokers, sometimes called information brokers, syndicated data brokers, or information product companies, are businesses that aggregate

(…) information from a variety of sources; process it to enrich, cleanse or analyze it; and license it to other organizations. Data brokers can also license another company’s data directly, or process another organization’s data to provide them with enhanced results. Data is typically accessed via an application programming
Artificial intelligence and democracy

interface (API), and frequently involves subscription type contracts. Data typically is not “sold” (i.e., its ownership transferred), but rather it is licensed for particular or limited uses. (Gartner Glossary, n.d.)

To illustrate this argument, the US-based data broker Experian claims to have data from over 300 million individuals and 126 million households, to be able to address 85% of the US population, and to link to 500 million email addresses (Experian, n.d.-a).

Traditional data brokers such as Acxiom and Experian acquire data from a large range of sources, “including digital platforms like Google, Facebook and Amazon, telecom service providers (SAP, for example, operates an analytics tool which analyses billions of consumer data points from mobile operator networks), media outlets, publishing houses, retailers and financial services like banks and credit agencies” (Acxiom Corporation, 2017).

Lastly, data can also be derived from leaks, hacks, and data breaches. Once this data is stolen, it is sold, shared, exchanged, made available online. In most cases, it is impossible to track its origin, which means that it can end up being used for political purposes. In 2019, hackers stole more than three billion internet credentials and other forms of personal data, and two-thirds of the victims were not aware of the theft (Lewis, 2019). The press has reported in recent years on data breaches related to electoral processes, for instance in France during the 2016 Presidential campaign, when emails from the Macron campaign team were hacked (Burgess, 2017), disinformation campaigns before, during, and after the EU Parliament elections (Greenberg, 2017) and about voting booth software vulnerabilities (Revell, 2017). However, voter data makes the front page of media outlets less frequently, yet they are also at risk. “In 2017, cybersecurity researchers at UpGuard identified a misconfigured database containing the personal details of 198 million US voters” (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019, p.14). A survey conducted in 2019 showed that cybersecurity experts were concerned that the political campaigning industry is not taking leaks, hacks, data breach risks, and more generally digital interference in electoral processes, seriously enough. And they called for changes in the security practices of individual political consultants that they considered poor (Miller, 2019).

Tracking Citizens Across Devices

Thanks to tracking instruments, political marketers have the capacity to follow citizens across devices and platforms. This allows them not only to target citizens with ads but also to collect data and link the data collected to their identity, which significantly increases the value of the data collected. Users login with Facebook or Google and access various applications through this
single identification process, allowing the social media platforms to follow users even when they navigate on third-party apps, platforms, or services. A 2019 report from the Interactive Advertising Bureau (IAB) and Winterberry Group indicates that digital marketers “prioritize ‘cross-channel’ initiatives above all others in 2019, maintaining a focus on the harmonization of audience experiences across media.”

This large amount of collected data is possible thanks to the capacity offered to advertisers to follow users across devices. Social media platforms have enabled the emergence of new profiling and behavioral tracking tactics and tools to improve advertising and monetize the attention of their users. Behavioral tracking describes the great variety of data derived from online human behavior, including the personal data we voluntarily disclose, the content we produce, and the associated metadata. Large databases are built throughout the world to store all the data that results from this “collect-it-all” or bulk collection strategy. The large datasets amassed by and through social media platforms stem result from all the activities that citizens conduct online: “[t]he breadth of social networks’ reach, the frequency with which they are used, and the quantity and character of information uploaded by users have made them a unique resource” (Thurman, 2018, p.76).

All online activities are monitored to the point where it was possible in 2014 to conclude that “[t]here’s no such thing as privacy on the Internet anymore” (Merkel, 2014). On our personal and professional computers and other browsing devices, discrete files, also known as “cookies,” track our every online movement from their location on our hard drive, and then report that information to remote servers via the internet. In addition to cookies installed on the machine, other tools are developed as part of the code of the website. These tools are designed to collect specific data about our online behavior, in order to improve the service and content available. These tools, including “spotlight ads,” “web beacons,” and “pixel tags,” collect for instance the amount of time users spend on each page, the device they use, what they click on. Some authors contend that even when the user has selected the “Do Not Track (DNT)” setting on their web browser, most websites choose to ignore this (Brodkin, 2015).

Cross-device recognition of users across all the channels and devices they use include cookies and IP address for instance (Levine, 2016). This allows political marketers to reach out to the potential voter when they are the most susceptible to receiving the message, and adapt it to the channel used. For instance, dynamic ads can adapt the message to the user online, tailoring each message to the profile of the citizen and where the ad is viewed (Schuster, 2015): “[g]ranular shopper data allows Criteo Dynamic Retargeting to tailor ads, bids, and product recommendations that drive maximum results.”
Criteo, a leader in cross-device recognition with its HQ in Paris, France, illustrates the cross-device technique with this statement on their website:

Criteo’s global and continuously growing identity graph connects online and offline shopper IDs across devices, browsers, apps, and environments for a more holistic view of each individual. Say Shopper A was looking at couches on their desktop. When they switch to Facebook on their phone, Shopper Graph enables you to recognize them there, and deliver a mobile ad on Facebook for your furniture store. (Criteo, n.d.)

