Big data analytics in e-government and e-democracy applications: privacy threats, implications and mitigation

Paola Mavriki* and Maria Karyda

Department of Information and Communication Systems Engineering,
University of the Aegean,
GR-83200, Karlovassi, Samos, Greece
Email: pmavriki@aegean.gr
Email: mka@aegean.gr
*Corresponding author

Abstract: Big data analytics can help governments and organisations to support democratic processes through improving efficiency, effectiveness and transparency. A growing body of research investigates privacy threats related to big data analytics, but their implications for democracy are still scarcely explored. Focusing on the democratic value of privacy, we identify privacy threats for citizens stemming from the use of big data in e-government and e-democracy applications. We analyse the challenges for e-government arguing among others that automatic decision-making may lead to discrimination compromising equality, a basic democratic value. We also explore the challenges of the privacy threats for e-democracy, arguing that decreased privacy facilitates manipulation, polarisation and disinformation. Finally, we critically examine relevant technical privacy enhancing solutions which may play a significant role in shielding democracy through allowing citizens to freely share, access and discuss information and content that is contrary to political, religious or social views of governments.

Keywords: privacy; democracy; e-democracy; big data analytics; privacy implications; privacy-enhancing technologies.

Reference to this paper should be made as follows: Mavriki, P. and Karyda, M. (2022) 'Big data analytics in e-government and e-democracy applications: privacy threats, implications and mitigation', *Int. J. Electronic Governance*, Vol. 14, Nos. 1/2, pp.58–82.

Biographical notes: Paola Mavriki is a PhD candidate at the Department of Information and Communications Systems Engineering, University of the Aegean. Her research interests include privacy and data protection; e-democracy, e-government, human behaviour in information security, information security awareness.

Maria Karyda is an Associate Professor with the University of the Aegean, Department of Information and Communication Systems Engineering. Her research interests include organisational aspects of information systems security management, privacy protection in digital social networks and security culture and awareness.

This paper is a revised and expanded version of a paper entitled 'Big data analytics: From threatening privacy to challenging democracy' presented at the *8th International Conference eDemocracy 2019, Safeguarding Democracy and Human Rights in the Digital Age*, Athens, Greece, 12–13 December.

1 Introduction

Digital information is being produced today at an unprecedented rate. People publish their feelings on Facebook accounts, tweet their opinions, call friends on cell phones, post photographs on Instagram and log locations with GPS on phones providing large amounts of data to big data organisations (Grimmer, 2015; Mai, 2016). Moreover, with increasing numbers of sensor-enhanced everyday objects and infrastructures, such as smart home controls, activity tracking applications and context-sensitive mobile devices, ubicomp systems are taking a central place in our environments (Väänänen-Vainio-Mattila et al., 2015) generating very large volumes of data. Big data analytics refers to tools and techniques that analyse and acquire intelligence from big data, and the process of researching massive amounts of complex data in order to reveal hidden patterns or identify secret correlations (Gandomi and Haider, 2015; Jain et al., 2016). Although the business sector is leading big-data-application development, the public sector has also started to derive insight to help support decision making in real time from fast-growing in-motion data from multiple sources, including the Web, biological and industrial sensors, video, email, and social communications (Kim et al., 2014). Big data can facilitate governments to achieve their goals by improving efficiency, effectiveness, transparency and accountability (Klievink et al., 2017), and support democracy (Lindner and Aichholzer, 2020). Security and fight against crime, healthcare system support, new forms of e-participation, usage of new technologies to enhance the quality and number of services provided by the public administration, are some of the many fields of application in public administration of big data analysis (Höchtl et al., 2016).

However, several years since the broad adoption of digital technologies, research shows that many e-government initiatives fail to deliver the promised benefits and be used by a large portion of citizens (Bindu et al., 2019), while the challenges posed by these technologies are gradually recognised by individuals, civil rights groups, governments and society (Körner, 2019). The revelations concerning leaked documents disclosing the big scale goals of data collection, analysis and use by the NSA and possibly other national security organisations, was one of the first incidents which brought to public attention the fragile balance between privacy risks and big data opportunities (Polonetsky and Tene, 2013). After the Snowden disclosures, several highly publicised events have drawn attention to the use of people's personal data by other actors and agencies, both legally and illicitly (Lupton and Michael, 2017). A well-known case fuelling even more the debate over technology's societal impact and risks to citizens' privacy and well-being was the unauthorised access to personally identifiable information of more than 87 million unsuspecting Facebook users to the data firm Cambridge Analytica (Isaak and Hanna, 2018). The 2017 'Democracy Index' published by the Economist Intelligence Unit found that over half of the countries surveyed

experienced a decline in their democracy 'scores'. In Manheim and Kaplan (2019) the authors claim that the increasing use of artificial intelligence is one of the factors to blame.

Scholar research has identified the impact big data technologies on democracy (Bradshaw and Howard, 2018; Leese, 2014; Magrani, 2020; Tene and Polonetsky, 2017; Zarsky, 2015), but the role of privacy is less investigated (Boehme-Neßler, 2016; Manheim and Kaplan, 2019). The ambiguous nature of the concepts involved in combination with the characteristics of big data, make privacy challenges more complex and difficult to identify. Though there are extended legislative initiatives in most western democracies to address privacy issues related to big data analytics (e.g., related decisions of the European Court of Human Rights, of the European Court of Justice, the EU General Data Protection Regulation etc.), recent research (Bindu et al., 2019) suggests that a greater understanding of privacy in complex such as e-government for example contexts, could help to increase the success of e-government initiatives and that there is a need for further privacy research (Liu et al., 2020; Munyoka and Maharaj, 2019).

In this direction, the aim of this paper is to identify privacy threats and implications, stemming from the use of big data technologies for e-government and e- democracy. We analyse privacy issues and implications posed by surveillance and monitoring technologies, crowdsourcing and political communication applications both for e-government and for e-democracy. Our analysis identifies and takes into consideration all stakeholders involved, including governments, elected officials, media (and the role of online portals), political parties and interest groups, civil society organisations, international governmental organisations and citizens/voters.

Our analysis is based on an extensive literature review related to the challenges of big data applications for e-government and e-democracy. Focusing on the societal and democratic value of privacy and drawing on Dahlgren's (Dahlgren, 2013) conceptualisation of the democratic public sphere, we analyse the implications of these threats for democracy by identifying the privacy issues involved due to the use of big data technologies. We argue among others that big data technologies, such as personalisation for example, when employed for political communication have the potential to challenge democracy through compromising basic democratic values such as fairness, accuracy, completeness, pluralism of views etc. Also, we explore challenges for e-government arguing for instance that surveillance may lead to social sorting and discrimination.

Finally, we analyse relevant techniques and methods for data protection in order to identify those appropriate for tackling the identified privacy threats and to help people to freely share, access and discuss information and content that is contrary to the political, religious or social views of governments protecting this way basic democratic values.

The remaining paper is organised as follows: the next Section analyses the relevant privacy literature with regard to e-government and e-democracy, while Section 3, focusing on the social and democratic value of privacy, identifies privacy threats associated to big data surveillance for e-government and privacy threats stemming from crowdsourcing and political communication big data applications for e-democracy. Distinguishing the challenges for e-government from the challenges for e-democracy, Section 4 analyses the political and social implications of the privacy threats on democracy. Section 5 examines solution to mitigate the implications on democracy

through enhancing data protection and the paper concludes in Section 6 with a discussion on open issues and further research.

2 Background: challenges of big data applications for e-government and e-democracy

2.1 E-government and e-democracy

The term e-government emerged in the late 1990s as a context within which to share experiences among practitioners. Since then it has evolved in a research field involving multiple disciplines and employing a variety of theories and methods (Dwivedi et al., 2012; Grönlund and Horan, 2005; Sundberg, 2019). Though there is no universally accepted definition, it is generally considered that e-government involves using information technology to improve the delivery of government services to citizens, businesses, and other government agencies (Palvia and Sharma, 2007). E-government entails transferring government activities online, aiming to increase delivery of information and services to citizens, business, and public administration (Ingrams et al., 2020; Netchaeva, 2002). According to Joseph and Johnson (2013) big data can increase e-government efficiency and effectiveness, helping it to evolve into t-government (transformational government) which is viewed as the ultimate evolutionary stage of e-government.

Similarly, no single or universally accepted definition of the term e-democracy exists in literature, whereas with the same meaning, terms such as e-participation, digital democracy or cyber democracy are also used. Norris (2010) defines e-democracy "as the use of information and communications technologies to provide citizens access to governmental institutions and officials (elected and appointed) and to facilitate or enable citizen participation in governmental activities, processes and decision-making, including remote (e.g., via the internet) voting". For Lindner and Aichholzer (2020) e-democracy is "the practice of democracy with the support of digital media in political communication and participation". E-democracy is usually based on the models of participatory and deliberative democracy.