A tracking pixel is a single-pixel transparent image that exists within some websites but that is placed by third-party entities (Bashykarla, Hankey, McIntyre, Rennó, & Wright, 2019). For instance, a Facebook pixel allows Facebook to track users. The information collected from them then allows the political party to optimize its advertising strategy on Facebook (Newberry, 2019). “The parties that had websites with Facebook pixels were: the Bulgarian Socialist Party, the Conservatives (UK), Forum for Democracy (the Netherlands), National Front (France), the Liberal Democrats (UK), the Nationalist Party (Malta), the New Austria and Liberal Forum, New Flemish Alliance (Belgium), Save Romania Union, Sinn Fein (Ireland), and Venstre (Denmark)” (Treffer, 2019).

A cookie can be defined as a small piece of data that allows a website and other entities to recognize users. It is particularly helpful when a user regularly visits a website and could benefit from a personalized view of the items on that site, or where the cart of the e-shop contains the last items selected but not purchased during their last visit. Each time a user visits a website, the browser sends the cookie identification back to the server in order to allow the website to adapt its content according to the user’s last visit (Falahrastegar, 2014). Cookies can be placed by the website itself (first party) or another entity (third-party):

- First-party cookies are placed by the website that a user visits, and their aim is to help websites remember the user’s preferences, including items in their shopping cart, login name, etc. (Federal Trade Commission, n.d.).
- Third-party cookies are placed on the website by another entity, such as an advertising network or analytics company: “For example, if an advertising company notices that you read a lot of articles about running, it may show you ads about running shoes – even on an unrelated site you’re visiting for the first time” (Federal Trade Commission, n.d.).

Mobile web browsers and mobile apps do not allow cookies to function the same way as on non-mobile devices. They offer a heterogenous context where
cookies cannot perform the same actions on all mobile web browsers and mobile apps:

- The majority of mobile web browsers accept first-party cookies. However, they react differently with third-party cookies. For instance, mobile safari on Apple mobile devices does not accept third-party cookies, whereas Chrome on Android mobile devices does (IAB, n.d.-a).
- Mobile apps store cookies within the “webview” of the app, which is used to display online content (i.e. website and ads). However, the apps cannot share cookie information with other apps nor with the mobile web browser of the device, since each app has its own dedicated space on the device called the “sandbox” environment (IAB, n.d.-a).

To overcome the difficulty of tracking mobile users via cookies, four main options are available:

- Client/Device Generated Identifier: is basically an identifier within the operating system of the mobile device, e.g. Apple’s Identifier for Advertisers (IDFA), Google’s Android ID, Universal Device ID (UDID), and MAC Address. Users may or may not be able to control or change a device-generated identifier.
- Statistical, or probabilistic, ID: “is a form of device recognition technology allowing advertisers to identify both mobile and multi-screen audiences in the absence of cookies or other deterministic data” (Shields, 2014) through data provided by the mobile device including device type, operating system, user-agent, fonts, and IP address.
- HTML5 Cookie Tracking: cookies are stored in the HTML5 local storage on the device.
- Universal Login Tracking: “Login With Facebook” is the social media’s universal login API, which allows its users to carry their profile information to other apps and websites, including Spotify, Airbnb, and Tinder (Matsakis, 2018).

However, mobile phones allow other forms of tracking. For instance, they allow beacons “to pinpoint the location of customers in stores and other places and to deliver messages to their mobile devices” (Maycotte, 2015). A beacon can register the presence of other devices nearby and can be used for instance during political rallies to identify attendees, and then combine this information with data from data brokers (Adarsh, 2019). Mobiles also allow geofencing techniques, which track the location of individuals based on Bluetooth, Wi-Fi, and radio frequencies (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019).
Categorization of Data Collected and Process

In the past, political marketers used focus groups, questionnaires, and surveys to assess what certain parts of the population liked, how they were thinking, and how they reacted to specific images and messages. At that time, it was not possible to survey the whole population in real time. Therefore, they could only either pre-test a message with a focus group, or post-test the impact of a message on a representative group of people. But the evaluation of the impact was bound to be incomplete (because one could not assess the whole segment of the population), and the impact itself “wasted” some of the resources of individuals who were not “potential” targets (because one could only target segments based on demographics and broadcast to segments of the population rather than to individuals).

Thanks to social media platforms and the big data collection they allow, political marketers now potentially have access to a much larger part of the population connected to the internet. Most importantly, they can see in real time the effects of their messages, both in terms of content and form. Political marketers can learn a lot more about individuals by collecting personal and metadata from the activities they carry out on different platforms and at different moments. In other words, they do not need to guess what the whole population likes from focus groups or surveys conducted with a representative sample. They can use the big data collected thanks to the wide use of social media platforms by citizens and the digital footprint they leave behind: “If the twentieth century engineers of consent had magnifying glasses and baseball bats, those of the twenty-first century have acquired telescopes, microscopes and scalpels in the shape of algorithms and analytics” (Tufekci, 2014).

The data broker company Experian claimed to hold data on approximately one billion people in Europe and the United States and earned over 4.6 billion USD in revenue in 2018 (Experian, n.d.-b). This data can prove helpful to political marketers to identify and reach out to specific audiences during political campaigns and stimulate parts of the population to go and vote.

Several forms of data are collected, and by a multitude of actors. Data is collected from websites, apps, physical stores, or other situations where customers share these data voluntarily. This category of data is called volunteered data. But data can also be collected from users by third-parties (observed data). Most data used for political campaign purposes are called consumer data. They correspond to

(... the customer information that you’ve collected in the first-party context – for example, information you collected from your websites, apps, physical stores, or other situations where customers shared their information directly with you. There are many types of customer data, some of the common data types are email
addresses, first names, last names, phone numbers, and country of residence. (Google, n.d.-b)

Social media platforms allow different types of data to be collected:

- Demographic data about the user: name, address, email address, gender, age, etc.;
- Data from the user’s social media accounts: contacts and friends in their network;
- Surveys or quizzes that inform about the preferences and interests of the user (e.g. the quiz of Cambridge Analytica “This is Your Digital Life,” that allowed the company to develop its psychographic profiling capacity) (Revell, 2018);
- Behavioral data stemming from the user responses to specific messages or text;
- Metadata (data about the data, such as time, origin, and destination of the message, etc.).

This large amount of data collected is possible thanks to the capacity offered to advertisers to follow users across devices (Acton, 2018). Data can be either provided by citizens themselves (volunteered data) or collected by a third-party (observed data). Furthermore, data can be divided between actual data (data generated by someone or some organization) and modeled/inferred data (new data that is produced from the analysis of the actual data, online activities, and behaviors) (Christl, Kopp, & Riechert 2017). The data collected is about the user (e.g. name, address, email address, gender, age, etc.), his online activities (e.g. content published), and about the data itself (i.e. metadata). Another source of data used for political campaigns is voter files, which are created within political parties to support their political communication, and composed of data from data brokers, surveys, online and offline consumer data.

Metadata is also called communications data. The UK High Court distinguished between three broad categories: subscriber data, service data, and traffic data. Metadata is the information about the communication, or said differently the data about the data (Agnew, 2003) and it includes three categories:

- Descriptive metadata: data for purposes of discovery and identification such as title, abstract, author, and keywords (Foulonneau & Riley, 2014).
- Structural metadata: how objects are put together, for example, a table of contents for a monograph (McDonough, 2018).
- Administrative metadata: information to help manage a resource, including access rights, file type, and other technical information (Baca, 2008) such as geolocation data and data about the device used.
In terms of data collection and analysis capacity, metadata is simpler than other forms of data (e.g. a picture, natural language in a discussion forum). Other forms of data are highly diverse and can be very unstructured. Natural language, or text, can have indirect meaning and innuendo, which is challenging for machines to identify, and is increasingly encrypted. Metadata on the other hand is standardized, mostly numeric, and appropriate to use for quantitative analysis. Moreover, metadata can be perceived as less intrusive than what users publish on social media platforms, or than their personal information, although it can also potentially be revealed as sensitive and intimate information. In other words, metadata can be as revealing as other types of data, but it is easier to process and aggregate (Watt, 2017).

Metadata is collected in addition to other forms of data to better know the audience, understand their behavior, and discern how effective the outreach is. For instance, digital marketers can be interested in finding out the following elements to improve their marketing strategy: number of times an ad was viewed, a newsletter downloaded, a topic discussed, a post recommended, shared, or cited. As Felten (2013) argues, the “(...) analysis of telephony metadata often reveals information that could traditionally only be obtained by examining the content of communications. That is, metadata is often a proxy for content” (Felten, 2013, p.14). What Felten wrote about telephony metadata applies to internet metadata as well, except that the latter, thanks to smartphones and social media platforms, allows the collection of much more data and metadata, ranging from “traditional communications data like email, text messages and phone calls to music listened to on Smartphones, geolocation data, etc.” (Bernal, 2016, p.256). In other words, metadata in the web 2.0 context is even more revealing than previously (Bernal, 2016).

STATE SURVEILLANCE

I think it very important that the mere fact of there being surveillance takes away liberty. The response of those who are worried about surveillance has so far been too much couched, it seems to me, in terms of the violation of the right to privacy. Of course, it’s true that my privacy has been violated if someone is reading my emails without my knowledge. But my point is that my liberty is also being violated, and not merely by the fact that someone is reading my emails but also by the fact that someone has the power to do so should they choose. (...) It’s no use those who have possession of this power promising that they won’t necessarily use it, or will use it only for the common good. What is offensive to liberty is the very existence of such arbitrary power. (Skinner, 2013 cited in Bernal, 2016, p.250)

Big data, Artificial intelligence (AI) and machine learning (ML) are at the center of a stream of surveillance technologies, which are increasingly adopted by governments. State surveillance can have several purposes, including
homeland security against terrorist threats or espionage. In past decades, governments have searched for new ways to collect more data for three main reasons. To analyze and make sense of all this data, and to identify individuals and potential threats on social media platforms (and online in general), states use MLAs, either developed in-house by their intelligence services, or provided by private military firms.