Grönlund and Horan (2005) argue that a democratic society includes three distinct but interrelated spheres: the political sphere, the administrative sphere, and civil society. In this context, e-democracy takes place between civil society and formal politics (Sundberg, 2019) and, according to Palvia and Sharma (2007) refers to online activities of governments, elected representatives, political parties and citizens for democratic processes. Types and tools of e-participation include e-petitions, e-initiatives and e-campaigning whereas decision-making is supported by applications such as e-consultations, e-participatory budgeting, e-voting crowdsourcing for law proposals, for policymaking, collaborative decision-making within political parties etc. Summarising, as governments often adopt technological solutions for reasons of efficiency and cost savings, they are not necessarily enhancing democratic processes. Hence, as Netchaeva (2002) argues, it is important to make a distinction between the notions of e-government and e-democracy. However, if Norris's (2010) definition is considered, e-government practices potentially may contribute to the government-citizen dialogue and support e-democracy.

2.2 Challenges of big data applications that support e-government and e-democracy

Governments are often using big data for surveillance purposes (Mitrou et al., 2016). In modern societies, surveillance is increasingly emerging as a key governing technique of state authorities, corporations and individuals: "the focused, systematic and routine attention to personal details for purposes of influence, management, protection or direction" (Lyon, 2014, p.14 in Friedewald et al., 2017).

Sophisticated big data technologies are continuously emerging, making the privacy threats and their implications more and more complex and difficult to identify. Combining facial recognition technology and big data, sophisticated artificial intelligence applications may theoretically identify today anyone, anytime, anywhere, and what they have done (Lin and Hou, 2020; Shamsi and Khojaye, 2018). Moreover, there is an increasing interest in recognising human mobility behaviours. The accurate real-time and offline group detection is beneficial for various domains such as crowd management or for analysing social engagement or isolation (Solmaz et al., 2020). However, the high amount of tracking of individual mobility has raised serious concerns about privacy in various contexts (Kondor et al., 2020).

Many governments in western democracies are employing digital consultation platforms, e-petition mechanisms, and other types of crowdsourcing platforms to support e-democracy initiatives (Perez et al., 2018) as exploiting their positive impact on e-participation (citizens taking part in brainstorming, discussing, developing, and implementing decisions) have been explored in literature (Aitamurto, 2012; Liu et al., 2020; Lee et al., 2014; Taeihagh, 2017). The associated privacy challenges however are scarcely investigated. One of the few related studies (Diamantopoulou et al., 2018) identifies challenges "that crowdsourcing, and in general, e-participation approaches impose with regard to privacy protection".

In recent years we have seen a growing interest in the role of data in election campaigns to support to political communication (Gibson, 2015; Grassegger and Krogerus, 2017; Kreiss and Jasinski, 2016; Tufekci, 2014; Zuiderveen Borgesius et al., 2018). Relevant research suggests that data are now shaping the way campaigners communicate with voters facilitating the efficient allocation of scarce resources (Anstead, 2017). Several examples of integration of electoral campaigns in the digital environment of Europe are coming from the UK (Anstead, 2017; Gibson, 2015). Authors (Bolsover and Howard, 2017; Dalton, 2016; Jenkins et al., 2016; Tufekci, 2014) argue that the use of big data analytics for political purposes poses privacy issues with impact on the fundamental characteristics of democracy such as fair elections.

Summarising, researchers agree that safeguarding people's personal information is critically important for the welfare of modern societies, but research evidence indicates that big data technologies challenge our current understanding of privacy and the regulatory approach to privacy protection. In general, most authors (e.g., Aral and Eckles, 2019; Bradshaw and Howard, 2018; Magrani, 2020; Tenove, 2020) focusing on the negative impact which big data technologies may have on democratic values, discuss threats as discrimination, manipulation, disinformation without further exploring the role of privacy. In addition, e-government and e-democracy are multidisciplinary research fields, using a variety of theories and methods. Sundberg (2019) argues that the use of personal data collected without permission raises several significant ethical issues

but the related privacy issues and implications are scarcely addressed. Bindu et al. (2019), Liu et al. (2020) and Munyoka and Maharaj (2019) further suggest that there is a need for greater understanding of privacy in complex settings such as e-government contexts.

In the following section, after discussing the democratic and societal value of privacy, we identify privacy threats stemming from big data applications for e-government and e-democracy.

3 Privacy threats associated to big data applications for e-government and e-democracy

3.1 The democratic value of privacy

Westin's (1967) conception of privacy addresses how people protect themselves by temporarily limiting access of others to themselves. For Westin (1967, p.7 in Margulis, 2003) "privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. [Moreover]... privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means...". Based on its political philosophy, every society sets a distinctive balance between the private sphere and the public order.

Westin (1967) argues that in political democracies "privacy provides opportunities for political expression and criticism, political choice, and freedom from unreasonable police interference; it provides opportunities for people and organisations to prepare and discuss matters in private; it allows non-political participation in family, religion, and in other forms of association". Briefly, privacy is an 'arena of democratic politics' (Westin, 2003). Adopting Calo's (2011) arguments we consider that privacy in this case aims to secure a zone of freedom and action for citizens. Privacy as a zone of freedom is justified because it is intrinsic to human dignity, or instrumental for realising self-development. The agents whom privacy protects might be a single individual, a distinctive social class (e.g., young Greeks or citizens of Athens) or a distinctive social role (e.g., voters).

3.2 Privacy threats related to big data surveillance applications in e-government

As Solove (2003) mentions, although the governments of contemporary democracies are far from the meaning of Big Brother as described in Orwell's novel, some authors argue that they may have similar capabilities. Big data intensifies certain surveillance trends associated with information technology with applications in various areas such as healthcare, human rights, control (e.g., security, antiterrorism) (Lyon, 2014; Taylor et al., 2016).

An important privacy threat related to surveillance refers to the '*chilling effect*': when observed, people's behaviour might be inhibited, making them less likely to attend political rallies or criticise popular views (Solove, 2005). Global surveillance may deter public intellectuals and other citizens from voicing their opinions in the public sphere (Eide, 2019).

Governments may conduct surveillance by analysing and exchanging large amounts of information on their citizens, for example by using data mining tools to identify individuals or groups 'of interest' (Brown, 2014) and even when the government does not aim for total social control, surveillance can still impair freedom and democracy (Solove, 2003). Suspects can be isolated by category while even if the suspects have their names cleared by judicial process, it can be difficult for persons with a 'record' to make a new start. Data in the Canadian Police Information Centre, for instance, remain there permanently. And when police include mental health problems in their records these can lead to denial of entry to Canadians trying to cross the border into the US. Attempted suicide calls, for example, have been uploaded to international databases with just this outcome (Lyon, 2014).

Even more, identifying individuals is not necessary for group profiling to occur and peoples' privacy may be violated without their identity being disclosed. Data analytics may reveal the characteristics of anonymous groups of people, either by inference based on the characteristics of a surveyed group within the larger dataset or by observed network structure. People are likely to interact with others who are similar to them, hence from people's communication networks their contacts 'ethnicity, gender, income, political views and more' may be identified (Taylor, 2017).

Also, groups can be identified and classified by recommendation systems from analysing and processing group photos using metadata embedded in images (the time stamp and GPS coordinates of the place where the photo was taken) (Chen et al., 2012). Surveillance's inhibitory effects are especially potent when people are engaging in political protest or dissent. People can face persecution, public sanction, and blacklisting for their unpopular political beliefs. Finally, surveillance can make associating with disfavoured groups and causes more difficult. For the observers, surveillance presents a profound array of powers that are susceptible to abuse (Fisher, 2004; Solove, 2003).

3.3 Privacy threats related to big data applications for e-democracy

3.3.1 Crowdsourcing

Privacy threats related to crowdsourcing vary depending on the used methods and the context in which the citizens' – sourcing methods are used. Government agencies, crisis response teams, NGOs etc. may collect data from open-source digital platforms aiming to develop new policies, help victims of natural calamities to find shelters, medicines and other emergency needs, etc. The collected vast amount of data may include detailed information of who the users are, their mobile numbers, IP address, geographical location etc. These information may be used to target based on the hypothetic health status, age, gender, race, religion, political ideology, sexual orientation, etc. (Halder, 2014). Some groups might be provided with different opportunities, or prices or levels of service than others.

Moreover, especially in political crowdsourcing context governments can avail GPS/GPRS- based data provided by citizens and misuse those to oppress those individuals or groups who are against governments (Halder, 2014). Thus, as Halder (2014) specifies, the contributors of crowdsourcing initiatives become potential victims of human rights violations by the secret agents of governments and sometimes even by oppositions or terrorist organisations. As most crowdsourcing platforms offer limited privacy settings to participants, these threats are more likely to occur.

Other means for collecting information, knowledge, ideas and opinions from the citizens are through various social media platforms. The 'citizen-sourcing' may be effectuated by processing content which has already been generated freely by citizens on various social networks such as Facebook and Twitter (Charalabidis et al., 2014; Diamantopoulou et al., 2018; Ghermandi and Sinclair, 2019). Although 'passive' crowdsourcing relies on public data, users are unaware of the fact that their information is collected and analysed. In addition, they also ignore the purposes for which their data is used. Consequently, there are questions whether the implicit consent to publish the data is sufficient or rather an expressed consent is needed. Even in the absence of individual privacy violations, use of users' data without consent may constitute a violation of the rights of research subjects (Ghermandi and Sinclair, 2019). Also, decision-making becomes less transparent when based on combining a wide range of data sources and thereby less accountable (Cuquet et al., 2017).

Moreover, mobile crowdsourcing is becoming an increasingly popular way to collect geo-located data from millions of contributors. Beside the privacy threats related to the collected sensed data (e.g., heartbeat rates, fingerprints) there is environmental information sensed by users which can be used for inferring even more information about them (Feng et al., 2018). In particular, the privacy of location has received great attention as personal information may be inferred and disclosed through de-anonymisation from the users' geolocations, such as their points of interests (e.g., home, workplace, favourite grocery stores) or their habits and preferences. These privacy risks are accentuated by the massive amount of geolocation traces that are being collected by mobiles services (Mineraud et al., 2015).

3.3.2 Privacy threats related to political communication

Big data technologies allow practitioners, researchers, policy analysts, and others to predict the spread of opinions, political trends, etc. (Ekbia et al., 2015). Campaigns need accurate predictions about the preferences of voters, their expected behaviours, and their responses to campaign outreach. In addition, campaigns may use data to construct predictive models to make targeting communication more efficient (Nickerson and Rogers, 2014). Based on existing publicly available information, big data analysis tools can generate a predictive model of what has a high probability of being PII, essentially imagining an individual's data (Citron and Pasquale, 2014). Kosinski et al. (2013) method for instance is able to predict political party affiliation about 85% using only Facebook "likes". Personal harms emerge from the inappropriate inclusion and predictive analysis of an individual's personal data without their knowledge or consent.

Solove (2013) refers to deducing extensive information from 'clues' as the 'aggregation effect' and claims that the kinds of predictions that can be made from this type of data are "far too vast and complex, and are evolving too quickly, for people to fully assess the risks…involved". Once released to the public, data cannot be taken back (Narayanan et al., 2016). A person may leave behind in several years thousands of pieces of data which does not affect her or him negatively. As time passes, data analytic techniques improve, and a newly created algorithm may process the previously harmless digital footprints deducing potentially harmful information (Narayanan et al., 2016; Solove, 2013).

In addition, according to Moerel and van der Wolk (2017) predictions can be used to intentionally influence future options of people. Use of predictions poses risks of limiting individuals' human rights and freedoms, insofar as people self-regulate, being aware of "a state of conscious and permanent visibility" (McDermott, 2017).

Another privacy threat, referred to as the incremental effect, is related to automated profiling and the accumulation of personal data. Fragments of data regarding an individual user may be linked piece by piece until an individual profile is entirely exposed (Tene and Polonetsky, 2012). Aggregated data may reveal facets of people's lives, but the data is often reductive and disconnected from the original context in which it was gathered, and this leads to distortion (Solove, 2005). Privacy threats may arise for which individual control offers no protection, for instance, in the case of the based on group classifications decisions (Hildebrandt, 2012; Mulligan et al., 2016; Winter, 2015), or in the case of inferring, through big data techniques, sensitive personal data from digital footprints or from data that individuals have intentionally disclosed (Kosinski et al., 2013; Tene and Polonetsky, 2012). When decisions are taken based on data mining and profiling, undesirable discrimination may occur, as big data analytics exposes sensitive behaviours or other personal information that could be used to disadvantage certain individuals or groups by corporations or governments. Citizens may experience political and economic discrimination (Winter, 2015). For example, a person may be profiled inaccurately as an extremist movement or party adherent or supporter. Also, a party could target particular voters with tailored information that maximises, or minimises, voter engagement (Zuiderveen Borgesius et al., 2018).

There are also concerns about the privacy of specific groups such as non-governmental organisations (NGOs) or journalists. The activists are not just individual targets, but, because of their work, they may be targeted as a group. "Their claims to group privacy are rooted in human rights law, which recognises the need for special safeguards for particular groups such as journalists" (Eijkman, 2017).

According to Raymond (2017) demographically identifiable information (DII) comprises all forms of data in which the identification, classification, and tracking of demographic groups; this includes personal identifiable information (PII), online data, geographic and geospatial data, environmental data, survey data, census data. In the case of the release of DII, group privacy risks may occur. As Taylor (2017) argues, the risks posed to groups by big data analysis are "particularly clear on the political level: if unwelcome movements can be predicted, authorities can step in before people become defined as refugees, asylum-seekers or other problematic categories that award the right to move".

Summarising, this section has identified privacy threats associated to surveillance big data technologies destinated to serve the common good in the e-government context and privacy threats associated to big data technologies for crowdsourcing and political communication purposed to support democracy in the e-democracy context. We show that big data technologies such as surveillance, profiling for personalisation and for targeting, data mining, automated-decision-making etc. pose a multitude of privacy threats, including disclosure of sensitive personal data. The impact of these threats is related to individual, society as well as for democracy, as will be discussed in the following section.

4 Political and social implications: impact on e -government and e-democracy

As previously analysed, privacy facilitates, among others, political expression and criticism, political choice, and freedom from unreasonable police interference. Privacy also allows citizens to form their political, democratic identities. In addition, individuals achieve human dignity and significance in sociocultural contexts and groups intermediate between the individual and the state. Consequently, individual and group privacy have a critical role in a democracy.

We consider the privacy stakeholders belonging to the context examined here, those which Clift (2003) names 'democratic actors': governments, elected officials, media (and major online portals), political parties and interest groups, civil society organisations, international governmental organisations and citizens/voters. On one hand, as Clift (2003) specifies, the only ones who are supposed to experience 'e-democracy' are the citizens. On the other hand, however, advanced and sophisticated technologies make possible the political exploitation of the citizens' personal data challenging basic dimensions of democracy as we discuss next.

4.1 Implications for e-government

Over the last century Western democracies have achieved a progressive political accomplishment of equality of outcomes and opportunities (e.g., equality before the law; equal employment opportunity; anti-discrimination legislation etc.). As Henman (2005) describes, targeted approaches to public policy are challenging these policies and principles that seek to achieve equal opportunity because targeting involves an inequality of access and choice. Targeting explicitly creates different opportunities and chances for different groups. For example, Computer Assisted Passenger Pre-Screening classifies passengers according to their risk of terrorist activity and allocates unequal levels of surveillance of passengers while it also may involve a greater imposition of surveillance and scrutiny of targeted groups.

In terms of the dimension of interaction of the Dahlgreen's public sphere, publics, according to Habermas and Dewey (in Dahlgren, 2013), exist as discursive interactional processes which is the basic premise of those versions of democratic theory that see deliberation as fundamental. With the advent of the internet, "civic interaction takes a major historical step by going online, and the sprawling character of the public sphere becomes all the more accentuated" (Dahlgren, 2013). Social media users spend a great deal of time curating online 'exhibitions' of different aspects of their identities. However social media surveillance reduces individuals' control over the information they disclose about their attributes in different social contexts, often to powerful actors such as the state or multinational corporations. This limits their ability to regulate their social interactions and identities (Brown, 2014). One of the greatest harms occurring from mass surveillance – particularly mass covert surveillance such as communications monitoring – as Brown (2014) and Goold (2010) claim is the potential chilling effect on political discourse, on the ability of both individuals and groups to express their views and on the possibilities for whistle-blowing and democratic activism.

In addition, new surveillance technologies can lead to '*social sorting*', where discrimination and privilege are entrenched through the unplanned consequences of data gathering and analysis (Brown, 2014). Moreover, harms may occur from unsophisticated

algorithms and faulty data generate high rates of false positives that might serve as a basis for baseless, stigmatising criminal investigations (Citron, 2007).

Another stream of challenges for e-government is related to automated decision-making. Many algorithms make decisions by finding associations, classifying, anomaly detection, and predicting. Overfitting and overgeneralisation may lead to poor classification/prediction and selection bias may occur as well (Monteith and Glenn, 2016). The automation of human decision-making is often justified by an alleged lack of bias in algorithms. An algorithm's design and functionality, however, reflects the values of its designer and intended uses while is difficult to detect social or technical bias. As (Mittelstadt et al., 2016) states "bias is a dimension of the decision-making itself, whereas discrimination describes the effects of a decision, in terms of adverse disproportionate impact resulting from algorithmic decision-making".

The structural dimension of the public sphere, includes among others the legal frameworks of classic democratic issues (Dahlgren, 2013). Political rights of participation are preconditions for elections while they are embodied in the unlimited validity of the right to freedom of speech and opinion and among others the freedom of speech, expression, of association, etc. (Merkel, 2004). Privacy' s public value stems also from its importance to the exercise of these political rights (Goold, 2010). Surveillance however, which is often employed for detecting and investigating possible criminal and terrorist behaviour, may disrupt this structural dimension.

Summarising, among others social media surveillance limits individuals' ability to regulate their social interactions and identities and it may inhibit the ability of both individuals and groups to express their views and on the possibilities for whistle-blowing and democratic activism. In addition, new surveillance technologies can lead to 'social sorting', where discrimination and privilege are entrenched through the unplanned consequences of data gathering and analysis.

4.2 Implications for e-democracy

In a democracy, the electoral regime has the function of making the access to public power positions of the state dependent on the results of open, competitive, fair elections (Merkel, 2004). But the new digital divide, between the big data rich and the big data poor (Boyd and Crawford, 2012) favours incumbents who already are in the possession of valuable data, also entrenched and moneyed candidates within parties, as well as the data–rich among existing parties (Tufekci, 2014).