Surveillance includes a first stage where data is gathered, a second stage where data is processed, and a final stage where the processed data is examined by a human intelligence specialist. The question is: when is mass surveillance happening? At which stage?

It is the mere existence of systems and norms that allow the collection of data about citizens that generates the menace of surveillance, not the human examination, nor the use of the data collected (Bernal, 2016). This is the view of the ECtHR when it argues that the existence of data gathering engages Article 8 of the ECHR directly: “Everyone has the right to respect for his private and family life, his home and his correspondence.” This implies that most forms of modern communication, especially social media platforms, can be designated as vehicles for mass surveillance.

However, the US and UK governments argue that surveillance only happens at the last stage; in that case, the data “seen” by humans is not “massive,” or “indiscriminate,” which means that it does not constitute mass surveillance (Watt, 2017). As for the second stage, this is done by algorithms that analyze the data collected. What algorithms do can be described as “the focused, systematic and routine attention” to data, which means it can also qualify as surveillance according to Lyon’s definition, who defined surveillance as the “focused, systematic and routine attention to personal details for the purposes of influence, management, protection or directions” (Lyon, 2007a, p.14).

**Bulk and Systematic Surveillance**

State surveillance can have several purposes, including homeland security against terrorist threats or espionage. In past decades, governments have searched for new ways to collect more data for three main reasons. To analyze and make sense all this data, and to identify individuals and potential threats on social media platforms (and online in general), states use MLAs, either developed in-house by their intelligence services, or provided by private military firms.

First, the emergence of new threats to national security from groups organized in networks has resulted in numerous attacks difficult to predict and respond to with traditional military means (e.g. terrorist attacks in New York, Washington, Madrid, London, Mumbai, Boston, Paris, Brussels, Istanbul, Nice, among others). With such types of attacks, deterrence does not work.
Hence, states have opted for developing their listening capacity and collecting more data to preempt these attacks. In France, the new data protection law allows intelligence agencies to both access and collect the metadata stored by telecom operators and hosting service providers, including location data, in real time, but only for the prevention of terrorism (Winston, 2017). “As the old joke goes, your phone is now equipped with 3-way calling: you, the person you called, and the government. Add in communications providers that sniff your messages, log your metadata, and track your activities, and the scope of the problem becomes clear” (Manheim & Kaplan, 2019, p.123).

Espionage can be defined as “the consciously deceitful collection of information, ordered by a government or organisation hostile to or suspicious of those the information concerns, accomplished by humans authorised by the target to do the collecting” (Demarest, 1996, p.326). It is not a new phenomenon in international relations: espionage was used by states as one method to gather intelligence from afar, including through electronic means (Buchan, 2016), in times of peace as well as conflict. Traditional espionage techniques include analogue phone-tapping, photography, listening devices, and so on. The Five Eyes alliance consists of a global surveillance arrangement of states comprising:

- USA’s National Security Agency,
- UK’s Government Communications Headquarters,
- Canada’s Communications Security Establishment,
- Australian Signals Directorate,

In addition, states use social media platforms for offensive tactics, including targeted surveillance, digital espionage, and disinformation campaigns. These platforms not only allow states to deepen the internal reach of their intelligence agencies, but also to support the deployment of offensive operations and to project their capacity outwards: “[a]lthough varying in resources and capabilities, many governments’ armed forces and intelligence agencies have developed aggressive external operations” (Deibert & Pauly, 2019, p.83).

The second reason for seeking to collect more data is that states wish to improve their tax collection capacity and prevent tax evasion (Rubinstein, Nojeim, & Lee, 2017). For instance, in France, Article 65 of the Customs Code allows customs authorities to issue requests for personal data when investigating tax evasion:

These requisitions may be issued by a customs official having the rank of at least “controller,” and do not need to be approved by a judge. Telecom operators, transport companies, and airlines are among the kinds of companies that can receive...
orders from customs authorities for the communication of data. (Winston, 2017, p.52)

Third, governments take advantage of the increased visibility of citizens on social media platforms. Hence state surveillance programs not only make use of the personal information and metadata gathered by commercial operators, but they also benefit from their profiling and analytical tools (Watt, 2017). Because of the vast datasets generated by users on social media platforms, these large tech companies developed the AI-based data analytics technology to scrutinize and quantify online human behavior, which has made society more visible (Aho & Duffield, 2020) on the one hand, and on the other, has enabled large-scale social engineering (Scott, 1998).

The thirst for data is well captured by the concept of “systematic access,” which can be defined as “direct access by the government to private-sector databases or networks, or government access, whether direct or mediated by the company that maintains the database or network, to large volumes of data” (Rubinstein, Nojeim, & Lee, 2017, p.6). The collection of data is done in bulk: “large-scale collection, retention and subsequent analysis of communications data” (Murray and Fussey, 2019, p.41). In other words, it consists of bulk interception, bulk acquisition of communications data, bulk equipment interference, “bulk personal datasets” (“BPDs”) (Watt, 2017). In France, an intelligence-gathering law that includes provisions for bulk data collection and MLA-enabled analysis was labeled “le Big Brother français” in 2013 and was passed in 2015. The controversial FRA: Lagen in Sweden includes similar provisions (Klamberg, 2010).

As mentioned before, states are interested in gathering information about the communication content produced and distributed in the world, but also about the communication itself: where it takes place, who the sender and receiver are, what the geolocation data are, and the device used to share that content, and so on. Hence, we can say that the focus of surveillance activities is as much the “metadata” or “communications data” as the “content” (Watt, 2017). In Germany, local police forces can ask telecommunications companies to provide personal and metadata from communications to and from a specific individual, in a precise location, and for a specific timeframe. Between 2008 and 2012, Berlin police made 410 “radio cell inquiries” that collected information pertaining to 4.2 million cell phone connections, as Schwartz (2012) reports.

Intelligence agencies do not only target direct threats. Thanks to MLAs, they aim to build “a pattern of life,” which is a very detailed profile of a citizen who presents a potential threat and any other individuals associated with him (MacAskill & Dance, 2013). This means that “[t]he agency is allowed to travel ‘three hops’ from its targets – who could be people who talk to people who
Artificial intelligence and democracy

talk to people who talk to you” (MacAskill & Dance, 2013). On Facebook, the
typical user has about 155 friends (first degree of separation), 25,327 friends
of friends (second degree of separation), and a staggering 4,031,568 friends of
friends of friends (third degree of separation) (Knapton, 2016). In the case of
the Cambridge Analytica scandal, the company stole data from the users of the
apps they developed, but also of their friends, which allowed them to finally
collect – illegally – personal data from over 83 million US citizens (Rosenberg,
Confessore, & Cadwalladr, 2018).

Limited Oversight of State’s Surveillance Practices

State surveillance is quite developed in several EU Member States, particularly
the upstreaming tactics, which consists of bulk collection of data as opposed to
specific disclosures of data from ISPs. This bulk data collection-based
surveillance is still possible in the age of GDPR since “even those nations
with otherwise comprehensive data protection laws, access for regulatory,
law enforcement, and national security purposes is often excluded from such
laws; alternatively, they are treated as accepted purposes for which access
is authorized under separate laws that may or may not provide adequate
safeguards against possible abuses” (Rubinstein, Nojeim, & Lee, 2017, p.6).
In the EU, France and Germany for instance have surveillance programs in
place. The German foreign intelligence agency, BND, has a surveillance hub
in Frankfurt that monitors traffic to, from and throughout the country (Spiegel,
2013). France uses surveillance methods that are similar to the NSA (Follorou
& Johannès, 2013). Since 2015, the government has acquired additional sur-
veillance capacity, including the opportunity to ask service providers to use
their algorithms to spot suspicious activity and share it with the government
(Toor, 2015).

One type of legal requirement from the state, known as data retention, is to
demand that some service providers collect specific categories of data from
users and for a pre-defined timeframe, often ranging from six months and two
years, in order to make that data available to governments upon request. Recent
demands from governments to ISPs have been to keep identifying information
and connection data available (including dialed number information) for
a period of time (Center for Democracy and Technology, 2011). Another
requirement from the state, called design mandate, is where the service pro-
vider must design their information systems so that they can provide the data
to the government in real time or near-real time.

The “third-party” doctrine contends that different laws apply when a gov-
ernment acquires information or data indirectly, not from the data subject, but
from a “third-party.” A famous case to illustrate this argument is about the
jailhouse informant, who receives information from a suspect, and shares it
with the authorities, against the will of the suspect. In addition, the “misplaced trust” doctrine argues that individuals sometimes share protected or confidential information with third-parties because of their misplaced trust in them (Manheim & Kaplan, 2019). These two doctrines lift restrictions in US law, in particular those contained in the Fourth Amendment, to collecting personal data from any non-governmental third-party entity and individual who has them. For instance, GPS-based geolocalization apps such as Google maps have our web search and travel histories, banks hold our financial information, and healthcare insurance companies store our medical history and other highly sensitive data.

The relationship between the state and social media platforms is complex. Snowden revealed in 2013 that the NSA had unrestricted access to the servers of some of the largest tech companies at the time: Microsoft, Yahoo, Google, Facebook, PalTalk, AOL, Skype, YouTube, and Apple (Greenwald & MacAskill, 2013). This relationship is also opaque, since it is often in the interest of both parties to keep the agreement and operations as quiet as possible, including sometimes prohibiting service providers from divulging the information (Rubinstein, Nojeim, & Lee, 2017). Moreover, “[o]versight and reporting mechanisms are either absent or limited in scope when they exist, and generally do not reach voluntary data sharing” (Rubinstein, Nojeim, & Lee, 2017, p.17). And even though most countries impose limits on government access to personal data through courts, committees, and oversight bodies, “[a] major question, of course, is whether those control and review mechanisms are strong enough in the face of technological change, the continuing trend of individuals storing more and more of their digital persona in cloud-based computing models, and more aggressive government demands” (Rubinstein, Nojeim, & Lee, 2017, p.17).

Moreover, the separation between the data collected and used by intelligence and national security agencies on the one hand, and the data from and for law enforcement and other government agencies on the other, is becoming more and more porous. The 9/11 and subsequent terrorist threats have caused this virtual wall to deteriorate. In many countries, data collected for national security has become more easily shared for other uses, and conversely, data collected by law enforcement agencies can be used for national security purposes.