Furthermore, the institutional core of political rights is the right to political communication and organisation, which are vital parts of a complete democratic regime (Merkel, 2004). However, the current big data ecosystem undermines those political rights which have the function both of enabling organised democratic elections and of furthering the unorganised pluralistic interests of complex societies. Data-driven campaigning might even be a form of cartelisation (Katz and Mair, 1995) where large parties erect barriers to protect their dominance from new entrants. The high expense of new campaigning techniques is a significant disadvantage for smaller and newer parties (Anstead, 2017).

According to Dahlgren (2013), a functioning public sphere is understood as aggregation of communicative spaces in society that permit (ideally unimpeded) the circulation of information, ideas, debates and also the formation of political will. These spaces also serve to facilitate communicative links between citizens and the power

holders of society. But the way modern mass media and social media function today obstruct this flow by becoming increasingly personalised and creating "*filter bubbles*". Media's capacity to circulate material that builds connections between otherwise diverse groups is not helped, but rather undermined, by the pressures toward personalisation (Couldry and Turow, 2014). Consequently, the own opinion is reinforced, while the ability to handle different points of view is reduced facilitating in this way the polarisation of society.

The 'filter bubbles' and financial means of potent influencers facilitates the spread of inaccurate, misleading or even wrong information. They become increasingly personalised, manipulative, and deceptive, spreading oversimplified messages or misinformation (Helbing and Klauser, 2019). Platforms' algorithms highlight popular information, especially if it has been shared by friends, surrounding us with content from relatively homogenous groups (Chesney and Citron, 2018).

For citizens it is increasingly hard to judge, which information can be trusted and why (Helbing and Klauser, 2019). 'Deep fakes' are not just a threat to specific individuals or entities. They have the capacity to pose threats to society in a variety of ways. Many actors will have sufficient interest to exploit the capacity of deep fakes to skew information and thus manipulate beliefs. The damage may extend to, among other things, distortion of democratic discourse on important policy questions, manipulation of elections etc. (Chesney and Citron, 2018).

Cases of misuse of technological affordances and personal data for political goals have been reported globally. Politicians may use microtargeting to manipulate voters. A party could use social media to expose xenophobic voters to information about the high crime rates amongst immigrants. Gorton (in Zuiderveen Borgesius et al., 2018) warns that microtargeting "turns citizens into objects of manipulation and undermines the public sphere by thwarting public deliberation, aggravating political polarisation, and facilitating the spread of misinformation". The targeted information does not even need to be true to maximise its impact. Political parties could also use microtargeting to suppress voter turnout for their opponents while voters may not even know a party's views on many topics (Zuiderveen Borgesius et al., 2018). In addition, data-driven politics is about communicating efficiently, talking to voters who are most useful to a campaign. However, inefficient targeting might lead to better democratic outcomes, as it could include more people in the electoral conversation. As systems become more efficient, that externality might be lost (Anstead, 2017).

In the concept of the public sphere (Dahlgren, 2013), the representational dimension refers to the output of the media, the mass media as well as 'minimedia' that target specific small groups via, for example, newsletters or campaign promotion materials. In these terms, questions and criteria may be raised about media output for political communication, including fairness, accuracy, completeness, pluralism of views, etc. All these characteristics of a democracy may be undermined through personalisation. Content is selected for citizens on the basis of criteria unknown to them and is calibrated not to their proximate selection decisions, but to big data–generated assumptions about where those citizens would want to focus their attention or where marketers need those citizens' attention to be focused (Couldry and Turow, 2014). At the end of 2017, the Cambridge Dictionary declared 'populism' the word of the year. At the same time, the dictionary defined populism as "political ideas and activities that are intended to get support of ordinary people by giving them what they want". The tools of democratic innovation (such as e-democracy, particularly a truly deliberative e-democracy) can reinforce

democratic participation. However, as De Blasio and Sorice (2018) mention, they often do the opposite, becoming effective enhancers (directly or indirectly) of populist tendencies.

It seems that the challenges for e-government associated to surveillance are to a large extent related to the inhibition of the ability of both individuals and groups to express their views and to automated decision-making which potentially may lead to discrimination. While the challenges for e-democracy, are significantly related to personalisation and targeting technologies used for political communication which between others may facilitate the spread of inaccurate, misleading or even wrong information aggravating political polarisation. Also, as Zarsky (2019) claims, much of the justification for voting in a democratic state is premised on the expectation of autonomous voters but it seems that voters' personal data may be a fuel of the manipulation. Accepting the notion that individuals could be systematically and easily manipulated shakes many of the foundational assumptions related to democracy.

Conclusively, the tools of democratic innovation (e.g., e-democracy) have the potential to reinforce democratic participation, but when misused, they often have the opposite effect. All above-described processes are fuelled by personal data consequently they are facilitated by the decreased levels of privacy. For instance, as Martin (2020) claims, manipulation is only possible because someone have intimate knowledge of individuals as to what renders them vulnerable in their decision making. Therefore, it may be concluded that data protection may play a key role in shielding democracy and solutions for enhancing privacy and data protection may mitigate the impact on society and democracy occurring from abuses of personal and sensitive data. Next, we examine several technical solutions which among others allow for example to political activists to maintain their anonymity in order to freely express their opinions.

5 Mitigating privacy threats

There is a tendency to refer to the right to data protection as an expression of the right to privacy, but the distinction between both rights in the EU Charter of Fundamental Rights is not purely symbolic (Kokott and Sobotta, 2013). The physical dimensions of privacy (the home and the body) are complemented by non-physical dimensions. This has two aspects: privacy with respect to relations and privacy with respect to information. Informational privacy is the dimension which became relevant in modern societies.

The Electronic Privacy and Information Centre (EPIC) distinguishes four models of privacy protection. Due space limitation we examine here one of these models of data protection: technological solutions for enhancing big data protection.

5.1 Privacy-enhancing technologies (PETs)

The privacy-enhancing technologies (PETs) aim at protecting the individual's privacy by the use efficient privacy preserving algorithms, applications, systems and services in various domains and environments. They provide anonymity, pseudonymity, unlinkability, and unobservability of users as well as of data subjects (Heurix et al., 2015; Kaaniche et al., 2020) by ensuring various levels of obfuscation of content and metadata (Makin and Ireland, 2020). A well-known definition (adopted later by the European Commission), was given by Borking and Blarkom et al. in (ENISA, 2014) as follows:

"Privacy-Enhancing Technologies is a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system".

In the last decade, numerous PETs were proposed for network traffic anonymisation, identity management, or anonymous data storage based on different building blocks, including cryptographic primitives or the separation of information (Heurix et al., 2015; Kaaniche et al., 2020). Kaaniche et al. (2020) classifies PETs into three different groups. The first group (user-side techniques) requires end-user to manage his identity protection by himself by installing specific softwares to control attributes disclosure up-to the certification of attribute properties. This group of PETs includes anti tracking technologies and privacy preserving certification. The second group (server-side techniques) refers to techniques where the server is involved in data processing by anonymising databases or by processing encrypted data at the request of costumers. This group includes statistical disclosure control techniques and self-destructing systems. Obfuscation and privacy preserving computation relate both user-side and server-side privacy enhancing techniques while they are set up according to the system's design purposes. The third group (channel-side techniques), is associated with the quality of the channel between the user and the server (mediated and/or encrypted) or the quality of the exchanged data which can be voluntarily debased. This group of techniques include secure communications and Trusted Third Party approaches.

Besides law enforcement, everyday citizens might be interested in PETs to avoid governmental and corporate surveillance. After Edward Snowden disclosures, interest and use of PETs such as Google search queries for Tor, as well as use of both the private search engine DuckDuckGo and the Tor network increased significantly. Several governments (e.g., in China and Syria) prevent users from accessing specific content online, using various techniques such as deep packet inspection, IP address geolocation and filtering systems to manage internet traffic. In order to gain access, Syrians use Tor, VPNs, web/sock proxies and BitTorrent (Makin and Ireland, 2020). Tor (The Tor Project | Privacy & Freedom Online,) one of the best-known anonymity technology an onion routing based, low latency network which is built over slightly over 6000 volunteer-provided relay servers (Welcome to Tor Metrics) (Wolf, 2018). Jardine's (2018) research results show that political repression emerges as a significant predictor of Tor network usage. The emerging relationship is consistently U-shaped, with political repression driving Tor network usage most in both highly liberal and highly repressive regimes. Consequently, the technology of Tor is useful for political dissidents and for citizens trying to exercise their basic political rights as well. In generally, PETs, may be solution for people to freely share, access and discuss information and content that is contrary to the political, religious or social views of governments (Makin and Ireland, 2020).

In regard to the identity management systems, they still rely on what Cameron (in (Dunphy and Petitcolas, 2018) called a decade ago a 'patchwork of identity one-offs' as they comprise several types of identity management systems that are restricted to specific domains and do not interact enough with one another. This centralisation however currently faces challenges due to the increasing regularity of data breaches that lead to reputation damage; identity fraud; and above all, a loss of privacy for all concerned. These breaches underline a lack of control and ownership that end users experience with their digital identities. According to Dunphy and Petitcolas (2018) the alternative solutions concerning the above-mentioned problems may involve decentralised

approaches such as the distributed ledger technology which are used for supporting Bitcoin and which do not require a central authority to validate transactions of its native cryptocurrency. "The distributed ledger itself is an append-only shared record of transactions that is maintained by entities on a peer-to-peer network, whereas the often-cited 'blockchain' is a cryptographic data structure that is often instrumental in DLTs and is constructed through cryptographic hashing of blocks of transactions". One of the benefits of this approach is that users cannot lose control of their digital identities if they lose access to the services of a particular identity provider/broker.