In Germany, the US legal concept of “wall” corresponds to the “Trennungsgebot,” or “Separation Rule” (Schwartz, 2012) which establishes an organizational and informational separation between national security and law enforcement bodies. This rule is also the result of Germany’s history and the considerable skepticism of the population against any form of transgression against privacy rights. This rule nevertheless does not completely prevent these bodies from working together. This argument is well illustrated
by the creation in 2006 of an “Anti-Terrordatei,” or “Anti-Terror Database” (Schwartz, 2012), which by 2011, had acquired personal data about 18,000 individuals from 38 distinct security authorities (Deutscher Bundestag, 2011). Although the database distinguishes between two forms of data (open and concealed), the agency that stored the data in the database always receives a notification of the search request by another agency and can “decide whether the applicable legal rules permit it to share further information with the inquiring agency” (Schwartz, 2017, p.63).

Similar to the case in Germany, in France there is a formal distinction between these two legal frameworks: criminal investigation and national security. However, the distinction between the two is also blurry. Following 9/11 and even more so after the 2015 Paris terrorist attacks, intelligence agencies were given more flexibility to collect data: “As one would expect, fewer safeguards surround data collection in the context of intelligence activities. For example, intelligence authorities do not need a judge’s permission to conduct data gathering, whereas similar data gathering by judicial police would require the authorization of a judge” (Winston, 2017, p.50).

In the 13 countries they surveyed, Rubinstein, Nojeim, and Lee (2017) recognized that “[t]he laws relating to access to communications and communications metadata seem to have grown out of an almost universal recognition of two competing propositions: that communications privacy is an essential right, and that the ability to intercept communications in real time or to access communications and associated data in storage is an important investigative technique for both criminal investigations and the protection of national security interests” (p.106).

The NSA and the BND have a long history of collaboration. The NSA provides “selectors” to the BND, consisting of clues such as IP addresses, telephone numbers, email addresses, or geo-coordinates, about which the BND can search for information and then provide the results back to the NSA. “According to the Zeit magazine, moreover, there are secret agreements in place among the NSA, BND, and the Federal Office for the Protection of the Constitution, under which the NSA provides technologies and goals for data gathering and analysis, and the German intelligence agencies collect the information” (Schwartz, 2017, p.88).

An expert report to the special committee of the German Bundestag investigated the activities resulting from the collaboration between the NSA and the BND, and concluded that it infringed some bilateral agreements between Germany and the United States as well as German law.

The then Federal Data Protection commissioner, Peter Schaar, feared that intelligence agencies might engage in “competence hopping” (Befugnis – Hopping), which means the intelligence agencies from different countries share their intelligence tasks among each other to evade legal restrictions.
Consequently, Determann and Guttenberg (2014) argued that “data stored and transmitted exclusively on European territory is not safer from US cyberspying than it would be in the United States.” To be noted also is that the law of the European Union “does not impose any meaningful limitations on government surveillance because the EU has limited jurisdiction over the foreign intelligence activities of its member states” (Determann & Guttenberg, 2014, p.885).

In France, the legal framework post-9/11 requires telecommunication providers and hosting services to retain “identification data.” And after the 2015 terrorist attacks, France adopted a new surveillance law that (1) created a single and coherent legal framework for intelligence activities; (2) created a Commission of independent experts, the Commission Nationale de Controle des Techniques de Renseignements (CNCTR), to oversee the collection of data from intelligence agencies; and (3) stipulated that the collection of data for intelligence purposes is subject to the authorization of the Prime Minister after reviewing the (non-binding) opinion of the CNCTR. Hence, since 2015, “France’s intelligence agencies have wide-ranging powers to collect data and conduct interceptions with no prior judicial approval. Those rights include the ability to analyze metadata of all French Internet users to detect suspicious patterns of behavior” (Winston, 2017, p.49).

In France, the Supreme Courts, or Conseil d’Etat and Conseil Constitutionnel, grant the government and legislative powers flexibility in drafting and implementing surveillance laws.

The French decisions do not attempt to determine whether the relevant surveillance measure represents the least intrusive means available to achieve the desired objective. When reading French court decisions on government surveillance, one cannot help but think that French courts apply a lighter version of the proportionality test than do the CJEU or the ECtHR. (Winston, 2017, p.50)

At the EU level, the Court of Justice of the European Union (CJEU) (2014) applies a strict proportionality principle to surveillance that infringes privacy for the sake of national security. The CJEU and the European Court of Human Rights (ECtHR) examine whether the activity is accepted by law, and consider the principle of “necessity,” which implies that the activity must be effective and the least intrusive on privacy for a purpose that is respectful of a democratic society.

Two decisions about bulk collection of data confirm the view of the CJEU: in the Digital Rights Ireland case, the Court argued that any collection of data that involves an entire population of citizens is a disproportionate breach of privacy, while in its Schrems decision, it contended that mass surveillance is not compatible with the EU Charter of Fundamental Rights.
A specific feature of the French surveillance system seems however to contradict the decision of the CJEU. The 2015 Surveillance Law allows intelligence agencies to require that telecom operators and hosting providers install black boxes, or Boites noires, on their networks (after obtaining the authorization of the Prime Minister and an opinion from the CNCTR) to analyze metadata and identify suspicious activities such as terrorist threats (Winston, 2017). However, the French Constitutional Court argued that this provision was not a disproportionate infringement on privacy since it was necessarily linked to anti-terrorist activities, was under the authorization of the Prime Minister, and focused only on metadata.

**Sentiment Analysis**

Social media and online platforms are a great source of information and communication between government and citizens. Their use has great impact on policy processes. “First, it impacts the first task of each cycle that is problem definition. In fact, by investigating what citizens are saying on social media platforms, policymakers can discover, when well informed, that there are problems in the society that need to be tackled” (Driss, Mellouli, & Trabelsi, 2019, p.568). For instance, natural language processing and sentiment analysis are useful computational techniques to analyze textual data posted on social medias by citizens. The extraction of opinions formulated on blogs, discussion forums or social networks can make citizens’ voice heard, but also how citizens reason and conceptualize specific political issues. It becomes a crucial asset for policy makers who are more and more aware of the lack of legitimacy and distrust of democratic institutions, yet public policies struggle to be tailored to citizens’ opinions. Consequently “social networks could also play a fundamental role to understand not only opinions, but also arguments supporting them” (Milano, O’Sullivan, & Gavanelli, 2014, p.33).

Social media platforms also enable their advertising clients to merge the consumer data from data brokers with the data collected on social media platforms. This is the case of Axiom, for instance, for the Facebook Marketing Platform. “With the combined power of reach and the wealth of consumer data available to them, internet platforms serve political campaigns on a global scale and provide them with tailored services” (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019). They also allow political marketers to identify new audiences through lookalike modeling (Facebook’s “Lookalike Audiences” and Google’s “Similar Audiences”). This technique allows a political party or campaign manager to identify citizens that “look” like the strongest supporters of their candidate: “Lookalike Audiences are lists of people to target with advertising who are similar to (or ‘look like’) the people currently engaging with your business” (Facebook, n.d.).
Lookalike modeling allows Facebook for instance to identify among its over 2 billion users the ones who are the most similar (according to some psychographic and demographic indicators) to a small number of your strongest supporters. This allows a broadening of the scope of the audience and ensures that all potential voters are targeted with ads: “By finding audiences that the marketer would otherwise be unable to identify, lookalike modeling becomes a key marketing tactic for new customer acquisition” (LiveRamp, n.d.). Moreover, customer database matching can also be useful: they offer their clients the possibility to upload the list of their supporters, identify the ones who use the platform, and then target them with ads. It is the case with Facebook’s “custom audiences” and Google’s “customer match”.

Some digital marketing companies have developed dedicated political marketing services, and in particular sentiment analysis for the nonprofit sector and governments. Bakamo Public, a branch of Bakamo Social, is specifically dedicated to governments and nonprofit organizations. German Foundation Friedrich Ebert Stiftung commissioned the social listening company Bakamo Social to map out how European citizens talk online about migration: the company conducted a digital listening analysis of social media comments between 7 July 2017 and 1 August 2018. Bakamo Social found five dominant narratives associated with migration: security, humanitarianism, demographics, economy, and identity. Its analysis managed to define the percentage of conversations on social media that related to each one of these five narratives. Bakamo Public presents their activities as such:

We created Bakamo Public to make social listening accessible to values-based organisations and governments. For our clients in government agencies, NGOs, international development organisations, and civil society, we deliver insights that answer the hardest questions, rooted in authentic voices from social media. Every day on social media, millions of people discuss in their own terms current events, social issues, politics, and their hopes and fears, all unscripted and brutally honestly. These conversations are imbued with reasoning, context, emotion, and narrative which on analysis can show why people behave, think, and feel as they do, even when they themselves don’t know. (Bakamo Social, n.d.)

Another illustration of the sentiment analysis done by Bakamo Public during political campaigns is their media landscape analysis during the 2017 French Presidential election. They analyzed 20 million social media posts and 8 million shared links in public social media conversations between 1 November 2016 and 22 May 2017. Their analysis showed that “[c]onversation around the elections was disrupted by a counter-narrative that positions traditional media and institutions as elitist, sets the stage for disinformation, and offers solutions contrary to the democratic and pluralistic social order. Within the realm of the
counter-narrative, fake news reinforces biases and may seem true” (Bakamo Social, 2017, p.2).

Sentiment analysis combines human and artificial intelligence. It is based on the automation of the monitoring and analysis processes. First, scrapers automate the extraction of data from social media platforms: this allows the gathering/collection of content such as posts, tweets, and anything related to a pre-set hashtag or from specific users, including social media interactions like retweets, commenting, sharing. The second step in automation is to identify the sentiment in the data collected using a form of AI:

Much of this analysis builds on recent advances in natural language processing (NLP), a kind of artificial intelligence that specialises in looking at large bodies of text. NLP is programmed not only to recognise positive and negative sentiments of certain words, or the linguistic context for the sentiment of a message, but also to develop new rules as it performs more and more analysis, making it “smarter” over time. (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019)

The number of digital marketing companies that use social media platforms to collect data about users is quite large. However, since this market is global, it is difficult to have an overview of all their offers, which contributes to the blurry and diverse characteristics of computational politics. Among these companies is the polling and survey company YouGov, which acquired Portent.IO to benefit from their social listening capacities. YouGov Signal, as it is now called “track[s] engagement metrics across every piece of data across all major social platforms on a real time basis” (YouGov, n.d.). YouGov is present in Europe, including in France, Germany, Spain, and Sweden.