Decentralised approaches for protecting privacy of activists are also proposed by Poblet (2018). There are today numerous examples of movements which leveraged social media for protest and coordination. Some of the examples are the Iranian Green Revolution of 2009 and the Arab Spring of 2011. In both cases state governments reacted quickly by blocking access to social networks and shutting down the internet while repression of bloggers and digital activists followed (Poblet, 2018). Social media on one hand satisfy a major communication need in any form of activism as allows spreading out open information to mass-scale audiences. On the other hand, however, they are deficient in ensuring private, secured communications between activists when organise their actions in hostile environments. Poblet (2018) analyses the adoption of distributed and privacy-enhancing technologies in the context of civic and political activism. According to the author "civic and political activism advocating widespread use of distributed technologies, encryption and privacy-enhancing protocols to bring political change is perhaps a sign of a (re)emerging trend of the internet: the horizontal, decentralised network of networks that Vint Cerf and Tim Berners-Lee, inventors of its core technologies, initially envisioned".

There are numerous privacy- enhancing tools for online and mobile protection, such as anti-tracking, encryption, protocols for anonymous communications, attribute based credentials and private search of databases etc., which could offer valuable support in avoiding unwanted processing of personal data. However, apart from a few exceptions, such as encryption which is widely used, PETs have not become a standard and widely used component in system design (Danezis et al., 2015; ENISA, 2014).

Problems related to PETSs stem from the uncoupling between the these technologies and the practice of systematic engineering most engineers when they need to deal with privacy issues they turn to crafting tailored solutions rather than choosing the systematic and economic application of existent solutions drawn from the state of the technique (Martin, 2020). Moreover, privacy played no significant role in the design of today's internet infrastructure hence actual anonymisation services for example are organised as separate overlay networks. Furthermore, many PETs cause a high overhead and they cannot be activated by default for instance by the internet service provider. Consequently, the loss of comfort outweighs the benefits of privacy and became unacceptable for many users.

In Harborth et al. (2017) the authors specify

> "that PETs are only able to reach the mass market when they are standardised and usable without any action of the user ("zero-effort") and work so efficiently that they do not cause any noticeable limitation on the quality of service (especially regarding latency and bandwidth). This is particularly important considering the privacy attitudes and behaviours of regular users. To reach this goal, PETs need to be firmly integrated in the internet's infrastructure".

Harborth et al. (2017) present the objectives for the three tackled technical areas of ISP-based, network overlay-based, and 5G network anonymisation techniques.

5.2 Privacy by design

Privacy by design was first presented by Cavoukian (2009), and included the notion of embedding privacy measures and privacy enhancing technologies directly into the design of information technologies and systems. Today, privacy by design is multifaceted concept which in legal documents it is described as a general principle while engineers are equating it with the use of specific privacy enhancing technologies (ENISA, 2014). According to this approach 'privacy must be incorporated into networked data systems and technologies, by default. Privacy must become integral to organisational priorities, project objectives, design processes, and planning operations. Privacy must be embedded into every standard, protocol and process that touches our lives' (Cavoukian, 2009).

Related to the potential for bias and discrimination in automated algorithmic decision-making, Tene and Polonetsky (2017) argue that a distinction should be drawn between 'policy neutral algorithms' which may reflect in some cases existing societal biases and inequities and 'policy directed algorithms' which are purposely engineered to correct for apparent bias and discrimination. For example, if a search engine does not show opposing viewpoints on an issue, it should let users know that they are seeing only one side of the respective issue. As seen in previous sections in this paper, various digital platforms pose concerns around algorithm-based digital content manipulation. Smart user interfaces design, without editing algorithmic results may be a solution for incentivise users away from prejudice and bias. For example, while not editing results of searches for 'beautiful babies' or 'beautiful women', which yield results overwhelmingly dominated by white individuals, Google adds a toolbar with buttons enabling a user to easily select images of 'African American', 'Hispanic', 'Native American', 'Arab', and other minorities. Hence, without editing the results of a policy-neutral algorithm, Google provides an opportunity for multicultural outcomes. Another tool is differential privacy, which ensures that with respect to any classification, an observer cannot determine whether or not a given individual was a member of a protected group (Tene and Polonetsky, 2017).

With regard to group privacy, traditional approaches are based on the idea of hiding the individual in a 'crowd', and aggregating data to protect information about an individual. These techniques derive from k-anonymity where a set of data records is obfuscated in a way that hides the record of any individual record among at least k other records. The differential privacy approach obfuscates data by adding enough noise so that any query on the data will return the same answer irrespective of whether the individual record is in the dataset or not. But these methods to protect privacy are focused on personal privacy and ignore or even may unintentionally reveal group privacy. K-anonymity, almost by design, reveals group characteristics, and thus may compromise group privacy. As Suh et al. (2018) mentions in k-anonymity, the algorithm attempts to hide the individual in a group of size at least k consequently this approach inherently identifies and hence reveals a group. Yang et al. (2017) propose a method of 'anti-data mining' on group privacy information which eliminates and destroys the minable characteristic and group specificity of original data.

Summarising, some of the technical solutions which may potentially play a significant role in shielding democracy are encryption, PETs like Google search queries

for Tor, the use of both the private search engine DuckDuckGo and the Tor network, decentralised approaches such as the distributed ledger technology which are used for supporting Bitcoin. Using the above means, people may freely share, access and discuss information. Also, related to the potential for bias and discrimination in automated algorithmic decision-making, a distinction between 'policy neutral algorithms' and 'policy directed algorithms' and smart user interfaces design, without editing algorithmic results may be a solution for incentivise users away from prejudice and bias.

6 Discussion and conclusions

Traditionally, data collection has been limited by human perception and cognition. Today as Zuboff (2019) writes, "three of the world's seven billion people are now computer-mediated in a wide range of their daily activities far beyond the traditional boundaries of the workplace". The amount of generated big data on a daily basis can be seen itself as a form of surveillance. Through big data technologies more personal and highly detailed data may be collected, inferred, processed and analysed than at any time in the history. Even more, these data may be stored for ever thus it may be analysed any time in the future for any goal with unknown today results and implications. All the above are changing the perspectives privacy and the implication of privacy threats have to be examined.

One of the most important questions relates to technology concerns its impact on democracy. As people understand technologies as functions for accomplishing tasks, they consider their consequences as unintentional social and environmental consequences. For that reason, the social structure aspect of information and communication technologies with a key role regarding democracy may be ignored, because people understand it as functions.

Privacy stakeholders include governments, elected officials, media (and major online portals), political parties and interest groups, civil society organisations, international governmental organisations and citizens/voters. This paper identifies privacy threats for individuals and groups stemming from big data applications for surveillance. Furthermore, we explore the challenges of these threats for e-government arguing, that automatic decision-making may disfavour various groups and individuals compromising equality, a basic democratic value. Also, we argue that without privacy is not possible for citizens to feel free and autonomous and we underline that social media surveillance, potentially may inhibit the ability of both individuals and groups to freely express their views undermining this way another democratic value.

Moreover, we identify privacy threats stemming from crowdsourcing applications showing that they vary depending on the used methods and the context in which the citizens'- sourcing methods are used. We indicate the challenges of these privacy threats for e-democracy underling among others the implications of targeting, procedure which also may lead to discrimination as some groups might be provided with different opportunities. The third issue we have examined is the challenges for e-democracy related to political communication. We have identified privacy threats and implications stemming from the use of big data technologies for political communications. We argue among others that personalisation may impede the free circulation of information, ideas, debates and the formation of political will may facilitate the polarisation of the society, the spread of inaccurate, misleading or even wrong information allowing manipulation

and distortion of democratic discourse. We indicate that personalisation may challenge e-democracy through compromising basic democratic values such as fairness, accuracy, completeness, pluralism of views. We also argue that one of the challenges for e-democracy of political targeting is that it undermines the democratic public sphere by thwarting public deliberation, aggravating political polarisation, and facilitating the spread of misinformation.

However, in the case of political targeting for example, in west democracies there are only few parties which have the necessary resources for microtargeting in order to reach individuals with persuasive messages. Also, in comparison with the US, European political parties have far less access to the types of data required to target voters and much stronger data protection laws. Consequently, it is easier and cheaper (at least up to now) to identify and target and manipulate groups than individuals. For example (Kramer et al., 2014) in an experiment with people who use Facebook, showed that big data owners may influence people on a mass scale (two groups of 155,000 participants). When positive expressions were reduced in the news feed, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. In our opinion, in the context of this paper, profiling and targeting are processes concerning groups rather individuals. Consequently, the consequences for democracy occurring from these processes have to be approached and investigated from a collective or group perspective as well. Moreover, considering the complexity of the privacy threats and the fact that there are privacy harms stemming from group privacy violations there is a need to determine in further research haw and whose interests and rights are affected.

The undoubtful benefits for humanity of big data analysis indicate that big data technologies are here to stay and will be increasingly used to support government and democracy. As Zarsky (2017) argues,

> "big data analysis of personal information both affects and is affected by the extent of data protection policy. On the one hand, big data analytics compromises the citizens' privacy rights hence stricter enforcement of privacy low is required. On the other hand, though, stringent data protection laws impede the flow of personal data, as well as the ways it could be analysed and used. In other words, stricter data protection and privacy laws compromise the growth of the big data industry and the benefits to be derived from it".

Further research is needed to determine in which point the benefits outweigh the risks.

We also examine several relevant privacy enhancing technologies proposed in literature. We indicate that these technologies may help people to freely share, access and discuss information and content that is contrary to the political, religious or social views of governments shielding this way basic democratic values. However, in order to be more useful, they have to be integrated in the internet's infrastructure. 'Privacy by Design', a legal requirement under GDPR, is a multifaceted and more spherical solution for enhancing privacy. It is an approach according to with privacy must be embedded into every standard, protocol and process. There is a developing understanding that innovation must be approached from the perspective of 'Privacy by Design'. According to Riva and Barry (2019) privacy by design approach would empower users with full control over accessible content and related profiling activities. Users should be able to select the type of interaction they want to perform while being empowered with more consistent information over their personalised content.

Overall, we indicate in this research the challenges for e-government related to privacy implications for democratic values stemming from big data surveillance, and the

challenges for e-democracy related to privacy implications associated to big data technologies for crowdsourcing and political communications. One of the key factors facilitating misinformation, manipulation, disinformation and discrimination is decreased privacy.

Finally, we probably cannot claim absolute rights against government surveillance and at the same time say wanting the government to ensure order. As Westin (2003) states, "managing these tensions among privacy, disclosure, and surveillance in a way that preserves civility and democracy, and copes successfully with changing social values, technologies, and economic conditions, is the central challenge of contemporary privacy definition and protection".

References

Aitamurto, T. (2012) *Crowdsourcing for Democracy: A New Era in Policy-Making Crowdsourcing for Democracy: A New Era In Policy-Making*, Publications of the Committee for the Future, Parliament of Finland 1,

Anstead, N. (2017) 'Data-driven campaigning in the 2015 United Kingdom general election', *The International Journal of Press/Politics*, Vol. 22, pp.294–313, https://doi.org/10.1177/1940161217706163

Aral, S. and Eckles, D. (2019) 'Protecting elections from social media manipulation', *Science*, Vol. 365, pp.858–861, https://doi.org/10.1126/science.aaw8243

Bindu, N., Sankar, C.P. and Kumar, K.S. (2019) 'From conventional governance to e-democracy: tracing the evolution of e-governance research trends using network analysis tools', *Government Information Quarterly*, Vol. 36, pp.385–399, https://doi.org/10.1016/j.giq.2019.02.005

Boehme-Neßler, V. (2016) 'Privacy: A matter of democracy. why democracy needs privacy and data protection', *International Data Privacy Law*, Vol. 6, pp.222–229.

Bolsover, G. and Howard, P. (2017) 'Computational propaganda and political big data: moving toward a more critical research agenda', *Big Data*, Vol. 5, pp.273–276, https://doi.org/10.1089/big.2017.29024.cpr

Boyd, D. and Crawford, K. (2012) 'Critical questions for big data', *Information, Communication and Society*, Vol. 15, pp.662–679, https://doi.org/10.1080/1369118X.2012.678878

Bradshaw, S. and Howard, P.N. (2018) *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*, The Computational Propaganda Project.

Brown, I. (2014) 'Social media surveillance, in: the international encyclopedia of digital communication and society', *American Cancer Society*, pp.1–7, https://doi.org/10.1002/9781118767771.wbiedcs122

Calo, R. (2011) 'The boundaries of privacy harm', *Ind. LJ*, Vol. 86, pp.1131–1162.

Cavoukian, A. (2009) *Privacy by Design: The* 7 *Foundational Principles*, Information and Privacy Commissioner of Ontario, Canada 5.

Charalabidis, Y., Loukis, E.N., Androutsopoulou, A., Karkaletsis, V. and Triantafillou, A. (2014) 'Passive crowdsourcing in government using social media', *Transforming Government: People, Process and Policy*, Vol. 8, pp.283–308, https://doi.org/10.1108/TG-09-2013-0035

Chen, Y-Y., Hsu, W.H. and Liao, H-Y.M. (2012) 'Discovering informative social subgraphs and predicting pairwise relationships from group photos', *Proceedings of the 20th ACM International Conference on Multimedia, MM '12*, ACM, New York, NY, USA, pp.669–678, https://doi.org/10.1145/2393347.2393439

Chesney, R. and Citron, D.K. (2018) *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security (SSRN Scholarly Paper issue ID 3213954)*, Social Science Research Network, Rochester, NY.

Citron, D.K. (2007) 'Technological due process', *Wash. UL Rev.*, Vol. 85, pp.1249–1313.

Citron, D.K. and Pasquale, F. (2014) 'The scored society: due process for automated predictions', *Wash. L. Rev.*, Vol. 89, p.1.

Clift, S. (2003) *E-Democracy, e-Governance and Public Net-Work*, Artículo en línea. Publicus. net.

Couldry, N. and Turow, J. (2014) 'Advertising, big data and the clearance of the public realm: marketers' new approaches to the content subsidy', *International Journal of Communication*, Vol. 8, pp.1710–1726.

Cuquet, M., Vega-Gorgojo, G., Lammerant, H. and Finn, R. (2017) *Societal Impacts of Big Data: Challenges and Opportunities in Europe*, arXiv preprint arXiv: 1704.03361.

Dahlgren, P. (2013) *The Political Web: Media, Participation and Alternative Democracy*, Palgrave Macmillan, New York, NY.

Dalton, R.J. (2016) 'The potential of big data for the cross-national study of political behavior', *International Journal of Sociology*, Vol. 46, pp.8–20, https://doi.org/10.1080/00207659.2016.1130410

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J-H., Metayer, D.L., Tirtea, R. and Schiffner, S. (2015) *Privacy and Data Protection by Design – From Policy to Engineering*. arXiv: 1501.03726 [cs], https://doi.org/10.2824/38623

De Blasio, E.D. and Sorice, M. (2018) 'Populisms among technology, e-democracy and the depoliticisation process', *Revista Internacional de Sociología*, Vol. 76, No. 4, p.e109, https://doi.org/10.3989/ris.2018.76.4.18.005

Diamantopoulou, V., Androutsopoulou, A., Gritzalis, S. and Charalabidis, Y. (2018) 'An assessment of privacy preservation in crowdsourcing approaches: towards GDPR compliance', Presented at the *2018 12th International Conference on Research Challenges in Information Science (RCIS)*, 29–31 May, Nantes, France, pp.1–9, https://doi.org/10.1109/RCIS.2018.8406643

Dunphy, P. and Petitcolas, F.A.P. (2018) 'A first look at identity management schemes on the blockchain', *IEEE Security Privacy*, Vol. 16, pp.20–29, https://doi.org/10.1109/MSP.2018.3111247

Dwivedi, Y.K., Weerakkody, V. and Janssen, M. (2012) 'Moving towards maturity: challenges to successful e-government implementation and diffusion', *SIGMIS Database*, Vol. 42, pp.11–22, https://doi.org/10.1145/2096140.2096142

Eide, E. (2019) 'Chilling effects on free expression: surveillance, threats and harassment', in Krøvel, R. and Thowsen, M. (Eds.): *Making Transparency Possible. An Interdisciplinary Dialogue*, Cappelen Damm Akademisk, Oslo, pp.227–242.

Eijkman, Q. (2017) 'Indiscriminate bulk data interception and group privacy: do human rights organisations retaliate through strategic litigation?', *Group Privacy*, Springer, Cham, pp.123–138.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V.R., Tsou, A. and Weingart, S., Sugimoto, C.R. (2015) 'Big data, bigger dilemmas: a critical review', *Journal of the Association for Information Science and Technology*, Vol. 66, pp.1523–1545, https://doi.org/10.1002/asi.23294

ENISA (2014) *Privacy and Data Protection by Design [WWW Document]*, URL https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design (Accessed 6 September, 2020).

Feng, W., Yan, Z., Zhang, H., Zeng, K., Xiao, Y. and Hou, Y.T. (2018) 'A survey on security, privacy, and trust in mobile crowdsourcing', *IEEE Internet of Things Journal*, Vol. 5, pp.2971–2992, https://doi.org/10.1109/JIOT.2017.2765699

Fisher, L.E. (2004) 'Guilt by expressive association: political profiling, surveillance and the privacy of groups', *Ariz. L. Rev.*, Vol. 46, p.621.

Friedewald, M., Burgess, J.P., Čas, J., Bellanova, R. and Peissl, W. (Eds.) (2017) *Surveillance, Privacy and Security: Citizens' Perspectives*, 1st ed., Routledge, London, https://doi.org/10.4324/9781315619309

Gandomi, A. and Haider, M. (2015) 'Beyond the hype: big data concepts, methods, and analytics', *International Journal of Information Management*, Vol. 35, pp.137–144, https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Ghermandi, A. and Sinclair, M. (2019) 'Passive crowdsourcing of social media in environmental research: a systematic map', *Global Environmental Change*, Vol. 55, pp.36–47.