During political campaigns, the data collected and analyzed help the political analysts and political marketers to evaluate the “brand” of the political candidate, and show them how the population on social media platforms perceives specific societal questions. This can inform the campaign team, enabling them to identify and then address a concern of the population that had not yet been detected, to adapt in real time a speech and more generally the communication about an issue according to the sentiment analysis, and to learn which words are used and how the issues are framed by citizens and influencers online.

Brandwatch (n.d.), another digital marketing company, offers social media listening and analytics services thanks to its data library of 1.4 trillion conversations. By acquiring the London-based SaaS start-up Qriously, this major social intelligence company can also launch targeted surveys, with global reach and near-immediate results. Qriously turned mobile ad-networks into a distribution platform for polling and quizzes.

OssaLabs is another sentiment analysis company that was developed under government research contracts and delivers advanced analysis capabilities (OssaLabs, n.d.-a). OssaLabs enables analysts to create queries about content,
that is, keywords (and combinations of keywords) that consist of individual words, phrases, hashtags, user handles, or URLs. Once the queries – and consequently the data collection – is in operation, OssaLabs also enables analysts to organize the data through three sets of groups: mention groups (to gather all data associated with one keyword and related terms), participant groups (to group together data from a specific user or account), and follower groups (to bring together data generated by followers of a user, such as a political user). Finally, dashboards enable analysts to review, filter, and visualize all the data and analysis outcomes, with functionalities such as Top Authors, Top Tweets, Word Cloud, Overall Sentiment, and Geospatial Analysis. OssaLabs claims that they can help political campaign research with real-time information about voter perception, and rally constituents and potential voters around top-of-mind issues (OssaLabs, n.d.-b).

However, sentiment analysis focuses only on users who interact and discuss politics on social media platforms. In other words, it leaves behind the conversations that take place in other contexts, such as at home or during the lunch break at work. Moreover, the claim of sentiment analysis companies that they can deduce, from the data collected online, how users will behave in the future (e.g. vote) is somehow misleading and not a reliable method (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019). Predicting the future behavior of individuals is complex and cannot be based only on past data as mentioned previously. It is nevertheless a significant advantage to “have the pulse” of the population on social media about various issues and public figures (for political marketing purposes) and for intelligence purposes.

CONCLUDING REMARKS

Data is helpful for online platforms to better understand users and commercialize their attention. It is also the raw material that feeds AI, that allows machine learning applications to refine their processing capacity and reach more precise results. Hence, thanks to the commodification of online behavior and attention, social media platforms not only increase their profits, but also develop new AI-based services and improve the existing ones. Hence, their capacity to collect and process data provides them with a double win: on one hand they make money as data brokers, and on the other they develop the future of technology with the data collected. Due to the unique combination of technological innovation and financial capacity, as well as the huge amount of data collected, social media platforms are in the best position to develop future technology and design it according to their interests and the interests of the shareholders. They have billions of users ready to give away their data and test new products and services in exchange for free services.
What is more, big data and AI-based surveillance tactics and tools provide states with an unprecedented access to the personal data of citizens. The limited oversight of these surveillance practices poses a substantial risk to privacy and democracy. Moreover, surveillance is not a sure way to increase the security of a nation. The question is whether it is indeed increasing security, which some analysts and experts question:

…over the last fifteen years, the bulk collection approach has cost lives, including lives in Britain, because it inundates analysts with too much data. It is 99 per cent useless, as attacks occur when intelligence and law enforcement lose focus on previously suspected terrorists and fail to find accomplices or others enabling fresh attacks. (Former NSA Technical Director William Binney cited in Bernal, 2016, p.251)

The reason is that the techniques used to detect commercial fraud cannot be applied with the same level of success to terrorist plots since terrorist attacks are rare. Consequently, the prediction systems have not learned from a large enough array of data and cases to be able to make precise predictions, which can increase the risk by flooding the system with false alarms.

Furthermore, Edward Snowden showed that public authorities, including security services, are not immune to function creep. In other words, the existence of data about citizens that is collected and stored on highly secure servers does not preclude this data from ever being misused, hacked, or corrupted. Consequently, this data represents a vulnerability for the surveillance system itself – and the state – as well as for the citizen. On the one hand, this data is a gold mine for foreign intelligence services and criminal groups to detect vulnerabilities and gather data to launch a cyberattack or disinformation campaign, or to blackmail political leaders. On the other hand, it is also a gold mine for cybercriminals to deceive and blackmail individuals. In this context, the new surveillance paradigm increases the degree of uncertainty and vulnerability in the citizen–government relations.

REFERENCES


Fuchs, C. (2012). The political economy of privacy on Facebook. Television & New Media, 13(2), 139‒159.


Artificial intelligence and democracy