Gibson, R.K. (2015) 'Party change, social media and the rise of 'citizen-initiated' campaigning', *Party Politics*, Vol. 21, pp.183–197, https://doi.org/10.1177/1354068812472575

Goold, B.J. (2010) *How Much Surveillance is Too Much? Some Thoughts on Surveillance, Democracy, and the Political Value of Privacy (SSRN Scholarly Paper issue ID 1876069)*, Social Science Research Network, Rochester, NY.

Grassegger, H. and Krogerus, M. (2017) *The Data That Turned the World Upside Down*, Luettavissa, http://motherboard.vice.com/read/big-data-cambridge-analytica-brexit-trump.Luettu, 28, 2017.

Grimmer, J. (2015) 'We are all social scientists now: how big data, machine learning, and causal inference work together', *PS: Political Science and Politics*, Vol. 48, pp.80–83, https://doi.org/10.1017/S1049096514001784

Grönlund, Å. and Horan, T.A. (2005) 'Introducing e-gov: history, definitions, and issues', *Communications of the Association for Information Systems*, Vol. 15, p.39.

Halder, B. (2014) *Measuring Security, Privacy and Data Protection in Crowdsourcing (SSRN Scholarly Paper issue ID 2568818)*, Social Science Research Network, Rochester, NY, https://doi.org/10.2139/ssrn.2568818

Harborth, D., Herrmann, D., Köpsell, S., Pape, S., Roth, C., Federrath, H., Kesdogan, D. and Rannenberg, K. (2017) *Integrating Privacy-Enhancing Technologies into the Internet Infrastructure*, arXiv: 1711.07220 [cs].

Helbing, D. and Klauser, S. (2019) How to Make Democracy Work in the Digital Age, in: Towards Digital Enlightenment, Springer, Cham, pp.157–162.

Henman, P. (2005) 'E-government, targeting and data profiling: policy and ethical issues of differential treatment', *Journal of E-Government*, Vol. 2, pp.79–98.

Heurix, J., Zimmermann, P., Neubauer, T. and Fenz, S. (2015) 'A taxonomy for privacy enhancing technologies', *Computers and Security*, Vol. 53, pp.1–17.

Hildebrandt, M. (2012) 'The dawn of a critical transparency right for the profiling era', in Bus, J. (Ed.): *Digital Enlightenment Yearbook 2012*, IOS Press, Amsterdam, pp.41–56.

Höchtl, J., Parycek, P. and Schöllhammer, R. (2016) 'Big data in the policy cycle: policy decision making in the digital era', *Journal of Organizational Computing and Electronic Commerce*, Vol. 26, pp.147–169, https://doi.org/10.1080/10919392.2015.1125187

Ingrams, A., Manoharan, A., Schmidthuber, L. and Holzer, M. (2020) 'Stages and determinants of e-government development: a twelve-year longitudinal study of global cities', *International Public Management Journal*, Vol. 23, pp.731–769.

Isaak, J. and Hanna, M.J. (2018) 'User data privacy: Facebook, Cambridge analytica, and privacy protection', *Computer*, Vol. 51, pp.56–59.

Jain, P., Gyanchandani, M. and Khare, N. (2016) 'Big data privacy: a technological perspective and review', *J. Big Data*, Vol. 3, p.25, https://doi.org/10.1186/s40537-016-0059-y

Jardine, E. (2018) 'Tor, what is it good for? political repression and the use of online anonymity-granting technologies', *New Media and Society*, Vol. 20, pp.435–452.

Jenkins, J.C., Slomczynski, K.M. and Dubrow, J.K. (2016) 'Political behavior and big data', *International Journal of Sociology*, Vol. 46, pp.1–7, https://doi.org/10.1080/00207659.2016.1130409

Joseph, R.C. and Johnson, N.A. (2013) 'Big data and transformational government', *IT Professional*, Vol. 15, pp.43–48, https://doi.org/10.1109/MITP.2013.61

Kaaniche, N., Laurent, M. and Belguith, S. (2020) 'Privacy enhancing technologies for solving the privacy-personalization paradox: taxonomy and survey', *Journal of Network and Computer Applications*, p.102807.

Katz, R.S. and Mair, P. (1995) 'Changing models of party organization and party democracy: the emergence of the cartel party', *Party Politics*, Vol. 1, pp.5–28, https://doi.org/10.1177/1354068895001001001

Kim, G-H., Trimi, S. and Chung, J-H. (2014) 'Big-data applications in the government sector', *Commun. ACM*, Vol. 57, pp.78–85, https://doi.org/10.1145/2500873

Klievink, B., Romijn, B-J., Cunningham, S. and de Bruijn, H. (2017) 'Big data in the public sector: uncertainties and readiness', *Information Systems Frontiers*, Vol. 19, pp.267–283.

Kokott, J. and Sobotta, C. (2013) 'The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR', *International Data Privacy Law*, Vol. 3, pp.222–228.

Kondor, D., Hashemian, B., Montjoye, Y-a.d. and Ratti, C. (2020) Towards Matching User Mobility Traces in Large-scale datasets. IEEE Transactions on Big Data, Vol. 6, pp.714–726, https://doi.org/10.1109/TBDATA.2018.2871693

Körner, K. (2019) *Digital Politics, A.I., Big Data and the Future of Democracy*, EU Monitor Digital Economy and Structural Change.

Kosinski, M., Stillwell, D. and Graepel, T. (2013) 'Private traits and attributes are predictable from digital records of human behavior', *Proceedings of the National Academy of Sciences*, Vol. 110, pp.5802–5805.

Kramer, A.D.I., Guillory, J.E. and Hancock, J.T. (2014) 'Experimental evidence of massive-scale emotional contagion through social networks', *PNAS*, Vol. 111, pp.8788–8790, https://doi.org/10.1073/pnas.1320040111

Kreiss, D. and Jasinski, C. (2016) 'The tech industry meets presidential politics: explaining the democratic party's Technological Advantage in Electoral Campaigning, 2004–2012', *Political Communication*, Vol. 33, pp.544–562, https://doi.org/10.1080/10584609.2015.1121941

Lee, D., Goel, A., Aitamurto, T. and Landemore, H. (2014) 'Crowdsourcing for participatory democracies: efficient elicitation of social choice functions', *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 2, No. 1, pp.133–142, Retrieved from https://ojs.aaai.org/index.php/HCOMP/article/view/13150

Leese, M. (2014) 'The new profiling: algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European union', *Security Dialogue*, Vol. 45, pp.494–511.

Lin, L. and Hou, Z. (2020) 'Combat COVID-19 with artificial intelligence and big data', *Journal of Travel Medicine*, Vol. 27, No. 5, July, taaa080, https://doi.org/10.1093/jtm/taaa080

Lindner, R. and Aichholzer, G. (2020) *E-Democracy: Conceptual Foundations and Recent Trends, in: European E-Democracy in Practice*, Springer, Cham, pp.11–45.

Liu, H.K., Tang, M. and Chen, K-H. (2020) 'Public decision making: connecting artificial intelligence and crowds', *The 21st Annual International Conference on Digital Government Research, Dg.o. '20*, Association for Computing Machinery, Seoul, Republic of Korea, pp.214–222, https://doi.org/10.1145/3396956.3396965

Lupton, D. and Michael, M. (2017) 'Depends on who's got the data': public understandings of personal digital dataveillance', *Surveillance & Society*, Vol. 15, No. 2, pp.254–268.

Lyon, D. (2014) 'Surveillance, Snowden, and big data: capacities, consequences, critique', *Big Data and Society*, Vol. 1, 2053951714541861.

Magrani, E. (2020) *Hacking the Electorate: Thoughts on Misinformation and Personal Data Protection*, Konrad Adenauer Stiftung, http://www.jstor.org/stable/resrep25290

Mai, J-E. (2016) 'Big data privacy: the datafication of personal information', *The Information Society*, Vol. 32, pp.192–199, https://doi.org/10.1080/01972243.2016.1153010

Makin, D.A. and Ireland, L. (2020) 'The secret life of PETs: a cross-sectional analysis of interest in privacy enhancing technologies', *Policing: An International Journal*, Vol. 43, No. 1, pp.121–136, https://doi.org/10.1108/PIJPSM-07-2019-0124

Manheim, K. and Kaplan, L. (2019) 'Artificial intelligence: risks to privacy and democracy', *Yale Journal of Law & Technology*, Vol. 21, No. 1, pp.106–189.

Margulis, S.T. (2003) 'On the status and contribution of Westin's and Altman's theories of privacy', *Journal of Social Issues*, Vol. 59, pp.411–429, https://doi.org/10.1111/1540-4560.00071

Martin, K.E. (2020) *Manipulation, Choice, and Privacy (SSRN Scholarly Paper issue ID 3491696)*, Social Science Research Network, Rochester, NY, https://doi.org/10.2139/ssrn.3491696

McDermott, Y. (2017) 'Conceptualising the right to data protection in an era of big data', *Big Data and Society*, Vol. 4, 2053951716686994, https://doi.org/10.1177/2053951716686994

Merkel, P.S. (2004) 'Embedded and defective democracies', *Democratization*, Vol. 11, pp.33–58, https://doi.org/10.1080/13510340412331304598

Mineraud, J., Lancerin, F., Balasubramaniam, S., Conti, M. and Tarkoma, S. (2015) 'You are AIRing too much: assessing the privacy of users in crowdsourcing environmental data', *2015 IEEE Trustcom/BigDataSE/ISPA. Presented at the 2015 IEEE Trustcom/BigDataSE/ISPA,* pp.523–530, https://doi.org/10.1109/Trustcom.2015.415

Mitrou, L., Drogkaris, P., Leventakis, G., Friedewald, M., Burgess, P. J., Čas, J., … and Peissl, W. (2016) 'Legal and social aspects of surveillance technologies: CCTV in Greece or an attempt to explain some divergent findings of PACT's survey', *Surveillance, Privacy and Security: Citizens Perspectives*, Routledge, London, pp.123–138.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016) 'The ethics of algorithms: mapping the debate', *Big Data and Society*, Vol. 3, 2053951716679679, https://doi.org/10.1177/2053951716679679

Moerel, L. and van der Wolk, A. (2017) *Big Data Analytics Under the EU General Data Protection Regulation*, This paper has been first published in Dutch by SDU Uitgevers in.

Monteith, S. and Glenn, T. (2016) 'Automated decision-making and big data: concerns for people with mental illness', *Curr. Psychiatry Rep.*, Vol. 18, p.112, https://doi.org/10.1007/s11920-016-0746-6

Mulligan, D.K., Koopman, C. and Doty, N. (2016) 'Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 374, p.20160118.

Munyoka, W. and Maharaj, M.S. (2019) 'Privacy, security, trust, risk and optimism bias in e-government use: the case of two southern African development community countries', *South African Journal of Information Management*, Vol. 21, pp.1–9.

Narayanan, A., Huey, J. and Felten, E.W. (2016) 'A precautionary approach to big data privacy', in Gutwirth, S., Leenes, R. and De Hert, P. (Eds.): *Data Protection on the Move: Current Developments in ICT and Privacy/Data Protection, Law, Governance and Technology Series*, Springer Netherlands, Dordrecht, pp.357–385, https://doi.org/10.1007/978-94-017-7376-8_13

Netchaeva, I. (2002) *E-Government and E-Democracy: A Comparison of Opportunities in the North and South. Gazette (Leiden*, Netherlands), Vol. 64, pp.467–477, https://doi.org/10.1177/1748048502064005601

Nickerson, D.W. and Rogers, T. (2014) 'Political campaigns and big data', *The Journal of Economic Perspectives*, Vol. 28, pp.51–73.

Norris, D.F. (2010) 'e-government. not e-governance. not e-democracy not now! not ever?', *Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance, ICEGOV '10*, Association for Computing Machinery, Beijing, China, pp.339–346, https://doi.org/10.1145/1930321.1930391

Palvia, S. and Sharma, S. (2007) 'E-government and e-governance: definitions/domain framework and status around the world', *Proceedings of the Fifth International Conference on e-Governance*, 28–30 December, Hyderabad, Available at: www.iceg.net/2007/books/1/1_369.pdf (Accessed 12 august, 2018).

Perez, O., Bar-Ilan, J., Gazit, T., Aharony, N., Amichai-Hamburger, Y. and Bronstein, J. (2018) 'The prospects of E-democracy: an experimental study of collaborative E-rulemaking', *Journal of Information Technology and Politics*, Vol. 15, pp.278–299, https://doi.org/10.1080/19331681.2018.1485605

Poblet, M. (2018) *Distributed, Privacy-Enhancing Technologies in the 2017 Catalan Referendum on Independence: New Tactics and Models of Participatory Democracy*, FM, https://doi.org/10.5210/fm.v23i12.9402

Polonetsky, J. and Tene, O. (2013) 'Privacy and big data: making ends meet', *Stan. L. Rev. Online*, Vol. 66, p.25.

Raymond, N.A. (2017) 'Beyond 'do no harm' and individual consent: reckoning with the emerging ethical challenges of civil society's use of data', *Group Privacy*, Springer, Cham, pp.67–82.

Riva, G.M. and Barry, M. (2019) 'Net neutrality matters: privacy antibodies for information monopolies and mass profiling | neutralidade da rede importa: anticorpos de privacidade para monopólios de informação e profiling em massa', *Revista Publicum*, Vol. 5, pp.7–35, https://doi.org/10.12957/publicum.2019.47199

Shamsi, J.A. and Khojaye, M.A. (2018) 'Understanding privacy violations in big data systems', *IT Professional*, Vol. 20, pp.73–81, https://doi.org/10.1109/MITP.2018.032501750

Solmaz, G., Fürst, J., Aytaç, S. and Wu, F-J. (2020) 'Group-in: group inference from wireless traces of mobile devices', *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, IEEE, Sydney, Australia, pp.157–168.

Solove, D.J. (2003) 'Reconstructing electronic surveillance law', *Geo. Wash. L. Rev.*, Vol. 72, pp.1264–1270.

Solove, D.J. (2005) *A Taxonomy of Privacy (SSRN Scholarly Paper No. ID 667622)*, Social Science Research Network, Rochester, NY.

Solove, D.J. (2013) 'Introduction: Privacy self-management and the consent dilemma', *Harvard Law Review*, Vol. 126, No. 7, pp.1880–1904.

Suh, J.J., Metzger, M.J., Reid, S.A. and El Abbadi, A. (2018) 'Distinguishing group privacy from personal privacy: the effect of group inference technologies on privacy perceptions and behaviors', *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2(CSCW), pp.1–22.

Sundberg, L. (2019) 'Electronic government: towards e-democracy or democracy at risk?', *Safety Science*, Vol. 118, pp.22–32, https://doi.org/10.1016/j.ssci.2019.04.030

Taeihagh, A. (2017) 'Crowdsourcing: A new tool for policy-making?', *Policy Sciences*, Vol. 50, pp.629–647.

Taylor, L. (2017) 'Safety in numbers? Group privacy and big data analytics in the developing world', *Group Privacy*, Springer, Cham, pp.13–36.

Taylor, L., Floridi, L. and van der Sloot, B. (2016) *Group Privacy: New Challenges of Data Technologies*, Springer, Cham.

Tene, O. and Polonetsky, J. (2012) 'Big data for all: privacy and user control in the age of analytics', *Nw. J. Tech. and Intell. Prop.*, Vol. 11, p.xxvii.

Tene, O. and Polonetsky, J. (2017) 'Taming the golem: challenges of ethical algorithmic decision-making', *NCJL & Tech.*, Vol. 19, p.125.

Tenove, C. (2020) 'Protecting democracy from disinformation: normative threats and policy responses', *The International Journal of Press/Politics*, Vol. 25, pp.517–537, https://doi.org/10.1177/1940161220918740

Tufekci, Z. (2014) *Engineering the Public: Big Data, Surveillance and Computational Politics*, First Monday 19.

Väänänen-Vainio-Mattila, K., Olsson, T. and Häkkilä, J. (2015) 'Towards deeper understanding of user experience with ubiquitous computing systems: systematic literature review and design framework', in Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P. and Winckler, M. (Eds): *Human-Computer Interaction – INTERACT. 2015*, *Lecture Notes in Computer Science*, Springer International Publishing, Cham, pp.384–401, https://doi.org/10.1007/978-3-319-22698-9_26

Westin, A.F. (1967) 'Privacy and freedom', *Atheneum*, New York 7.

Westin, A.F. (2003) 'Social and political dimensions of privacy', *Journal of Social Issues*, Vol. 59, pp.431–453, https://doi.org/10.1111/1540-4560.00072

Winter, J. (2015) 'Algorithmic Discrimination: Big Data Analytics and the Future of the Internet, in Winter, J. and Ono, R. (Eds.): *The Future Internet: Alternative Visions, Public Administration and Information Technology*, Springer International Publishing, Cham, pp.125–140, https://doi.org/10.1007/978-3-319-22994-2_8

Wolf, G. (2018) 'In pursuit of privacy: an introduction to anonymization technologies', *Current Trends in Computer Sciences and Applications*, Vol. 1, pp.12–13.

Yang, F., Tian, T., Yao, H., Zhao, X., Zheng, T. and Ning, M. (2017) 'Anti-data mining on group privacy information', *International Conference on Human Centered Computing*, August, Springer, Cham, pp.481–491.

Zarsky, T. (2015) *Understanding Discrimination in the Scored Society (SSRN Scholarly Paper No. ID 2550248)*, Social Science Research Network, Rochester, NY.

Zarsky, T. (2017) *Incompatible: The GDPR in the Age of Big Data (SSRN Scholarly Paper No. ID 3022646)*, Social Science Research Network, Rochester, NY.

Zarsky, T. (2019) *Privacy and Manipulation in the Digital Age (SSRN Scholarly Paper No. ID 3321172)*, Social Science Research Network, Rochester, NY.

Zuboff, S. (2019) 'Surveillance capitalism', *Esprit*, Vol. 5, pp.63–77.

Zuiderveen Borgesius, F., Moeller, J., Kruikemeier, S.Ó.F.R., Irion, K., Dobber, T. and Bodó, B., de Vreese, C.H. (2018) *Online Political Microtargeting: Promises and Threats for Democracy (SSRN Scholarly Paper No. ID 3128787)*, Social Science Research Network, Rochester, NY.

Websites

The Tor Project | Privacy & Freedom Online [WWW Document] (n.d.) URL https://torproject.org (Accessed 22 March, 2021).

Welcome to Tor Metrics [WWW Document] (n.d.) URL https://metrics.torproject.org/ (Accessed 22 March, 2021).